



## OPEN ACCESS

## EDITED BY

Iwan Wicaksono,  
University of Jember, Indonesia

## REVIEWED BY

Nia Erlina,  
Ganesha University of Education, Indonesia  
Singgih Bektiarso,  
University of Jember, Indonesia

## \*CORRESPONDENCE

Pamela Flores  
✉ [pamela.flores@bio.lmu.de](mailto:pamela.flores@bio.lmu.de)

RECEIVED 31 January 2025

ACCEPTED 31 March 2025

PUBLISHED 30 April 2025

## CITATION

Flores P, Boone WJ and Neuhaus BJ (2025)  
Development and evaluation of an instrument  
to examine young children's knowledge of  
the biological concept of structure and  
function.  
*Front. Educ.* 10:1569123.  
doi: 10.3389/educ.2025.1569123

## COPYRIGHT

© 2025 Flores, Boone and Neuhaus. This is  
an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Development and evaluation of an instrument to examine young children's knowledge of the biological concept of structure and function

Pamela Flores<sup>1\*</sup>, William J. Boone<sup>2</sup> and Birgit J. Neuhaus<sup>1</sup>

<sup>1</sup>Department of Biology Education, Faculty of Biology, Ludwig-Maximilians-Universität in Munich, Munich, Germany, <sup>2</sup>Department of Educational Psychology, Miami University, Oxford, OH, United States

One important line of inquiry pursued by researchers has focused on the development of instruments to assess students' knowledge of different scientific concepts. In the field of life science, the vast majority of instruments discussed in the literature assess high school and primary school students. Very few assessment tools exist that focus on younger children. Given that the development of conceptual understanding is a central aim of early science education, there is a need for instruments that measure preschool children's knowledge in an age-appropriate manner. In this paper, we present an instrument that measures young children's knowledge of the biological concept of "structure and function." We made use of Rasch psychometric techniques to assess the measuring functioning of the instrument, including the analysis of dimensionality, item and person reliability, step ordering, and the range of item difficulty in relation to the range of person ability. Our analysis revealed that the instrument exhibited strong psychometric properties. The results indicate that children's conceptual knowledge can be characterized through two different cognitive activities: (1) recognizing the relation between biological structures and their respective functions, and (2) explaining these relations, and that these cognitive activities are related to each other. Further, the results reveal a great variance in children's abilities and contribute to the theory of previous studies regarding the link between children's previous experiences and their conceptual understanding. Overall, these results indicate that our instrument provides an appropriate tool to measure young children's conceptual understanding of structure and function. Implications for young children's science education are discussed.

## KEYWORDS

early science education, biology education, kindergarten, conceptual understanding, structure and function, disciplinary core ideas, test instrument

## Theoretical background

Conceptual knowledge has been defined as the knowledge about general principles and the relations and connections between specific facts or basic elements within a discipline (De Jong and Ferguson-Hessler, 1996; Krathwohl, 2002; Förtsch et al., 2020; Förtsch et al., 2018; Hussein, 2022; Nahdi and Jatisunda, 2020). As stated by Van Boxtel et al. (2000), conceptual knowledge "is reflected in the way students participate in activities that require the use of the concepts. Students have to become able to use scientific concepts to describe, explain and manipulate phenomena" (p. 312). Krathwohl (2002) defines several cognitive activities that

refer to “what is to be done with or to” that knowledge (p. 213). These cognitive activities constitute the mental process of constructing meaning, and therefore reflect a person’s conceptual knowledge (Mayer, 2002). Of particular relevance for our investigation are the two cognitive activities named *recognize* and *explain*. *Recognize* refers to a person’s ability to identify a piece of information as consistent with their own knowledge base. *Explain* refers to a person’s ability to construct and use cause-effect models when giving meaning to an observed phenomenon (Mayer, 2002). Similarly, in the field of early science, Tolmie et al. (2016) identified three core abilities, including the ability to make accurate observation, the ability to recognize and reason about causal connections, and the ability to explore the mechanism that explains these connections (see also Lin et al., 2020).

The development of conceptual knowledge is a crucial aim of science education, not only at primary and secondary school (Kultusministerkonferenz, 2004; NGSS Lead States, 2013), but also at the preschool level (Anders et al., 2018; Staatsinstitut für Frühpädagogik München, 2024; Steffensky, 2017). In the latter, the idea is not for children to achieve certain pre-defined, specific learning goals before entering primary school, but rather to develop an initial understanding of the scientific concepts they naturally encounter in their everyday lives (Anders et al., 2018; Eshach, 2006; French, 2004; Gelman and Brennehan, 2004; Möller and Steffensky, 2010; Steffensky, 2017).

In the domain of life sciences, three core concepts have been identified and used to structure and promote students’ learning in Germany: system, development, and structure and function (Kultusministerkonferenz, 2004). The concept of structure and function refers to the relation that exists between certain features of an organism and the purpose they serve, and thus represents one of the most important characteristics of living beings. Examples of this are the relation between a duck’s feet form and its ability to swim, or the relation between the structure of a plant’s xylem system and its ability to transport water in an upward direction. Given that young children constantly gather experiences with animals, plants and their own bodies in everyday life and show an intrinsic motivation to learn about these topics, the concept of structure and function seems to be naturally appropriate for science learning at the preschool level.

The importance of conceptual knowledge as a science learning goal translates into a need for valid and reliable assessment tools. Few instruments have been developed to measure children’s knowledge of the concept of structure and function. The instruments that do exist, focus mainly on primary and secondary school students. Förtsch et al. (2018), for example, developed a paper-pencil test to measure the structure-and-function knowledge of students at the secondary level. The test consisted of factual knowledge tasks, in which participants were asked to name one or more biological structures, and conceptual knowledge tasks, where participants were asked to describe one relation, e.g., describe a specific biological structure based on a given function (Förtsch et al., 2018). An instrument focusing on primary school students was developed by Kümpel (2019). In this test, three knowledge levels were considered: (1) the factual level, which concerns children’s ability to recall certain terms (e.g., the names of an animal’s body parts), (2) the relational level, which involves children’s ability to describe the relation between a specific biological structure and its function, and (3) the conceptual level, which involves participants’ understanding of a general principle (e.g., participants are asked to give an explanation about why birds have different beak shapes). Reiser et al. (2024) developed a “Measure instrument for the

Understanding of Structural-functional Correlations of the Locomotor System (MUSCLS).” This instrument is designed for children aged 10–14 years and consists of 10 drawing task items that capture children’s conceptual knowledge of the contraction of muscles, the antagonist principle of musculature, and the action of muscles over a joint.

This type of paper-and-pencil tests, although useful for the primary and secondary school context, cannot be implemented at the preschool level, as young children usually cannot write nor read and their linguistic skills are still evolving. Therefore, for early science education efforts, instruments must be able to address young children’s knowledge in an age-appropriate manner. A small number of studies have addressed preschool-aged children’s knowledge of the biological concept of structure and function using different assessment methodologies. Samarapungavan et al. (2009) developed the Science Learning Assessment. This instrument was utilized in the studies of Samarapungavan et al. (2008) and Samarapungavan et al. (2011) to investigate preschoolers’ learning of different biological concepts after participating in a project about the life cycle of monarch butterflies. It has also been used by Booth et al. (2022) to measure 5- and 6-year old children’s scientific literacy. Regarding the concept of structure and function, this instrument includes items in which children are asked to name the function of a butterfly’s different body parts, e.g., its legs and mouth. Thus, they are asked to match specific biological structures with their respective functions. Anderson et al. (2014) investigated preschool children’s knowledge of structure and function in plants by making use of three sources of evidence (the “Draw-A-Plant” instrument, a plant survey, and semi-structured interviews). In the “Draw-A-Plant” instrument, children’s drawings were rated based on whether they included certain structural elements of a plant, e.g., its leaves and roots, and certain factors needed to survive, e.g., water and the sun. In the survey, participants were asked to select from a set of pictures, the pictures that depicted plants. Students were also asked to select from another set of pictures, the pictures that depicted things that plants need to survive. This was followed by an interview in which children were asked about their drawings and their survey responses, with the aim of further detailing children’s reasoning behind their selections and answers. A study of Ahi (2017) focused on measuring young children’s understanding of the structure–function relationships of the digestive system. Participants were provided with an illustration of the outline of a human body. During a one-on-one interview, they were asked to draw and describe the path they think the food follows after being eaten, name the organs that they think are part of the process and describe the function the organs fulfil. Westerberg (2024) developed a comprehensive assessment tool to examine preschool children’s science and engineering knowledge and skills. It reflects a three-dimensional model that consists of disciplinary core ideas, science and engineering practices, and crosscutting concepts. Here, structure and function is defined as one of several crosscutting concepts. During the implementation, children are read aloud a question or prompt and are asked to respond by selecting from four illustrated response options. The instrument consists of 48 items, out of which only three refer to the relation between structure and function.

These instruments are useful tools for comprehending children’s understanding of the relation between structure and function, but they have three noteworthy limitations. First, the instruments only cover either a very specific content, such as plants or the human digestive tract, or a domain-general conceptualization that does not

allow for differentiation from other crosscutting concepts. Second, they generally require participants to merely match structures and functions and do not focus on investigating students' reasoning behind their selection. And third, the methodologies used do not take into consideration the different degrees of difficulty that can exist between items. There is therefore a need for an instrument to examine young children's knowledge of structure and function that focuses on the domain of life sciences and covers a wide range of organisms, reveals children's reasoning, and takes into account different levels of difficulty.

## This study

In this study, we propose that a young student's conceptual knowledge of structure and function can be identified through two different but related cognitive activities: *recognize*, which refers to a person's ability to identify the relation between a specific biological structure and its respective functions, and *explain*, which refers to the ability of a person to describe and explain which specific characteristics of a given structure allow the structure to fulfil its function. Based on this, we assess young children's knowledge of the biological concept "structure and function" by measuring children's ability to *recognize* structural-functional relations by matching biological structures with the functions they serve, and children's ability to *explain* each relation. For this assessment, items are presented in a two-tier structure, in which each tier targets one cognitive activity. This test format is a common approach to measure students' knowledge and students' reasoning (Treagust, 1988). In this type of item structure, the 1st tier item is a multiple-choice or true/false question, whereas the 2nd tier item requires participants to justify their 1st tier answers by either providing an open response or choosing from a set of possible reasons the response they think is most similar to their own response (Liu et al., 2011; Treagust and Mann, 1998; Treagust, 1988). This item structure enables researchers to investigate whether participants' reasoning is based on a conceptual understanding of the topic being addressed. A number of studies have made use of two-tier instruments to measure high school students' knowledge of several biological topics, e.g., photosynthesis and plant respiration (Haslam and Treagust, 1987), gas exchange (Treagust and Mann, 1998) and plant growth and development (Lin, 2004). To the best of our knowledge, however, the instrument we present is the first instrument that makes use of the two-tier test format to measure the conceptual knowledge of preschool-aged children.

Our instrument addresses some of the limitations of current assessment tools (Ahi, 2017; Anderson et al., 2014; Samarapungavan et al., 2009; Westerberg, 2024), as it focuses on the structure and function relations within life sciences, covers a wide range of organisms and requires children to not only match structures and functions but also to describe and explain relationships. Further, we make use of Rasch psychometric technique to assess the measuring functioning of our instrument (Wright and Masters, 1982; Wright and Stone, 1979). The Rasch analysis includes the analysis of item fit, item and person reliability, step ordering, and assessing the range of item difficulty in relation to the range of person ability. Our discussion also considers the implications of the item ordering and spacing presented in the Wright maps.

The instrument and analysis presented here are a component of the doctoral dissertation completed by the lead author of this article (Flores, 2022).

## Methods

### Sample and procedure

The instrument was administered to a sample of 59 preschool children from 5 different preschools located in and around Munich, Germany (Flores, 2022). The participating children had an average age of 6 years and 3 months ( $SD = 0.44$ ). The interviews lasted on average 15 min and were audio recorded. To conduct the interviews, the interviewers received a script that contained the interview. The interviewer also received the drawings to be used for each interview question.

### Instrument development

Below we present the steps which were taken to create the final instrument. The instrument development process included a pilot data collection, which was then utilized to help us develop a final set of instrument items. Following the collection of data with the final instrument, we utilized psychometric techniques to evaluate instrument functioning and to compute the item difficulties. In doing so, we followed instrument development recommendations outlined in Boone et al. (2014).

### Pilot instrument

The first version of the instrument was administered to 74 preschool children, 31 1st grade and 46 2nd grade children. It contained seven items, each of which consisted of one multiple choice question with two or three options (a 1st-tier question) and one open question in which children were asked to justify their first answers (a 2nd-tier question).

This data collection provided several important insights which were used to inform the development of the item pool for the final instrument. First, the data collection allowed for a general appraisal of how feasible and age-appropriate the test was in terms of the type of interview, duration of the interview, and the use of the two-tier item structure. The pilot results also shed light on preschool children's understanding of the relation between structure and function, and, with this, the pilot results revealed the level of difficulty necessary to assess preschool-aged children's conceptual knowledge. Most importantly, participants' 2nd-tier answers, i.e., students' justifications of their 1st-tier answers, were qualitatively analyzed. The results were used as the basis to define the categories that were used to code children's answers in the final instrument.

Following analysis of the pilot data, it was decided that of the seven pilot items, five items would be used in the final version of the instrument. Two pilot items were kept as they were for the final instrument, and three pilot items were edited for the final instrument. In the three edited items, the wording of the questions was improved and/or one more option was included in the multiple-choice questions so that all 1st-tier questions had the same number of options. After the pilot data collection and analysis, new pictures were drawn for all new items by the author of this paper.

### Final instrument

The final instrument consists of nine items (two pilot items which were not edited for the final instrument, three items which were piloted and then edited based upon pilot results, and four new items)

that present specific structural-functional relationships found in a diverse group of organisms, such as insects, birds, mammals, and plants (see Table 1). The items present organisms well known to young children, e.g., ducks and squirrels, and behaviors that young children can relate to their own everyday lives, such as eating and moving. The items were reviewed by two experts in biology education and ten graduate students in biology education completing their last year of university.

Each item starts with a short introduction about the behavior of an organism. After this, children are asked about the relation of a specific biological structure of the given organism and the function it serves, and the students are required to answer by selecting one of three options. This way, they are required to match the structure with its function (*recognize* the relation). Afterwards, the interviewer asks them to justify their 1st-tier answers (*explain* the relation). Figure 1 shows one example item. In this item, the interviewer starts by telling the child that when they were at the lake, they saw a fish looking for food toward the bottom of the lake. The interviewer then asks the child what type of mouth do they think the fish had, and lets them choose one of the three options presented in Figure 1. Finally, the interviewer asks the child why they think the fish had that type of mouth.

For the coding of the 1st tier, test-takers score 1 point if they select the correct answer out of the three given options. These scores constitute the variable labeled *recognize*.

In order to evaluate students' 2nd-tier answers, we defined eight categories to which participants' statements can be categorized (see Table 2). Categories I-III refer to responses in which children display conceptual understanding by referring to a relevant structure, function, or relation between them. Categories IV-VII refer to statements in which children do not make use of their conceptual knowledge when justifying their 1st-tier answers, and category VIII refers to the cases in which participants' 1st-tier answer is incorrect but the 2nd-tier answer reveals an understanding of the structural-functional relationship. For example, if a student's answer to item Nr. 1 was "because with this mouth they can find food on the ground," that answer would be coded as category III. If the answer was "because I like the color blue," it would be coded as category VI. Statements that are assigned to categories I-III receive 1 point, whereas statements that correspond to categories IV-VIII do not. These scores students receive using Table 2 concern the variable called *explain*.

The 2nd-tier answers of 11 participants (17% of the sample) were categorized by two independent raters. The resulting interrater reliability showed very good values ( $\kappa = 0.87$ , 95% CI [0.78, 0.97],  $p < 0.001$ ). Thus, following coding of student answers, each student had a *recognize* score and an *explain* score for each item.

The final instrument, including the interview script for each item, the corresponding pictures and the coding scheme, can be found in the Supplementary File 1 (English version) and Supplementary File 2 (German version).

TABLE 1 List of the items and the organism, structure, and function presented.

Item nr.	Organism	Structure	Function
1	Fish	Position of its mouth	Looking for food from the lakebed
2	Duck	Webbed feet	Swimming
3	Mosquito	Stinging mouth	Stinging through flesh
4	Flower	Shape, colors and parts	Attracting insects
5	Seed	Wing-shape	Flying away
6	Conifer needle	Wax layer	Protecting against cold
7	Squirrel	Nest with two entries	Escaping from predators
8	Mouse	Length of its tail	Balancing
9	Fleas, Rabbits, Kangaroos	Form of the hind legs	Jumping

"Yesterday I was at the lake and saw a fish in the water. This fish was swimming at the bottom of the lake and was looking for food on the lakebed. Here you can see three pictures of fishes, and if you look carefully, you can notice that they all have a different types of mouth."

1<sup>st</sup>-tier question: "Which mouth do you think the fish had that I saw yesterday?"

2<sup>nd</sup>-tier question: "Why do you think it's good for that fish to have that mouth?"

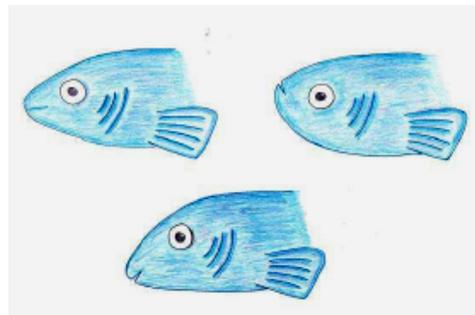


FIGURE 1 Item Nr. 1 – script and drawings (translated from German).

TABLE 2 Coding categories for children's 2nd-tier answers.

Category	Content of 2nd-tier answers
I	Structure
II	Function
III	Structure and Function
IV	Previous experiences (e.g., seen it on TV)
V	Fantasy
VI	No relation to question/ incomprehensible answer
VII	No answer/ "Do not know"
VIII	Structure and Function but wrong 1st-tier answer

## Instrument evaluation

Rasch analysis was performed using the program Winsteps (Linacre, 2021b). Rasch analysis is a psychometric approach to evaluate the measurement functioning of an instrument (McLaughlin et al., 2023; Rasch, 1960; Vasseleu et al., 2021). Unlike traditional approaches, which compute the measured trait of an individual immediately and solely using raw scores, the Rasch approach takes the differing degrees of difficulty between items into consideration. When Rasch techniques are used, the individuals' nonlinear raw scores are converted into linear "person measures." There are many reasons for using Rasch techniques. One is that person abilities and item difficulties are expressed on the same scale so that they can be directly compared to each other.

The Rasch approach can be used not only to compute linear person measures and item measures but Rasch facilitates the analysis of numerous qualities of an instrument to evaluate instrument functioning. For example, Rasch allows the evaluation of (1) the dimensionality of the construct (whether all items measure a single trait), (2) the item and person reliabilities (whether the measures of item difficulty and of person ability are overall consistent), (3) the step ordering (whether the average measure of respondents who answer an item correctly is consistently higher than the average measure of those who do not answer it correctly), and (4) Wright Maps (Boone et al., 2014).

Wright Maps are particularly useful in that the maps provide a way with which the results of an instrument evaluation can be visualized. On one single linear measurement scale, Wright Maps can present both the items according to their difficulty level as well as the respondents according to their ability level (Boone, 2016). Through the evaluation of the Wright Maps it is possible to assess the location of items along the difficulty scale, ceiling and floor effect, and test item targeting (which helps determine whether the items generally are at the correct level of difficulty for a given group of participants; Finger et al., 2012).

## Results

### Psychometric results

#### Dimensionality (item fit)

The dimensionality of the construct is evaluated through the mean-square (MNSQ) Infit and Outfit values of each item, with ideal values being close to 1. Table 3 displays the MNSQ Item Infit and Outfit values of each item and tier as well as the corresponding mean values. All values are located within the range of 0.5–1.5, which is considered satisfactory

in studies with small sample sizes (Linacre, 2002; Wright and Linacre, 1994).

Another important piece of information, which can be revealing in terms of investigating an instruments' functioning, is the number of computational iterations of Winsteps that are necessary to obtain good estimates from the data when running the Rasch software. According to Linacre (1987), "lack of convergence is an indication that the data do not fit the model well". This means fewer iterations of Winsteps to converge indicates better fit of the model to the data. In our data set, only 4 iterations were necessary for the variable *recognize* analysis, and only 6 iterations were necessary for the variable *explain* analysis. This is considered a very small number of iterations.

### Item and person reliability

Table 4 presents the item and person reliabilities for the Rasch *recognize* analysis and the Rasch *explain* analysis. The values of item reliability are close to those considered satisfactory (0.90 or higher, according to Malec et al., 2007). In contrast to this, the person reliability shows rather low values in both variables. This is not unexpected given the low number of respondents. Our discussion generally emphasizes observations concerning item characteristics and patterns.

### Step ordering

In a well-functioning instrument, it is expected that the average ability of respondents who answer an item correctly should always be higher than that of respondents who answer incorrectly (often this is called an investigation of step ordering). Table 5 presents data to facilitate this analysis. For each item the average of respondents followed the pattern one expects to observe in a well-functioning instrument. A higher person ability in logits indicates better test performance.

### Ceiling effect, floor effect, targeting

A common part of a Rasch analysis is to assess if there is a Floor Effect (do any students get a 0 for every item) and a Ceiling Effect (do any students get a 1 for every item). The goal in a well-functioning instrument is that less than 5% of the respondents' responses should be at the Floor or Ceiling (Fisher, 2007). In our data set, in the variable *recognize*, only 1 child was at the Ceiling (1.6%), and none at the Floor. In the variable *explain*, no one was at the Ceiling, and 10 were at the Floor (16.1%). This was not surprising, as the variable *explain* was expected to be more difficult. Test item targeting concerns the goal of having an appropriate level of average item difficulty with regard to the average ability level of respondents. A well-functioning test should not be too difficult or too easy for respondents. A rule used by Finger et al. (2012) is that the difference between average item difficulty and average person ability should be less than one logit. For both *explain* and *recognize*, the difference was less than one logit, suggesting good test item targeting.

### Wright maps

Wright Maps of the variables *recognize* and *explain* are presented in Figure 2. These reveal that there is a wide distribution of the items along the difficulty scale, although there are some overlapping items (e.g., between Items Nr. 5 and 9 in *recognize*) and some gaps between items (e.g., between Items Nr. 7 and 8 in

TABLE 3 MNSQ Infit values, MNSQ Outfit values, item values.

Item nr.	1st tier ( <i>recognize</i> )			2nd tier ( <i>explain</i> )		
	Measure (logits)	MNSQ Infit	MNSQ Outfit	Measure (logits)	MNSQ Infit	MNSQ Outfit
1	1.53	1.17	1.42	0.72	1.09	1.01
2	-0.13	0.83	0.71	-0.50	0.90	0.84
3	-0.43	0.94	0.81	-1.13	1.14	1.07
4	-2.46	0.95	0.91	-1.81	0.85	1.15
5	-0.88	1.09	1.41	-1.02	0.98	0.94
6	1.00	1.02	1.08	0.14	0.94	0.80
7	1.71	0.94	0.86	1.13	0.84	0.59
8	0.66	0.92	1.01	2.46	1.04	0.53
9	-1.00	1.02	1.02	0.03	1.27	1.47
Mean	0.00	0.99	1.03	0.00	1.01	0.93

TABLE 4 Item and person reliability for the variables *recognize* and *explain*.

	1st tier ( <i>recognize</i> )	2nd tier ( <i>explain</i> )
Item reliability	0.92	0.81
Person reliability	0.41	0.13

*explain*). The maps reveal that there is a good overlap between the range of item difficulty and the range of person ability in both variables.

### Correlations

There is a significant correlation between the student *recognize* measures and the student *explain* measures ( $r = 0.78, p < 0.01$ ), which suggests that the variables might be related to each other.

### Discussion

In this paper, we present the development and evaluation of a new instrument that measures young children’s knowledge of the concept of structure and function in the domain of life science. This instrument provides measures not possible with previous instruments that aim to measure related knowledge from very young children (Anderson et al., 2014; Ahi, 2017; Samarapungavan et al., 2008; Samarapungavan et al., 2011; Westerberg, 2024). In our new instrument, the content of the questions does not refer to a single animal or plant but to a wide variety of organisms, including plants, insects, and several vertebrates. Our instrument is unique in that the two-tier item structure allows for the measurement of two different cognitive activities that reflect children’s conceptual knowledge, i.e., their ability to *recognize* the relation between specific structures and functions, and their ability to *explain* these relations. To the best of our knowledge, this is the first time this item format has been implemented to measure preschool children’s understanding of a scientific concept. Third, through the implementation of Rasch techniques, the analysis took into consideration the differing degrees of item difficulty and, for example, allowed the construction of the informative Wright Maps.

### Evaluation of the instrument

Evaluation of the psychometric properties of the instrument utilizing Rasch technique included an evaluation of dimensionality, step ordering, item and person reliability, Ceiling Effect, Floor Effect, test targeting, and Wright Maps. The analysis steps we took are similar to those taken in other studies (e.g., Dorfner et al., 2019; Jüttner et al., 2013).

The analysis of dimensionality showed that the mean square infit and outfit values of all items were located within the acceptable range for small sample sizes and that only a small number of iterations of Winsteps were necessary for both variables *recognize* and *explain*, which provides added evidence that the data fits the Rasch model. The item reliability of both variables showed satisfactory values above 0.80, whereas the person reliability values were rather low. Even though the low person reliability was unavoidable in the context of this implementation (see limitations), the high item reliability indicates that our instrument can reliably provide an item ordering for the *recognize* items and the *explain* items. The analysis of step ordering showed the expected pattern, as the average person ability value of test-takers that answered each item correctly was consistently higher than that of test-takers that answered incorrectly; another indicator of good instrument functioning.

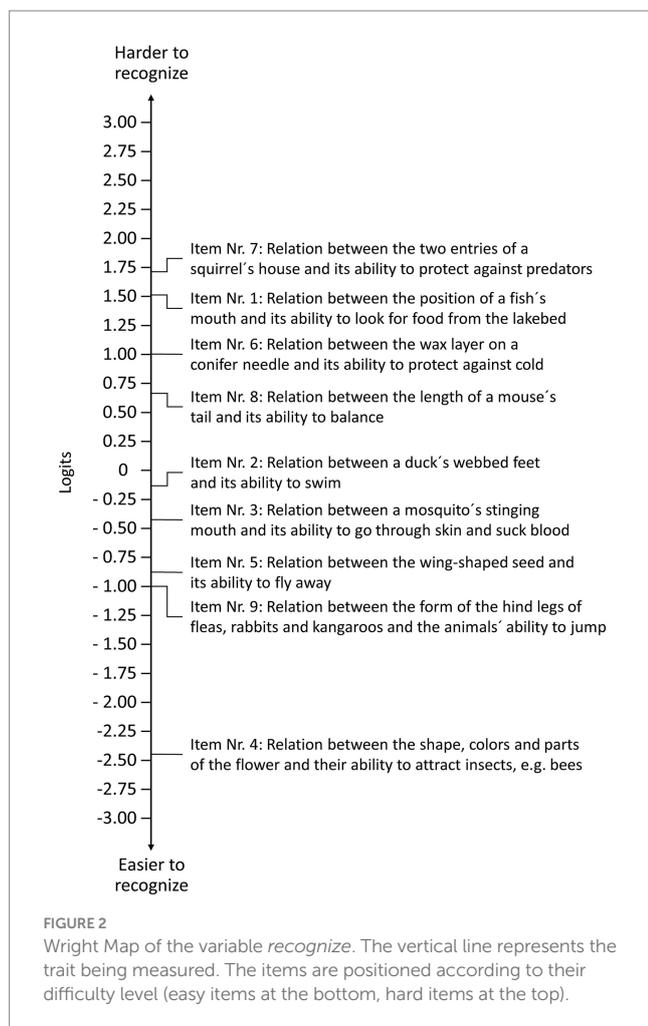
The Wright Maps were evaluated with regard to the position of items along the difficulty scale, ceiling and floor effect, and test item targeting. Items were located with a good distancing between each other, except for a few exceptions. Only a small expected amount of students were at the Floor or Ceiling. For both *recognize* and *explain*, the difference between average item difficulty and average person ability is less than one logit, suggesting good test item targeting. These analyses all reveal that our instrument provides an appropriate tool to measure young children’s understanding of the concept of structure and function as reflected by the two cognitive activities *recognize* and *explain*.

### Key findings about preschool children’s conceptual knowledge

Our findings further confirm previous studies in the field of early science education which document that preschool-aged children possess a basic understanding of the scientific concepts that are

TABLE 5 Average ability of respondents who answered correctly and incorrectly each item and tier.

Item nr.	1st tier ( <i>recognize</i> )		2nd tier ( <i>explain</i> )	
	Average ability of respondents whose answer was correct (logits)	Average ability of respondents whose answer was incorrect (logits)	Average ability of respondents whose answer was correct (logits)	Average ability of respondents whose answer was incorrect (logits)
1	1.38	0.50	0.50	-1.50
2	1.33	-0.24	0.32	-1.98
3	1.21	-0.19	-0.03	-2.07
4	0.92	-0.50	-0.05	-2.54
5	1.04	0.03	0.09	-2.13
6	1.45	0.28	0.51	-1.74
7	1.73	0.37	1.02	-1.50
8	1.45	0.12	1.25	-1.22
9	1.06	-0.15	0.08	-1.58



commonly part of science education in school, specifically the biological concept of structure and function (Ahi, 2017; Anderson et al., 2014; Samarapungavan et al., 2008; Samarapungavan et al., 2011; Westerberg, 2024). Further, this study contributes to the existing literature with four key findings based on our empirical results. First,

children's understanding of the concept can be characterized through two different cognitive activities, i.e., their ability to *recognize* the relation between structure and function and their ability to *explain* these relations in a concept-based manner. Second, the significant correlation of student *recognize* measures and student *explain* measures suggests that the cognitive activities *recognize* and *explain* are related to some degree. These results go in line with the framework used by Tolmie et al. (2016) and Lin et al. (2020) to characterize early science core abilities, which includes the abilities to recognize and reason about causal connections and to explore the mechanisms that explain these connections.

The student *recognize* ability measures has a mean of 0.83 logits with a maximum measure of 3.97 and a minimum measure of -1.64 logits. The student *explain* ability measure has a mean of -1.02 logits with a maximum measure of 2.64 and a minimum measure of -3.93 logits. This leads to the third key finding, as this empirically demonstrates a great variance in preschool-aged children's abilities to match structures with their respective functions and to explain these relationships. We suggest this variance may be explained by children's varied language skills. This aligns with the idea that a certain level of linguistic skills is fundamental to participate in the social interactions and communicative process that characterize guided learning situations (Akerson et al., 2000; Lemke, 1990; Van Boxtel et al., 2000; Vygotsky, 1978).

The fourth key finding refers to which structural-functional relations are easier or harder to recognize and explain and why. In the Wright Map of the variable *recognize*, the items appear in the following sequence from easiest to hardest: Item Nr. 4, 9, 5, 3, 2, 8, 6, 1, 7 (see Figure 2). The easiest relations to recognize are thus the one between the parts of a flower and the function of attracting insects (Item Nr. 4), between the form of certain animals' hind legs and the function of jumping (Item Nr. 9), and between the wing-shaped form of a seed and the function of flying away (Item Nr. 5). The hardest relations to recognize are the one between the structure of a squirrel's nest and the function of fleeing from predators (Item Nr. 7), between the position of a fish's mouth and the function of feeding (Item Nr. 1), and between the structure of a conifer needle and the function of protection from the cold (Item Nr. 6). In the Wright Map of the variable *explain*, the sequence is the following, from easy to difficult: Item Nr. 4, 3, 5, 2, 9,

6, 1, 7, 8 (see Figure 3). Similar to the variable *recognize*, the Items Nr. 4 and 5 are positioned among the easiest relations to explain, and Items Nr. 6, 1, and 7 are among the hardest relations to explain.

A possible explanation for the ordering of these items, from easy to more difficult, may lie on children’s previous contact with the specific content depicted in the questions. As mentioned by French (2004), young children’s understanding of biological concepts builds through their everyday interaction with the world around them. Barrutia and Díez (2019), for example, argue that children’s previous opportunities to observe only the visible parts of plants might be the reason why roots were rarely represented in their drawings. Similarly, Reiber et al. (2019) points out that children’s understanding of human organs and their functions up to the age of nine is limited to the organs that they can perceive directly with their own senses. Previous studies show that preschool-aged children understand structure and function relations that they could observe, e.g., the role of different body parts of the monarch butterflies (Samarapungavan et al., 2008; Samarapungavan et al., 2011), but are not able to recognize such relations in cases that they were not able to observe, e.g., the gas exchange of plants or the role of the intestines in the digestive system (Ahi, 2017; Anderson et al., 2014).

Looking at the difficulty sequences, it becomes clear that children’s previous contact and observation of the specific examples may in fact explain the ordering of the items. While spending time in nature,

children may already have observed bees flying towards specific parts of the flowers (Item Nr. 4), for example. They may have rabbits as pets or have seen kangaroos in movies and books, and are familiar with their way of moving (Item Nr. 9). In contrast to this, children most probably never had the opportunity to observe a squirrel escaping a predator (Item Nr. 7) or a fish feeding underwater (Item Nr. 1), and the wax layer on a conifer needle is not observable with bare eyes (Item Nr. 6). Looking at the item sequences in both Wright Maps, one can see that Items Nr. 8 (the relation between the structure of a mouse’s tail and the function of balancing) and Nr. 9 (the relation between the form of certain animals’ hind legs and the function of jumping) seem to be easier to recognize than to explain. This suggests that even though children are familiar with the relation between a certain structure and its function, they are not always able to define what specific characteristic of the structure allows it to fulfil its function. They thus may know that a long tail is important for balancing or that long and strong hind legs allow an animal to jump, but not fully understand why it is so. This differentiation is what makes it so important to analyze these two different cognitive abilities separately.

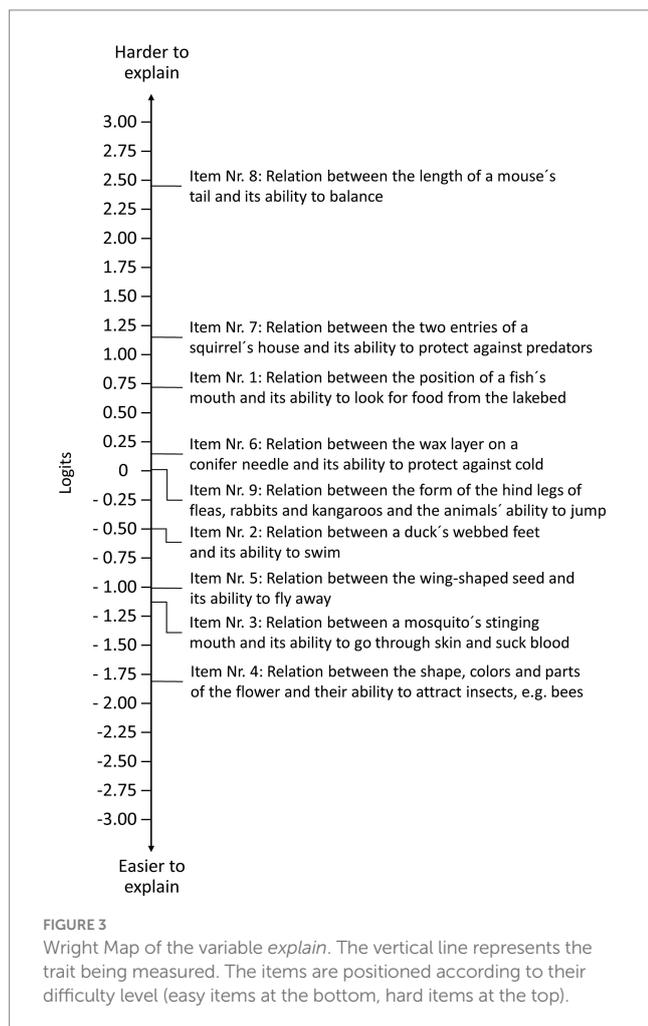
The sequence of the items along the difficulty scale found in this study contributes to the theory gathered by several previous studies about the link between children’s previous experiences with certain organisms and their conceptual understanding (Ahi, 2017; Anderson et al., 2014; Barrutia and Díez, 2019; Reiber et al., 2019; Samarapungavan et al., 2008; Samarapungavan et al., 2011). This theory-based explanation of the pattern of items in the Wright maps is, in turn, another important indicator of a well-functioning instrument (Boone et al., 2014).

### Limitations

The main limitation of this study is the low person reliability that was found during the instrument evaluation. According to Linacre (2021a), person reliability values depend on the variance of sample ability, the sample-item targeting, the length of the test, and the number of categories per item. In the Rasch analysis we observed a wide range of sample ability and an acceptable sample item targeting. We suspect the low person ability may be related to the number of items and the number of ratings categories per item. The number of categories per item is inevitably low due to the nature of the test format, as both the 1st-tier and the 2nd-tier answers are meant to be coded dichotomously. With nine items, the length of the test could contribute to person reliability values. This, however, was also inevitable during this implementation because children were tested for a variety of skills and participated in a learning activity, so it was important to keep the tests short as to not overreach their attention span. In future studies we might explore removing some items with similar difficulty and authoring items to fill some gaps in the Wright map. However, we feel with those very young students, there is truly a limit to the number of items which can be presented.

### Implications for research and praxis

The instrument development and evaluation presented here serves as a basis upon which future research can be built. Given that the main limitation of this study might be the length of the test, the first step in future



implementations, both in research and praxis, would be to increase the number of items. To achieve this, the nine items presented here can be used as models for the formulation of new items that depict the same type of content and structure following the insights from our instrument development and evaluation. To prevent the test from taking too long, the items could be divided and given in two administrations and some items with similar difficulty, one of two items could be removed to make room for items to fill gaps in the distribution of items on the two Wright maps.

The administration of our instrument with 59 preschool children revealed that the format of one-on-one interviews with items that include a short introduction to the content, two short questions and accompanying drawings, is an excellent method to assess the conceptual knowledge of young children. We therefore recommend researchers consider using this format with children that cannot write and read and thus are not able to participate in other test formats, such as paper-pencil tests. For future use of such instruments, we recommend that the interviewer memorizes the script beforehand to give an authentic feeling of casual conversation while ensuring that the questions are formulated in the same way with all the interviewees. Further, we suggest recording the interaction on video or audio to minimize the notetaking during the interview.

This instrument has been used to investigate the effect of guided inquiry on preschool children's conceptual knowledge (Flores, 2022). In the future, it can also serve the investigation of other research questions. With a wider age range, it could be used to characterize the development of children's conceptual knowledge, e.g., from preschool (6-year-olds) to the end of primary school (10-year-olds), as well as to identify the different types of preconceptions the different age groups hold about the relation between structure and function. An interesting aspect that could not be addressed in the framework of this study is the distribution of the categories of children's answers to the 2nd-tier question, that is, how many children refer to the structures and/or functions, to previous experiences, to fantasy, etc. Our instrument, in combination with other instruments, could be used to identify potential predictors of conceptual knowledge. This may include investigating the hypothesized link between children's previous contact with certain organisms and their understanding of the relation between its structures and function (Ahi, 2017; Anderson et al., 2014; Barrutia and Díez, 2019; Reiber et al., 2019; Samarapungavan et al., 2008; Samarapungavan et al., 2011), or exploring the link between children's domain-general scientific reasoning skills and their domain-specific conceptual knowledge (Klemm and Neuhaus, 2017; Koerber and Osterhaus, 2019; Sodian et al., 1991).

The evaluation of children's conceptual knowledge is not only relevant in research but also in the practice of classroom science education. This instrument offers an opportunity for preschool and primary school teachers to examine their students' level of understanding in a way that feels as natural as a conversation for the children. There are several scenarios in which this instrument could be of benefit to teachers. One important example is the transitioning of students from preschool to 1st grade. Further, teachers could evaluate their children's level of conceptual knowledge before science lessons in order to fine tune their lessons accordingly, or administer the instrument after science lessons in order to evaluate students' learning progress.

The ordering of items along the difficulty scale in the Wright map suggests that, when conceiving lessons about scientific concepts for young children, it is important to start with topics that are embedded into students' everyday life. In the field of life

science, and specifically regarding the concept of structure and function, this translates into using animals and plants students may have already been in contact with, either in their own lives or through books or movies, such as colorful flowers (Item Nr. 4) or rabbits and kangaroos (Item Nr. 9). Later on, teachers could make use of examples of relations between structure and function that young children may have never been able to observe directly, such as the entrance of a squirrel's nest (Item Nr. 7) and the position of a fish's mouth under water (Item Nr. 1).

Finally, of great interest for the teachers' teaching of preschool science is the fact that children's ability to recognize structure–function relations does not necessarily mean that they are able to explain them. This could be used as a starting point in science lessons, by first making children aware of this gap in their understanding, producing a so-called cognitive conflict (Nachreiner et al., 2015), and then motivating students to explore the specific mechanisms of such relations in order to not only recognize them but also be able to explain them correctly.

In general, our findings demonstrate that young children already possess an important basic understanding of the scientific concept of structure and function, especially regarding animals and plants they have been able to observe in their everyday lives. This indicates that concept-based science lessons are not only appropriate but also beneficial for preschool-aged children, as these types of lessons allows them to further develop their basic conceptual understanding, which is arguably the main goal of early science education.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

PF: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. WB: Formal analysis, Methodology, Supervision, Writing – review & editing. BN: Conceptualization, Data curation, Formal analysis, Methodology, Supervision, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was kindly supported by the Elite Network of Bavaria under Grant K-GS-2012-209.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any

product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2025.1569123/full#supplementary-material>

## References

- Ahi, B. (2017). Thinking about digestive system in early childhood: a comparative study about biological knowledge. *Cogent Educ.* 4:1278650. doi: 10.1080/2331186X.2017.1278650
- Akerson, V. L., Flick, L. B., and Lederman, N. G. (2000). The influence of primary children's ideas in science on teaching practice. *J. Res. Sci. Teach.* 37, 363–385. doi: 10.1002/(SICI)1098-2736(200004)37:4<363::AID-TEA5>3.0.CO;2-%23
- Anders, Y., Hardy, I., Pauen, S., Ramseger, J., Sodian, B., and Steffensky, M. (2018). Early science education—goals and process-related quality criteria for science teaching. Barbara Budrich.
- Anderson, J. L., Ellis, J. P., and Jones, A. M. (2014). Understanding early elementary children's conceptual knowledge of plant structure and function through drawings. *CBE—Life Sci. Educ.* 13, 375–386. doi: 10.1187/cbe.13-12-0230
- Barrutia, O., and Diez, J. R. (2019). 7 to 13-year-old students' conceptual understanding of plant nutrition: should we be concerned about elementary teachers' instruction? *J. Biol. Educ.* 55, 196–216. doi: 10.1080/00219266.2019.1679655
- Boone, W. J. (2016). Rasch analysis for instrument development: why, when, and how? *CBE—life sciences. Education* 15:rm4. doi: 10.1187/cbe.16-04-0148
- Boone, W. J., Staver, J. R., and Yale, M. S. (2014). Rasch analysis in the human sciences. Springer.
- Booth, A. E., Shavlik, M., and Haden, C. A. (2022). Exploring the foundations of early scientific literacy: Children's causal stance. *Dev. Psychol.* 58, 2302–2309. doi: 10.1037/dev0001433
- De Jong, T., and Ferguson-Hessler, M. G. (1996). Types and qualities of knowledge. *Educ. Psychol.* 31, 105–113. doi: 10.1207/s15326985Sep3102\_2
- Dorfner, T., Förtsch, C., Boone, W., and Neuhaus, B. J. (2019). Instructional quality features in videotaped biology lessons: content-independent description of characteristics. *Res. Sci. Educ.* 49, 1457–1491. doi: 10.1007/s11165-017-9663-x
- Eshach, H. (2006). Science literacy in primary schools and pre-schools. Dordrecht, The Netherlands: Springer.
- Finger, R. P., Fenwick, E., Pesudovs, K., Marella, M., Lamoureux, E. L., and Holz, F. G. (2012). Rasch analysis reveals problems with multiplicative scoring in the macular disease quality of life questionnaire. *Ophthalmology* 119, 2351–2357. doi: 10.1016/j.ophtha.2012.05.031
- Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions* 21:1095.
- Flores, P. (2022). *Early science education – exploring preschool children's basic conceptual knowledge along with their involvement and preschool teachers' professional competence* (Doctoral dissertation, Ludwig-Maximilians-Universität München). doi: 10.5282/edoc.31481
- Förtsch, C., Dorfner, T., Baumgartner, J., Werner, S., von Kotzebue, L., and Neuhaus, B. J. (2020). Fostering students' conceptual knowledge in biology in the context of German National Education Standards. *Res. Sci. Educ.* 50, 739–771. doi: 10.1007/s11165-018-9709-8
- Förtsch, C., Heidenfelder, K., Spangler, M., and Neuhaus, B. J. (2018). How does the use of core ideas in biology lessons influence students' knowledge development? *Zeitschrift Didaktik Naturwissenschaften* 24, 35–50. doi: 10.1007/s40573-018-0071-1
- French, L. (2004). Science as the center of a coherent, integrated early childhood curriculum. *Early Child. Res. Q.* 19, 138–149. doi: 10.1016/j.ecresq.2004.01.004
- Gelman, R., and Brenneman, K. (2004). Science learning pathways for young children. *Early Child. Res. Q.* 19, 150–158. doi: 10.1016/j.ecresq.2004.01.009
- Haslam, F., and Treagust, D. F. (1987). Diagnosing secondary students' misconceptions of photosynthesis and respiration in plants using a two-tier multiple choice instrument. *J. Biol. Educ.* 21, 203–211. doi: 10.1080/00219266.1987.9654897
- Hussein, Y. F. (2022). Conceptual knowledge and its importance in teaching mathematics. *Middle Eastern J. Res. Educ. Soc. Sci.* 3, 50–65. doi: 10.47631/mejress.v3i1.445
- Jüttner, M., Boone, W., Park, S., and Neuhaus, B. J. (2013). Development and use of a test instrument to measure biology teachers' content knowledge (CK) and pedagogical content knowledge (PCK). *Educ. Assess. Eval. Account.* 25, 45–67. doi: 10.1007/s11092-013-9157-y
- Klemm, J., and Neuhaus, B. J. (2017). The role of involvement and emotional well-being for preschool children's scientific observation competency in biology. *Int. J. Sci. Educ.* 39, 863–876. doi: 10.1080/09500693.2017.1310408
- Koerber, S., and Osterhaus, C. (2019). Individual differences in early scientific thinking: assessment, cognitive influences, and their relevance for science learning. *J. Cogn. Dev.* 20, 510–533. doi: 10.1080/15248372.2019.1620232
- Krathwohl, D. R. (2002). A revision of bloom's taxonomy: an overview. *Theory Pract.* 41, 212–218. doi: 10.1207/s15430421tip4104\_2
- Kultusministerkonferenz, K. (2004). Bildungsstandards im Fach Biologie für den mittleren Schulabschluss. Neuwied: Luchterhand. (Analog für Chemie und Physik).
- Kümpel, N. (2019). Förderung des konzeptuellen Wissens durch den Einsatz von Basiskonzepten und problemorientierten Kontexten im Heimat- und Sachunterricht der Grundschule [Unpublished doctoral dissertation, Ludwig-Maximilians-Universität München].
- Lenke, J. L. (1990). Talking science: Language, learning, and values. Ablex Publishing Corporation.
- Lin, S.-W. (2004). Development and application of a two-tier diagnostic test for high school students' understanding of flowering plant growth and development. *Int. J. Sci. Math. Educ.* 2, 175–199. doi: 10.1007/s10763-004-6484-y
- Lin, X., Yang, W., Wu, L., Zhu, L., Wu, D., and Li, H. (2020). Using an inquiry-based science and engineering program to promote science knowledge, problem-solving skills and approaches to learning in preschool children. *Early Educ. Dev.* 32, 695–713. doi: 10.1080/10409289.2020.1795333
- Linacre, J. M. (1987). Rasch estimation: iteration and convergence. *Rasch Measure. Transact.* 1, 7–8.
- Linacre, J. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measure. Transact.* 16:878.
- Linacre, J. (2021a). Winsteps® Rasch measurement computer program User's guide. Version 5.1.7. Portland, Oregon: Winsteps.com.
- Linacre, J. (2021b). Winsteps® (version 5.1.7.) [computer software]. Portland, Oregon: Winsteps.com.
- Liu, O. L., Lee, H.-S., and Linn, M. C. (2011). Measuring knowledge integration: validation of four-year assessments. *J. Res. Sci. Teach.* 48, 1079–1107. doi: 10.1002/tea.20441
- Malec, J. F., Torsher, L. C., Dunn, W. F., Wiegmann, D. A., Arnold, J. J., Brown, D. A., et al. (2007). The mayo high performance teamwork scale: reliability and validity for evaluating key crew resource management skills. *Simul. Healthc.* 2, 4–10. doi: 10.1097/SIH.0b013e31802b68ee
- Mayer, R. E. (2002). Rote versus meaningful learning. *Theory Pract.* 41, 226–232. doi: 10.1207/s15430421tip4104\_4
- McLaughlin, J. E., Angelo, T. A., and White, P. J. (2023). Validating criteria for identifying core concepts using many-facet rasch measurement. *Front. Educ.* 8:1150781. doi: 10.3389/feduc.2023.1150781
- Möller, K., and Steffensky, M. (2010). "Naturwissenschaftliches Lernen im Unterricht mit 4- bis 8-jährigen Kindern. Kompetenzbereiche frühen naturwissenschaftlichen Lernens." in *Didaktik für die ersten Bildungsjahre. Unterricht mit 4- bis 8-jährigen Kindern*. ed. M. Leuchter (Seelze: Friedrich Verlag), 163–178.
- Nachreiner, K., Spangler, M., and Neuhaus, B. J. (2015). Begründung eines an Basiskonzepten orientierten Unterrichts [justification of oriented basic concepts teaching]. [teaching based on biological core ideas. A theoretical foundation.]. *Der mathematische und naturwissenschaftliche Unterricht [The Math and Science Education]*, 68, 172–177.

- Nahdi, D. S., and Jatisunda, M. G. (2020). "Conceptual understanding and procedural knowledge: a case study on learning mathematics of fractional material in elementary school" in *Journal of physics: Conference series*, vol. 1477. (IOP Publishing), 042037.
- NGSS Lead States (2013). *Next generation science standards: For states, by states*. National Academies Press.
- Rasch, G. (1960). Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. Nielsen & Lydiche.
- Reiber, A., Schweitzer, K., Klepser, R., Granström, M., Weitzel, H., and Bender, S. (2019). "Chilean and German primary pupils' conceptions of the musculoskeletal system – a comparison" in *Diversity in qualitative research. Qualitative psychology Nexus*. ed. G. H. In, vol. 15 (Tübingen: Center for Qualitative Psychology e.V), 13–30.
- Reiser, M., Binder, M., and Weitzel, H. (2024). Development and validation of a measure instrument for the understanding of structural-functional correlations of the locomotor system (MUSCLS). *J. Biol. Educ.* 1–20, 1–20. doi: 10.1080/00219266.2024.2351380
- Samarapungavan, A., Mantzicopoulos, P., and Patrick, H. (2008). Learning science through inquiry in kindergarten. *Sci. Educ.* 92, 868–908. doi: 10.1002/sce.20275
- Samarapungavan, A., Mantzicopoulos, P., Patrick, H., and French, B. (2009). The development and validation of the science learning assessment (SLA): a measure of kindergarten science learning. *J. Advanced Acad.* 20, 502–535. doi: 10.1177/1932202X0902000306
- Samarapungavan, A., Patrick, H., and Mantzicopoulos, P. (2011). What kindergarten students learn in inquiry-based science classrooms. *Cogn. Instr.* 29, 416–470. doi: 10.1080/07370008.2011.608027
- Sodian, B., Zaitchik, D., and Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Dev.* 62, 753–766. doi: 10.2307/1131175
- Staatsinstitut für Frühpädagogik München (2024). *Der Bayerische Bildungs- und Erziehungsplan für Kinder in Tageseinrichtungen bis zur Einschulung*. Weinheim: Beltz.
- Steffensky, M. (2017). *Naturwissenschaftliche Bildung in Kindertageseinrichtungen. Eine Expertise der Weiterbildungsinitiative Frühpädagogische Fachkräfte (WiFF)*. München: Deutsches Jugendinstitut. Available at: [https://www.weiterbildungsinitiative.de/fileadmin/Redaktion/Publikationen/old\\_uploads/media/WEB\\_Exp\\_48\\_Steffensky.pdf](https://www.weiterbildungsinitiative.de/fileadmin/Redaktion/Publikationen/old_uploads/media/WEB_Exp_48_Steffensky.pdf)
- Tolmie, A. K., Ghazali, Z., and Morris, S. (2016). Children's science learning: a core skills approach. *Br. J. Educ. Psychol.* 86, 481–497. doi: 10.1111/bjep.12119
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *Int. J. Sci. Educ.* 10, 159–169. doi: 10.1080/0950069880100204
- Treagust, D. F., and Mann, M. (1998). A pencil and paper instrument to diagnose students' conceptions of breathing, gas exchange and respiration. *Aust. Sci. Teach. J.* 44, 55–59.
- Van Boxtel, C., Van der Linden, J., and Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learn. Instr.* 10, 311–330. doi: 10.1016/S0959-4752(00)00002-5
- Vasseleu, E., Neilsen-Hewett, C., Ehrich, J., Cliff, K., and Howard, S. J. (2021). Educator beliefs around supporting early self-regulation: development and evaluation of the self-regulation knowledge, attitudes and self-efficacy scale. *Front. Educ.* 6:621320. doi: 10.3389/feduc.2021.621320
- Vygotsky, L. S. (1978). "Interaction between learning and development" in *Readings on the development of children* (New York: Scientific American Book), 34–40.
- Westerberg, L. (2024). *Developing and evaluating an assessment of preschoolers' science and engineering knowledge*. Doctoral dissertation, Purdue University
- Wright, B. D., and Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measure. Transact.* 8, 370–371.
- Wright, B. D., and Masters, G. N. (1982). *Rating scale analysis*. MESA Press.
- Wright, B. D., and Stone, M. H. (1979). *Best test design*. MESA Press. Available at: <https://research.acer.edu.au/cgi/viewcontent.cgi?article=1000&context=measurement>