Concreteness and Abstraction in Mathematics Education: A Taxonomy of the Semantic Landscape

Julia Chatain [1,2*], Charlotte Müller [2,3*], Keny Chatain [4], Leon Calabrese [5], Manu Kapur [2]

[1*]Dept. of Computer Science, ETH Zurich, Zurich, Switzerland.
[2]Dept. of Humanities, Social and Political Sciences, ETH Zurich, Zurich, Switzerland.
[3]Collegium Helveticum, ETH Zurich, Zurich, Switzerland.
[4]Institut Jean Nicod, École Normale Supérieure, Paris, France.
[5]Faculty of Law, University of Basel, Basel, Switzerland.

*Corresponding author(s). E-mail(s): jchatain@ethz.ch; chmueller@ethz.ch;

Abstract

Concreteness and abstraction are key in educational research in mathematics, both when discussing the nature of mathematics in itself, but also when exploring how to learn and teach mathematics. However, while the terms *concrete* and *abstract* are often used in the field, they are not always used with the same meaning. For example, the word *concrete* can be used to mean *specific*, *relatable*, *visual*, *tangible*, while the word *abstract* can be used as *general*, *rigorous*, *vague*, *symbolic*. While several scholars have emphasized the need for a multidimensional and fine-grained framework to articulate these differences in meaning, we further argue that it is crucial to precisely define how the terms *concrete* and *abstract* are actually used by the research community. Towards this goal, we offer three contributions. First, we empirically and systematically identify the various meanings of *concrete* and *abstract* used in the literature. Second, we offer a data-informed taxonomy to organize this semantic landscape and support research inquiry. And third, we offer templates for the design of future mathematics education interventions and studies.

Keywords: mathematics education, concreteness, abstraction, taxonomy, verbal dispute

1 Introduction

Concreteness and abstraction are central concepts in mathematics. First, when considering the nature of mathematics, some researchers and philosophers argued that mathematics is abstract, such as proponents of the formalism perspective (Ernest, 1985). In contrast, others argued that it is concrete, such as proponents of the embodied perspective (Lakoff and Núñez, 2000). Finally, others argued that mathematics is at the interface between concreteness and abstraction (Courant and Robbins, 1996). Second, when considering how to teach and learn mathematics, scholars argued that abstract representations promote better transfer (Kaminski, Sloutsky, and Heckler, 2008), but also that well-designed concrete examples may actually perform as well (Trninic, Kapur, and Sinha, 2020), or that a progression from concrete to abstract representations, e.g. *concreteness fading*, should be favored (Bruner, 1974; Fyfe, McNeil, Son, and Goldstone, 2014).

Through these few examples, we can already understand how central the terms *concrete* and *abstract* are to mathematics and mathematics education. It thus follows that the meaning of these terms needs to be unambiguous in order for scholars to be able to discuss their role in doing and learning mathematics. For example, when scholars debate on whether mathematics is concrete or abstract, it is sometimes unclear whether they actually debate about the nature of mathematics, or about the *meaning* of the terms concrete and abstract. Moreover, the terms *concrete* and *abstract* may not accurately capture all the nuances that scholars are investigating in mathematics education, but may be simplifications that, while being practical, may also create confusion.

These issues have been raised before, as previous work attempted to offer definitions (Wilensky, 1991; Löhr, 2023) for these key terms, and identified that the relationship between *concrete* and *abstract* should not be viewed as a dichotomy, but rather as a spectrum (Fyfe and Nathan, 2019). Furthermore, scholars even argued that an object can be both abstract and concrete (Coles and Sinclair, 2019; Wilensky, 1991): for example, while a mathematical symbol may be *abstract* because it is *formal*, it may also be *concrete* to a mathematician because it is *familiar*. Overall, previous work advocated for a multidimensional framework to define abstraction and concreteness, encompassing the various meanings of the terms (Trninic et al., 2020; Chatain, Varga, Fayolle, Kapur, and Sumner, 2023; Belenky and Schalk, 2014; Fyfe and Nathan, 2019; Petersen and McNeil, 2013), and even offered initial suggestions of such frameworks (Belenky and Schalk, 2014; Fyfe and Nathan, 2019). However, to this date, there is no comprehensive framework accounting for *all* the meanings of *concrete* and *abstract* that are effectively used by scholars in the field of mathematics education.

In this work, we offer a novel approach to the question by following a data-driven approach to map the semantic landscape of concreteness and abstraction in mathematics education. As a community, reflecting on the key terms we are using and defining them precisely can support scientific inquiry and even meaningfully advance the field (Chalmers, 2011; Vermeulen, 2018; Balcerak Jackson, 2014; Hirsch, 2005; Jenkins, 2014).

To build towards this, we aim to achieve two goals with this work:

G1 Identify the different meanings of the terms *abstract* and *concrete* in the field of mathematics education. Specifically, we want to answer: What do scholars mean when they use the terms *abstract* and *concrete*?
G2 Define a data-driven taxonomy that organizes the various meanings of *abstract* and *concrete* in the field of mathematics education.

Specifically, our goal is not to converge towards one best definition of concreteness or abstraction, but rather to support fine-grained scientific inquiry through a taxonomy of the semantic landscape acknowledging and organizing the various meanings of the terms used by scholars, and their relevance for mathematics education research.

We offer the following contributions:

C1 Empirical evidence of the various meanings of the terms *abstract* and *concrete* used in mathematics education;
C2 A taxonomy of the semantic landscape of *abstract* and *concrete*, informed by both theory and data, and supporting scientific inquiry;
C3 Templates to reflect on and classify learning interventions along different axes of concreteness and abstraction.

In the future, we hope that this work will help reflect on and organize past research, for example seemingly conflicting empirical results (Kaminski et al., 2008; De Bock, Deprez, Van Dooren, Roelens, and Verschaffel, 2011; Trninic et al., 2020; Burns and Hamm, 2011; Shurr, Bouck, Bassette, and Park, 2021), support in depth and nuanced scientific inquiry, and generate critical conversations and questions to further advance the field. In addition, we hope that our framework will improve qualification and design of future studies and interventions.

2 Related Work

Philosophers of mathematics used the words *abstract* and *concrete* extensively to describe the nature of mathematics. However, instructors and educational researchers used the same terms to describe artifacts and methods used to teach and learn mathematics. In this section, we detail various examples to highlight not only how abstraction and concreteness are crucial in mathematics practice and education, but also the diversity of usage. As such, we provide preliminary evidence of the need to map out the semantic landscape of concreteness and abstraction in mathematics education.

2.1 Concreteness and abstraction in mathematics education

The notions of concreteness and abstraction are central to mathematics education, at several levels: 1) when considering the nature of mathematics and 2) when considering how to learn and teach mathematics. In the following, we provide an overview of both accounts.

2.1.1 The nature of mathematics

In this section, our goal is not to state whether mathematics is abstract or concrete in nature, but rather to show that both terms have been used to describe mathematics, in various ways.

Power of generalization

Often, mathematics is described as an abstract discipline, concerned with the identification of powerful generalizations (Dreyfus, 2020): "There is broad agreement that the essential characteristics of mathematical knowledge are its generality and abstractness" (Bishop, Mellin-Olsen, & van Dormolen, 1991, p. 61), and "generality can be constructed through abstraction of the essential invariants in the context of a system of actions" (Dörfler, 1991, p. 73). Abstraction is often considered in a vertical paradigm, where concrete objects, often real-world, contextualized objects are considered low level, and abstraction "peels away" superficial details towards higher level ideal objects, which capture the essence of a concept (DiSessa, 2018, p. 21). In this paradigm, scholars speak of horizontal transfer, i.e. transfer within the same level of abstraction, and vertical transfer, towards a more ideal or a more contextualized representation (Bossard, Kermarrec, Buche, and Tisseau, 2008). However, it has also been argued that abstraction is rather an "imperfect and somewhat unstable translation between symbolic systems", following a horizontal paradigm (Wagner, 2019, p. 1). In this paradigm, abstracting does not mean creating and identifying ideal objects but rather building translations and transitions between concrete objects. Artigue (2009) suggests that this has implications for education as a an approach following a horizontal paradigm may support cognitive flexibility.

Symbol systems and formalisms

The formalism view of mathematics states that mathematics is a "meaningless game played with marks on paper, following rules" (Ernest, 1985, p. 606). Generally, mathematics is effectively communicated and manipulated using symbol systems and rules that are inherently abstract and meaningless if not interpreted (Harnad, 1990; Penrose, 1991). While these symbol systems may be considered as the essence of mathematics, they may also be viewed as representations only: shapes and forms that could be redesigned to directly embed meaning, for example leveraging more iconic representations (Harnad, 1990).

Connection to the real world

It is quite common for novices to believe that mathematics is an abstract subject in nature, disconnected from the real world, and, as such, difficult to learn (Schoenfeld, 1992). In contrast, Dossey (1992) argued that the nature of mathematics is strongly connected to the real and concrete world: "Aristotle's view of mathematics was not based on a theory of an external, independent, unobservable body of knowledge. Rather, it was based on experienced reality, where knowledge is obtained from experimentation, observation, and abstraction" (p. 40). Similarly, Lakoff and Núñez (2000) suggested that mathematics is embodied, i.e. the result of our sensorimotor experiences with the world. Finally, Courant and Robbins (1996) viewed mathematics as neither abstract nor concrete, but rather as a connection between the two realms: “Mathematics hovers uneasily between the real and the not-real; its meaning does not reside in formal abstractions, but neither is it tangible. [...] Mathematics links the abstract world of mental concepts to the real world of physical things without being located completely in either” (p. 0).

Rigid or malleable

Related to the question of the connection with the real world lies the question of the process of developing mathematics as a discipline: “Some see mathematics as a static discipline developed abstractly. Others see mathematics as a dynamic discipline, constantly changing as a result of new discoveries from experimentation and application (Crosswhite et al., 1986)” (Dossey, 1992, p. 39). Here, the abstractness of the discipline is connected to its rigidity, and is opposed to a certain malleability tied to the process.

A human and humane discipline

Another way of looking at mathematics is falliblism (Ernest, 1985), which considers mathematics as “what mathematicians do and have done” (p. 608). Through this lens, mathematics includes imperfections tied to all human creation (Ernest, 1985; Lakatos, 1976; Davis, Hersh, and Rota, 1981). Moreover, falliblism emphasizes the “immense practical usefulness of the subject” (Ernest, 1985, p. 609). This view is shared by others who consider mathematics to be a concrete discipline, which is anchored in our lived experiences: “Mathematics is a product of the neural capacities of our brains, the nature of our bodies, our evolution, our environment, and our long social and cultural history” (Lakoff & Núñez, 2000, p. 9). From this perspective, mathematics is very concrete, in the sense that it is relatable and connected to real world experiences. Importantly, according to Ernest (1985), falliblism is the only perspective that supports a humane approach to the discipline. Indeed, although mathematics used to be considered elitist, there has been a push for equity in mathematics, acknowledging the importance of considering mathematicians of different genders and backgrounds (Zevenbergen, 2001). The falliblism perspective thus supports, by definition, diversity in mathematics. Instead of focusing on abstract ideals, falliblism highlights mathematics as a concrete discipline, aggregating diverse perspectives and including imperfections tied to real-world experiences and human creations.

Affect and aesthetics

While some might feel indifferent towards mathematics and find it a rather cold and abstract topic, others may build strong emotional connections towards it: “Mathematics as an expression of the human mind reflects [...] the desire for aesthetic perfection” (Courant & Robbins, 1996, p. 0), or again “Mathematics is one of the most profound and beautiful endeavors of the imagination that human beings have ever engaged in” (Lakoff & Núñez, 2000, p. 5). In particular, aesthetics were often associated with concrete experiences offered to students learning mathematics. These experiences targeted their subjective sensitivity, and were often related to perceptual richness (Sinclair, 2001). However, scholars also considered how aesthetics and beauty can be tied to abstraction, for example, in proofs and theorems using symbolic and formal representations (Sinclair, 2001; Sinclair, 2011).

Overall, there are many views and philosophies of mathematics (Ernest, 1985; Wilkinson, 2021), and conflicting perspectives on the abstract and concrete nature of mathematics.

2.1.2 Teaching and learning mathematics

Beyond the question of the nature of mathematics lies the question of how to teach and learn mathematics. Generally, the question of the role of concreteness and abstraction in learning mathematics is a crucial one, underlying several inquiries such as: Can concrete examples impair learners' ability to abstract? At what point should educators introduce concrete representations when teaching mathematical concepts? Are abstract representations preventing learners from relating to the materials?

In this section, our goal is not to show whether abstraction or concreteness is best to support sense-making of mathematics, but rather show that, once again, both terms have been used in various ways in the field.

Conceptual metaphors to map concrete to abstract

From a cognitive perspective, proponents of the conceptual metaphor theory argued that any imperceptible concept must be understood in terms of an experiential concept (Lakoff and Núñez, 2000; Kövecses, 2010) or - importantly for the present review - abstract concepts must be understood in terms of concrete concepts. In short, as humans gain their first knowledge from how they exist and interact with the world, they build on this knowledge by relating new information to these initial experiences called image schemas. In mathematics, one of the easiest examples might be the conceptual metaphor *Categories are Containers* (Lakoff and Núñez, 2000). Understanding categories is not trivial at all, but it helps to conceptualize them as containers that are a bounded regions in space (the category) containing objects (category members). Principles such as the excluded middle, the modus ponens, the hypothetical syllogism or the modus tollens are then simply metaphorical applications of a spatial logic present in the Container schema (Lakoff and Núñez, 2000).

One concrete or abstract representation

Several studies were conducted to explore the impact of the use of concrete and abstract representations in mathematics education. Kaminski et al. (2008) showed that introducing concrete examples representing mathematical objects with *real-world* items such as tennis balls, pizza slices, or cups, may impair students' performance on abstract transfer tasks using more abstract representations. However, follow up studies showed that slightly adjusting the concrete examples to make them more *relatable* (Trninic et al., 2020) or adapting the posttest to include transfer tasks, which align better with the learning task (De Bock et al., 2011), canceled this effect, meaning that the effectiveness of concrete representations may lie more in how relatable they are rather than in the sole fact that they exist in the real world.

Multiple concrete or abstract representations

The use of several representations, either abstract or concrete, was also explored. For example, Schalk, Saalbach, and Stern (2016) showed that introducing *general* formalisms (i.e. abstract) instead of using solely *specific* cases (i.e. concrete) resulted in worse transfer. In this context, Rau (2017) discussed the necessary conditions for the effectiveness of multiple visual representations. She defined the representational dilemma: Students must make sense of both the representations and the concepts they represent (Rau, 2017; Ainsworth, 2006; McElhaney, Chang, Chiu, and Linn, 2015). Making sense of a visual representation means understanding the mapping between this representation and the concept it represents. The success of this process, therefore, lies in the quality of this mapping, but also in the interaction between the learner and the representation, including affective mechanisms and prior knowledge relevance.

Concreteness fading

Concreteness fading is a pedagogical approach leveraging several representations organized in a concrete to abstract progression, originally recommended by Bruner (1974). Specifically, the concreteness fading design suggests that "concepts and procedures should be presented in three progressive forms: (1) an enactive form, which is a physical, concrete model of the concept; (2) an iconic form, which is a graphic or pictorial model; and finally (3) a symbolic form, which is an abstract model of the concept" (Fyfe et al., 2014, p. 11). This approach was extensively explored in the literature, exploiting various representations and sequencing (Suh, Lee, and Law, 2020). Recently, the generalizability of concreteness fading to other sciences was questioned as what is an abstract and concrete representation differs greatly between STEM domains (Kokkonen and Schalk, 2021). We suggest that already within mathematics education, there is variety in meaning of the terms.

Learning as a process of concretion

Rather than at the representation level only, concreteness and abstraction was further mentioned as part of the process of learning. Papert (1980) related learning to the process of changing something "otherwise abstract", like equations, into a "comfortable friend", like gears (p. xviiiff). More generally, Dewey (1910) argued that understanding is changing something abstract, i.e. "strange and perplexing", into something concrete i.e. "plain and familiar" (p. 118).

As concreteness and abstraction are central to mathematics education, it is crucial to understand what these terms mean. However, while all these works deal with these concepts, they often mean different things, sometimes even resulting in different learning outcomes (Kaminski et al., 2008; De Bock et al., 2011; Trninic et al., 2020). In the following, we present why mapping the semantic landscape of concreteness and abstraction in mathematics education would offer an opportunity for growth in the field.

2.2 Mapping the semantic landscape

Past work has highlighted the importance of discussing and improving the definitions of the terms *concrete* and *abstract*. Several such definitions have been suggested, for example centering on physicality or familiarity (Wilensky, 1991; Löhr, 2023). However, scholars further argued that the terms *concrete* and *abstract* describe a spectrum, rather than two disjoint categories (Fyfe and Nathan, 2019), and could even cover a finer-grained multidimensional space rather than an unidimensional definition (Trninic et al., 2020; Chatain et al., 2023; Belenky and Schalk, 2014; Fyfe and Nathan, 2019; Petersen and McNeil, 2013). Indeed, Trninic et al. (2020) mentioned the need to "replac[e] the abstract–concrete dichotomy with a more comprehensive framework" (p. 13). Similarly, Chatain et al. (2023) argued that there is a "need for a more rigorous definition of 'abstraction' and 'concreteness' in the field of mathematics education" (p. 10) as, in their study, different forms of concreteness resulted in different learning outcomes and overall motivation. Similarly, Coles and Sinclair (2019) wrote that some objects can simultaneously be categorized as both abstract and concrete: "While the alphanumeric inputs of programming languages such as Logo may seem abstract, being symbolic as they are, they can also be seen as concrete for some children inasmuch as some children will have had encounters with these symbols that give them a direct and visible reference (Wilensky, 1991)" (Coles & Sinclair, 2019, p. 467).

In an attempt to address this need for a multidimensional definition, scholars formulated conceptual frameworks, to which we compare our approach and resulting taxonomy in Section 5.3 (Belenky and Schalk, 2014; Fyfe and Nathan, 2019). In the following, we argue for the relevance of our approach, which is systematic and data-driven. Specifically, our goal was to build a taxonomy of the semantic landscape of concreteness and abstraction that is **comprehensive** and **informed by actual usage** of the terms in the literature.

Generally, the importance of clarifying and detailing the meanings of key terms within a scientific domain has been extensively discussed in philosophy of language. Indeed, Chalmers (2011) has elaborated the concept of verbal disputes: disagreements that arise partly or wholly in virtue of a disagreement about the meaning of certain words. The parties involved are often unaware that their disagreement is rooted in these differences in word meanings. Different kinds of verbal disputes have been discussed in the literature, such as "pragmatic verbalness", i.e. "talking past each other", or "doxastic verbalness", i.e. "not really disagreeing" (Becker, 2022). A closely related concept is the concept of terminological inconsistency, in which the same term is unknowingly used with different meanings or connotations by the same or different individuals. While there is, to our knowledge, no explicit discussion of the concept of terminological inconsistencies in the philosophy of language literature, it seems plausible that such inconsistencies, like verbal disputes (Vermeulen, 2018), can be resolved by clarifying usage. Either way, such disputes and inconsistencies can constitute obstacles to understanding and scientific inquiry, because they distract from more substantive issues, e.g. in the present case, the mechanisms explaining ease or difficulty of learning.

Our paper is driven by the definition of what Chalmers (2011) calls a "*broadly* verbal dispute":

> "A dispute over S is (broadly) verbal when, for some expression T in S, the parties disagree about the meaning of T, and the dispute over S arises wholly in virtue of this disagreement regarding T."

In our context, S could be a statement such as "math is difficult to learn because it is abstract", where T is "abstract." A party believing that "abstract" stands for "generalizable" might disagree with S, while a party believing that "abstract" means "unfamiliar" might agree with S. Similarly, if S is "concrete representations improve learning outcomes in mathematics" and T is "concrete", parties considering the meaning of "concrete" to be "relatable" or "informal" might disagree about S. However, when discussing the terms more precisely, and engaging in shared reflection, both parties would realize that (1) they do not actually disagree, and (2) there is a need to better describe the meanings of the terms. Again, the latter is already the case as many scholars in the field argued for and have suggested multidimensional frameworks. In this paper, and as illustrated in Section 5.2.1, we ought to support and guide this resolution process by mapping out the semantic landscape of the terms "concrete" and "abstract".

Indeed, because verbal disputes and terminological inconsistencies turn on some disagreement regarding language, they can be resolved when all parties to the dispute agree to use certain words in specific and more precise ways. By systematically spelling out different ways in which the words *abstract* and *concrete* are used and understood in the literature, we hope to dispel any verbal dispute and prevent terminological inconsistencies connected to the use of these polysemous terms in educational research. Furthermore, we believe that our overview of the semantic landscape can serve further scientific inquiry by enabling identification of trends, gaps, and critical questions, but also by helping to organize and compare both past work and future work.

3 Method

This paper aims to provide an overview of how educational researchers conceptualize the terms *abstract* and *concrete* as well as to generate a taxonomy that structures our findings in a way that is useful to researchers, for scientific inquiry, but also when designing learning activities that aim to concretize in any form. We chose a qualitative approach that can be divided into three main steps (Figure 1):

- Data preparation: The creation of the corpus including choosing an appropriate database, designing the search string, and removing duplicates, books and theses.
- Data coding: The identification and coding of words associated with/used synonymous with *abstract*/*concrete* in the individual papers.
- Taxonomy definition: The organization of the generated codes associated with concreteness and abstraction into a taxonomy.

We hold an experientialist perspective, meaning that to us, "language is a tool for communicating meaning" (Clarke & Braun, 2021, p. 164). Importantly, we assume that the language used by the authors of the papers in our corpus, in conjunction with the words *abstract* or *concrete*, provides information about how the authors conceptualized these terms.

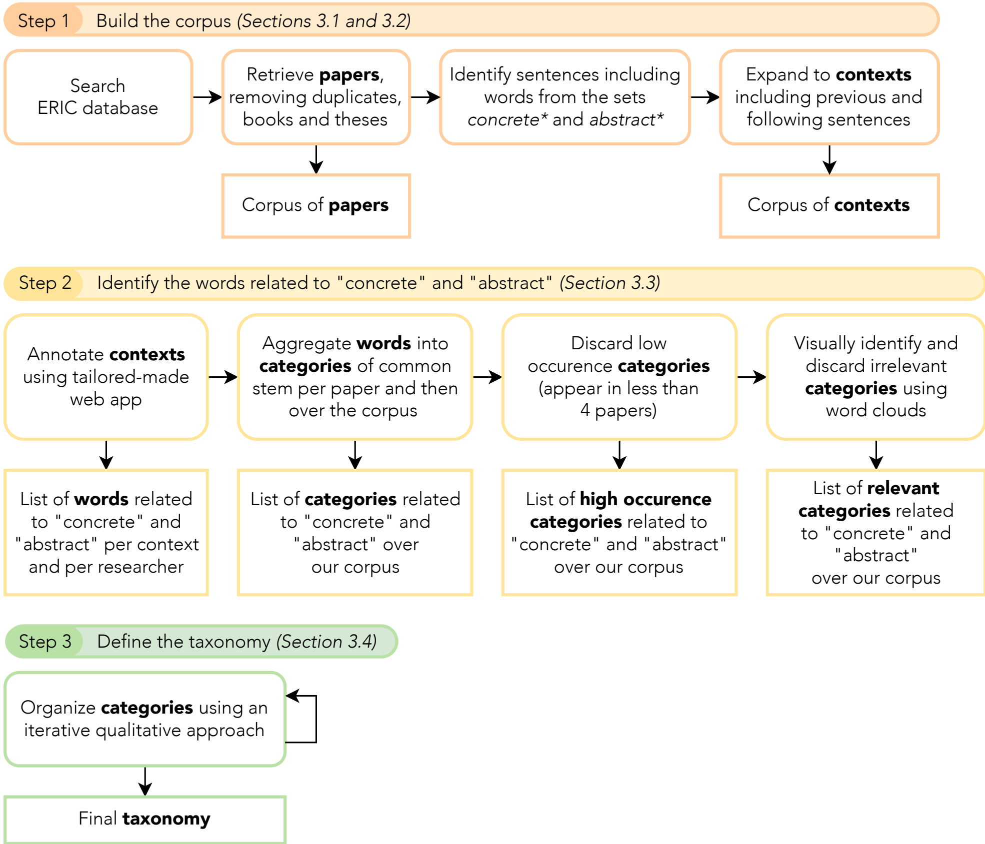

Fig. 1 Overview of our process from data collection to taxonomy definition.

Further, we hold a realist perspective insofar as the researchers generating this taxonomy are not removed from the process but are active participants. Therefore, we acknowledge the individual backgrounds of the four involved researchers (Muma, 1991): As part of the initial process of personal reflexivity, we report their profiles in Table 1, as well as their perspective statements in Table 2 (Clarke and Braun, 2021). Importantly, R1 and R2 offer an external perspective as they have little to no experience in the field of mathematics education, while R3 and R4 offer a more informed perspective.

We aimed for a high degree of trustworthiness as defined by Lincoln and Guba (Lincoln and Guba, 1985; Stahl and King, 2020). According to them, trustworthiness involves four dimensions: credibility, dependability, transferability, and confirmability. First, we increased *credibility* by triangulating coder perspectives. Every data point was analyzed twice. The perspectives were diverse in the sense that two coders were experts in the field and two coders provided outsider perspectives (see Table 2). Second, we increased *dependability* by documenting and continuously discussing the analysis process. Specifically, in the second iteration of the analysis process, R3 and R4

Table 1 Profiles of the researchers involved in the data collection, coding, and analysis process at the time of the coding phase.

Id	Gender	Highest degree achieved	Collection	Coding	Analysis
R1	M	Research assistant	Yes	Yes	No
R2	M	Post-doctoral researcher	No	Yes	No
R3	F	Doctoral researcher	No	Yes	Yes
R4	F	Doctoral researcher	No	Yes	Yes

Table 2 Perspective statements for researchers involved in the data coding

Id	Perspective statement
R1	R1 has a Bachelors degree in Psychology and Political Science and has assisted in the coding of multiple projects. He has little to no experience in the field of concreteness and abstraction in mathematics education.
R2	R2 is an expert in linguistics, specifically semantics, with little to no experience in the field of concreteness and abstraction in mathematics education, and an interest in analytical philosophy.
R3	R3 is an expert in embodied mathematics and embodied interaction for mathematics learning, with experience in the field of concreteness in mathematics education, and a positive bias towards a horizontal paradigm (i.e. relational) of abstraction in mathematics.
R4	R4 is an expert in embodied (haptic) (quantum) chemistry learning, with experience in the field of concreteness in a chemical context (i.e. perceivable) and previous encounters with concreteness in mathematics education as applied in quantum chemistry.

organized the codes independently of each other before coming together and forming the third iteration of the taxonomy. Third, we increased *transferability* by contrasting the proposed taxonomy with other established frameworks. Finally, we increased *confirmability* by offering a description of the process in Appendix A as well as further examples of data excerpts in the supporting material. That being said, we do not claim objectivity, as we offer a specific perspective.

3.1 Data preparation

First, R1 assembled a corpus of papers focusing on concreteness and abstraction in mathematics education. To do so, we focused on the ERIC database[1] due to our interest in research situated in educational contexts, and on peer-reviewed journal articles only. We used the following search query:

Search string applied for the initial corpus on October 31, 2022: `abstract:((concrete OR concreteness OR concretize OR abstraction OR abstract) AND mathematics AND (High-school OR Higher Education) AND (Learning OR education OR teaching) -‘‘Machine Learning’’) OR title:((concrete OR concreteness OR concretize OR abstraction OR abstract) AND mathematics AND (High-school OR Higher Education) AND (Learning OR education OR teaching) -‘‘Machine Learning’’`

Search string applied after first revision round on February 07, 2025, to be added to the initial corpus: `abstract:( (concrete OR concreteness OR concretize`

[1] https://eric.ed.gov/, accessed in 2023

```
OR abstraction OR abstract) AND mathematics AND (primary OR secondary)
AND (Learning OR education OR teaching) AND pubyearmax:2022 -‘‘Machine
Learning’’ ) OR title:( (concrete OR concreteness OR concretize OR
abstraction OR abstract) AND mathematics AND (primary OR secondary)
AND (Learning OR education OR teaching) AND pubyearmax:2022 -‘‘Machine
Learning’’
```

We removed machine learning articles through the query as we are only interested in human learning in this work. Furthermore, we later excluded works that investigated “abstracts”, as in paper abstracts (see description of the data coding in Section 3.2). These were the only content-related exclusion criteria implemented.

From this search, we collected 441 references. We then cleaned our sample to remove duplicates and exclude books or theses, resulting in a total of 401 papers. From this sample, we could retrieve the content of 368 papers through open or institutional access, which constitutes our final sample (detailed in supplementary materials).

We then converted all documents to a text format in the following way: first, we used the Poppler library (Noonburg and Cid, 2005) to convert the PDF to a plain text format; we then removed the abstract and references from the output using regular expressions. Noticing that, in the output of these two steps, many words of interest were split across lines due to hyphenation, we performed a final manipulation: for any line ending in a hyphen, we tried combining the string before the hyphen with the beginning of the next line; if the result belonged to a dictionary of English, we removed the hyphen and joined the lines.

Using a regular expression based on the roots of the words *concrete* and *abstract*, we identified the morphologically related words present in our corpus. We summarize the set of the selected morphologically related words in Table 3, excluding “Abstract” as section title. In the rest of this section, we use *concrete** and *abstract** to describe these sets.

Table 3 Sets of selected words morphologically related to “concrete” and “abstract”.

Set name	Content
*concrete**	concreteness, concrete, concretise, concretize, concretises, concretizes, concretised, concretized, concretising, concretizing, concretely
*abstract**	abstraction, abstract, abstracts, abstracted, abstracting, abstractly

Using these sets, we identified the contexts of these words in the papers. Specifically, we looked for the words of *concrete** and *abstract** in the papers, and extracted the sentence containing the word, as well as the previous and following sentences to compose a *context*. If the previous or following sentence also contained a word from our sets, we extended the context to one more sentence in that direction until no word from our sets was found. This allowed us to have complete self-contained contexts and avoid redundancy within our dataset. Through this procedure, we collected 4309 contexts.

3.2 Data coding

Once we identified the relevant contexts in the papers, we needed to identify the words most likely to relate to those from *concrete** and *abstract**. To do so, we first tried an automated statistical approach to identify the words that co-occurred most often with abstraction and concreteness. Specifically, we gathered all words whose frequency of occurrence in contexts containing the words *abstract** and *concrete** was significantly higher than their frequency of occurrence in any context. Significance testing was done using Fischer's exact test (Fisher, 1970). This analysis tended to produce very low-frequency terms (such as *brandy*), probably due to violated assumptions of our test, word occurrences being not independent from each other.

As such, we decided to focus on a manual selection of these words. We first attempted to create a guide for the coding that would result in rater-independent codes. However, we found that the different perspectives of the coders were relevant for the coding process and that we are actually interested in the different codes that a coder may relate to the words of interest. Therefore, we formulated a task that allowed some room of interpretation to the coder. The task was formulated as follows:

"Identify the intended meaning of the words concrete and abstract in this text. Accordingly, select in `green` the words that are more related to concrete and in `purple` the words that are more related to abstract. If a word is not related to either, leave it blank."

To support the annotation process, we designed and implemented a browser application where researchers can review each context and highlight, by clicking or touching, the words related to *concrete** and *abstract** within one context (Figure 2). The contexts are presented in a random order, starting with the contexts with the lowest review count. R1, R2, R3 and R4 performed these annotations, until all contexts had two reviews and each researcher reviewed half of the contexts. This resulted in a workload of approximately 1860 contexts per researcher, an amount that allows us to confidently claim convergence of code generation. In addition, R3 and R4 reviewed the corpus added after the first revision round on February 07, 2025, resulting in additional 592 contexts per reviewer. It is important to note that this second analysis was performed after the taxonomy generation. However, we did not find any additional codes that did not fit our generated framework.

Our tool also allows to label a specific context as irrelevant. We used this option for contexts that could not be parsed properly, that should not have been parsed, or that are not in English. For example, some papers reviewed "abstracts" of papers.

Before starting data annotation, all four researchers discussed their understanding of the task, and applied it on several contexts from our corpus until reaching a common understanding. Again, this common understanding was not intended to result in equal codes, but rather in an equal approach to the task. Following our qualitative approach, the goal was not to resolve disagreements regarding the words selected by each researcher, but to reach a common understanding of the task and the intent (Clarke and Braun, 2021).

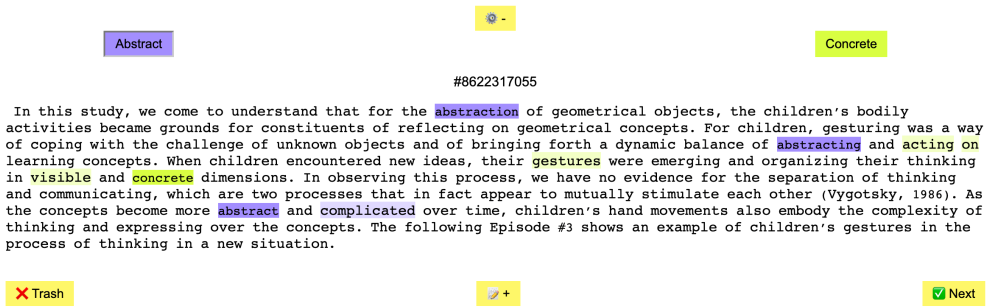

Fig. 2 Screenshot of the application used to select the words. The words from *concrete** and *abstract** are highlighted in opaque colors. The researcher can then highlight related words in transparent colors. To change the highlighting color, researchers can press the "Abstract" and "Concrete" buttons above the text. If the data is incorrect, researchers press the "Trash" button. Some notes can be added by pressing the notepad. Upon completion, researcher press the "Next" button. The task description can be toggled using the gear button. Finally, a unique identifier for the current context is displayed and can be used as reference for future discussions.

3.3 Pre-processing

We aggregated the annotations of all researchers per paper. For each paper, we aggregated the annotations per common stem using the NLTK English stemmer (Bird, Klein, and Loper, 2009). We then used the most common word of each category as the representative. This means that even if a word, or words with a common stem, appeared several times within one paper, it is only counted once. We then aggregated annotations over our entire dataset, using the same process: aggregate per common stem, and use the most common word as a representative. In this step, we assigned the count of the whole category to the representative. For example, the category $(virtually, 5), (virtual, 7), (virtuality, 1)$ would become $(virtual, 13)$.

The dataset was then refined by visually inspecting the word clouds produced by the individual researchers as well as the total word cloud (provided in the Appendix A). From this, we defined outliers such as "algebra" or "mathematics" that were found in many contexts but that did not contribute a special meaning to the words abstract or concrete. Finally, we decided on a cut-off point of minimal occurrence for a word to be added to the final dataset by plotting the corpus that would remain if we only considered words that occurred a certain number of times or more (see Figure 3). We selected a threshold of 4 occurrences minimum per word, as the corpus stabilized after 4 to 5.This means that while a large number of words were only coded in one to three papers and hence were not representative for the field, the remaining codes were found at higher frequency. As this method is ambiguous to a certain extent, we chose the more generous cut-off. We provide data examples per code as supplementary material.

3.4 Taxonomy definition

To organize the taxonomy, we followed an expert approach rather than an automated approach. Indeed, we explored automated approaches such as GloVe to evaluate semantic proximity (Pennington, Socher, and Manning, 2014), followed by a Principal

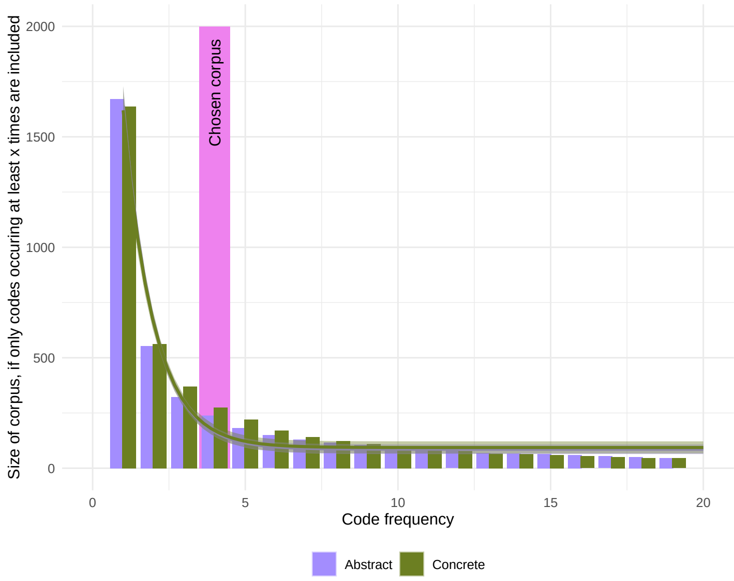

Fig. 3 Words which occurred less than four times were excluded from the analysis. One occurrence is defined as found in one paper. The curve is fitted exponentially.

Component Analysis (PCA) to organize the meanings into categories (Pedregosa et al., 2011). However, this approach is not suitable for our small dataset. Moreover, according to our multidimensionality hypothesis, certain words may be used with different meanings, preventing such analysis to provide productive results.

R3 and R4 built the final taxonomy based on expertise. To guide this process, we preliminarily discussed and identified the following goals: The resulting taxonomy should (1) accurately describe the semantic space covered by our data, (2) be organized to highlight the different meanings used in the literature, and (3) serve scientific inquiry, such as definition of future design and comparisons of studies.

In this regard, we adopted a qualitative approach, and performed a reflexive thematic analysis (Braun and Clarke, 2006; Clarke and Braun, 2021). Specifically, we followed (1) an inductive orientation to data, and acknowledged our personal bias in Table 2, (2) a semantic approach to meaning as per our task description, (3) a critical framework focused on the topics of concreteness and abstraction, and (4) a relativist theoretical framework as we acknowledge different realities as per our hypothesis of multiple meanings being used in the field (Clarke and Braun, 2021).

The detailed procedure is described in Appendix A. We started by grouping the codes into groups of related meaning. At this stage, these groups were quite broad, such as "kind of external representation" or "subjective characteristics". We discussed disagreements until both parties were convinced of the solution. Importantly, this process was not about one researcher being wrong and the other being right, but about finding a solution that works for both parties. During that process, we noticed that some of the codes seemed to describe the target concepts to be learned, some the representations of them, and some the relationship of the learner with this representation.

Therefore, we first tried to arrange our found groups in a triangle between the learner, the representation, and the concept. R3 and R4 performed this step independently (see Figures A5 to A8). During the discussion of conflicts, we found that we could not distinguish clearly between a certain code describing the relationship between the learner and the concept or the representation and the concept. This was also the source of most disagreements in the taxonomies created by R3 and R4 in this step. Specifically, we found that arguably the relationship between the learner and the concept will always be mediated by the representations of the concept known to the learner. We therefore modified this taxonomy into a linear model learner - representation - concept. Finally, we refined the groups into axes that would be sufficiently specific to be used in research and practise. We iteratively grouped codes until both R3 and R4 agreed every code was assigned a group. Moreover, we assigned all codes within an axis to either a *concrete* or an *abstract* ending of this axis. This is in alignment with previous work that stated that such terms are relative rather than dichotomous (Fyfe and Nathan, 2019). Therefore, we arranged the different meanings of *abstract* and *concrete* along the linear dimension learner - representation - concept and for every meaning, we present an axis between *abstract* and *concrete* supplemented with a design question to make the taxonomy easier to use. For two axes, we only found codes for one end of the spectrum. These axes are indicated by dotted lines in Figure 4.

4 Results

In this section, we describe our results, including the taxonomy, as well as a preliminary outlook on past research leveraging the taxonomy.

4.1 Taxonomy

In this section, we present our final taxonomy, depicted in Figure 4. For each axis, we also provide one example of codes comprised in our corpus. Moreover, we include prompting questions to help categorize representations along each axis. In addition, we complement this contribution with a blank template, in Appendix B (Figure B11), that can be used to organize studies or interventions according to this taxonomy.

As justified in Section 3.4, we describe the learning process as follows: To learn a concept ($\mathscr{C}$), a learner ($\mathscr{L}$) interacts with a representation ($\mathscr{R}$) of $\mathscr{C}$, designed through mapping. *abstract* $\leftrightarrow$ *concrete* axes are then organized at the interaction level, if they relate to properties of the representation ($\mathscr{R}$) as experienced by the learner ($\mathscr{L}$), or at the mapping level, if they relate to properties for how the representation ($\mathscr{R}$) is defined in relation to the concept ($\mathscr{C}$). Each axis is defined according to the representation ($\mathscr{R}$) as it is the only element that can be acted upon when designing interventions or studies. Indeed, while one cannot change the concept nor the learner, one can make design choices for a representation in relationship to the underlying concept or the target learner.

This distinction also enables the distinction between *subjective* and *objective* axes. Indeed, all axes at the interaction level depend on the representation and on the learners, and are therefore related to the subjective experience of the learners. While these axes can be designed for, thanks to previous empirical evidence and design recommendations, only actual measurement with target learners can ensure that the desired effect indeed happened. For example, one might decide to design a more engaging representation by creating a physical and interactive object. However, whether or not this representation increases *interest* depends on the learner. Similarly, while a designer might choose to design a highly *sensorial* representation, including visual and audio components, whether or not these cues will be perceived by the learner depends on the learners' sensory abilities but also on how and whether they choose to engage with the representation. In contrast, the axes at the mapping level only depend on the representation and the represented concept, and are therefore objective axes, solely depending on the designer's decisions. For example, the designer may decide to use real world examples (*situatedness*), or to design an experiential representation (*experientiality*) of the concept, and this does not depend on the learner.

In alignment with our goals, the axes of our taxonomy are organized such that the axes differ in meaning. However, they may relate to each other, as making certain design decisions along one axis may impact another one. For clarity, we discuss these relationships separately in Section 5.1. Finally, for consistency, we named each axis according to most common codes found at the concrete end of the spectrum.

Crucially, one representation may now be abstract along one dimension and concrete along another. For example, a simulation of an excavator is virtual (*physicality* → *abstract*) and situated in real life (*situatedness* → *concrete*). Similarly, the representation may be concrete for one learner and abstract for another. A mathematician will find equations familiar (*familiarity* → *concrete*) while for the high school student, the equations might seem unfamiliar (*familiarity* → *abstract*).

In the following, we describe the included axes and provide two exemplary data excerpts from our corpus for each axis.

4.1.1 Interaction level: Between the learner and the representation

At this level, we discuss axes that are related to the perception of the representation by the learner: *familiarity*, *interest*, *simplicity*, *relevance* and *sensoriality*. Specifically, to consider designing according to these axes, one must know who the targeted learners are as these axes refer to subjective experiences of the learners. We provide two data examples per axis. For more data examples per code, refer to the supplementary material.

Familiarity

Familiarity relates to how familiar or unfamiliar the representation is to the learner. The axis is based on cases in which we found that authors used concreteness in relation to a representation being meaningful, relatable, or accessible to the learner. Specifically, from their past experiences, how much can the learner relate to the chosen representation? Have they encountered similar representations before, in this context or another? Familiarity may either mean that the representation utilizes references that the learner is familiar with, independently of any mathematical endeavor, for example the use of pizzas to represent groups, or that the learner is familiar with

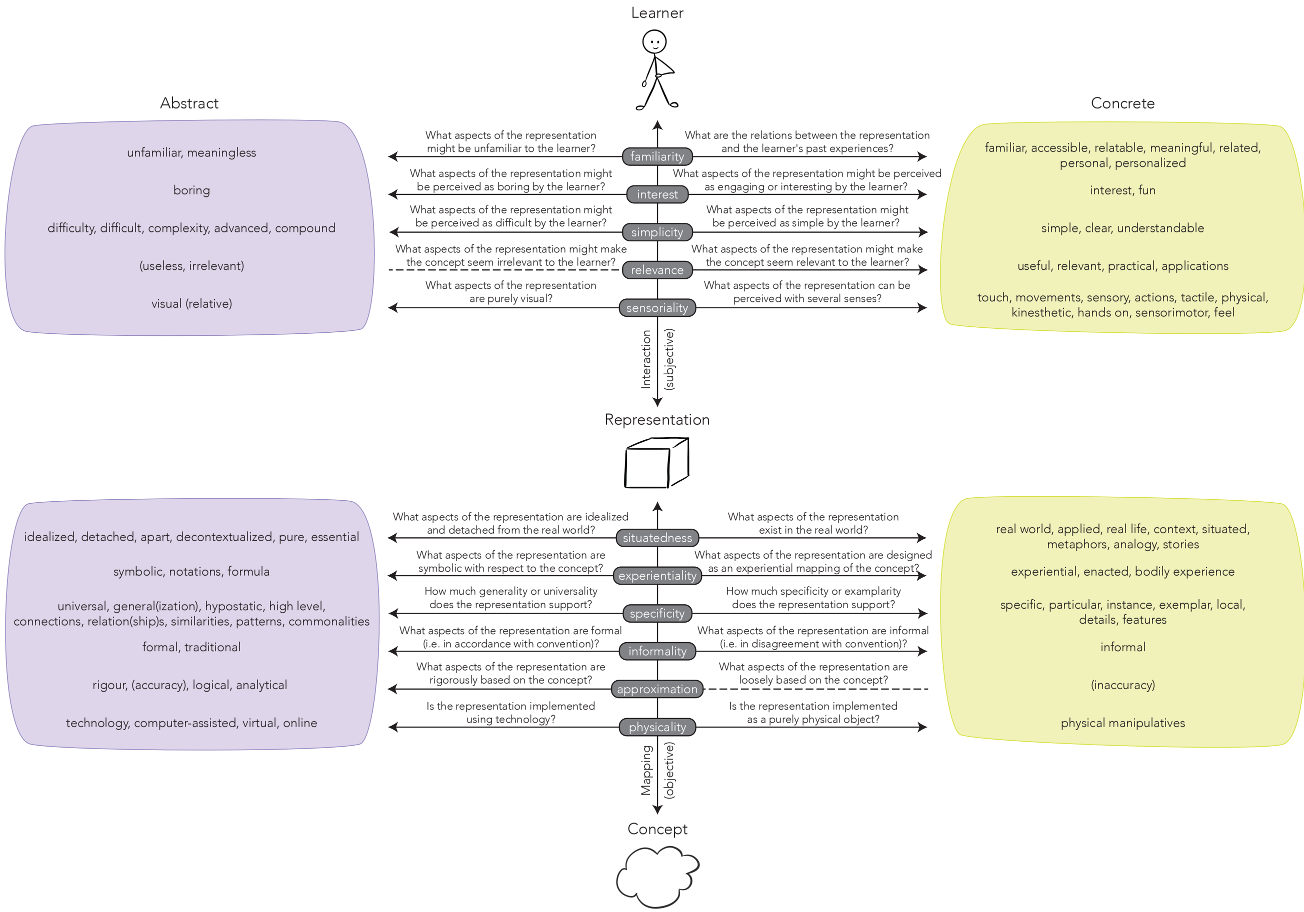

Fig. 4 The final taxonomy. Dotted lines or codes in brackets indicate that this end of the spectrum was not found as codes in the data but was added during the analysis process. The green end of the spectrum indicates a concrete end, purple an abstract end. Words in the purple and green bubbles are examples of codes from our corpus.

this specific representation and the way it maps to the underlying concept, for example, how students finishing high school are familiar with graphs as representations of functions and understand the relationship between the two.

For example, Rich and Yadav (2020) write "Familiar problems are generally considered to be at a lower level of abstraction than unfamiliar problems, [...]" (p. 196). Similarly, Altintas and Ilgün (2017) write "Maths would get more difficult and meaningless, Maths would get very abstract." (p. 963).

Interest

Interest relates to how interesting, entertaining or engaging, as opposed to boring, the learner may find the representation. For example, how aligned is the representation with the learner's areas of interest? How engaged is the learner when interacting with the representation? This axis can refer to both the initial reaction of the learners, that is, do they get interested in the representation and wish to interact with it and make sense of it, but also prolonged and repeated engagement, that is, do they interact with it for a long time and do they come back to it on their own.

Already now, we may contrast this axis against the familiarity axis. If a student is unfamiliar with a representation, but finds it interesting and wants to interact with it, this representation may be considered abstract or concrete, depending on what aspect one focuses on.

For example, Altintas (2018) writes "According to the students' views, when students are told stories about a set of abstract boring subjects and concepts, [...]" (p. 258). Woodhouse (2012) highlights the subjective aspect of this axes when he writes that "[a] familiar theme that runs throughout his work is the need to relate abstract ideas to the concrete experience and interests of students, so as to avoid their becoming inert" (p. 4).

Simplicity

Simplicity relates to how simple the representation is for the learner to understand, as opposed to how difficult or complex. The complexity could be intrinsic to the concept and thus translated to the representation, but it could also be extraneous and inherent to the representation itself. Importantly, here, complexity is defined as perceived by the learner, based on their abilities, but also on their estimation of their abilities (self-efficacy). As a designer, this might mean deciding to leave certain aspects of the concept out when designing the representation, as they are not relevant yet and might be too complex for the learners based on their current knowledge, but also to refrain from choosing representations that may make learners feel like they are not capable of solving the problem.

For example, Lang and Pagliaro (2007) write "[...] by defining them in simple understandable terms and by using concrete terms high in visual, pictorial or sensory connotations" (p. 457). Similarly, Ganesh and Middleton (2006) write "Multiple mathematical representations that are visual and symbolic allowing students to make the move from simple to more complex and abstract notions are recommended for understanding [...]" (p. 125).

Relevance

Relevance refers to how useful or practical the learner will believe the concept is when interacting with the representation. Here again, this axis focuses on the perception of the learner: a certain mathematical concept may be useful in the sense that it has a wide range of possible applications, but the representation may not be highlighting this sufficiently in relation to learners' past experiences. Does the chosen representation make learners feel like they can use this mathematical concept in their own lives, be it to solve other school problems, or out-of-school practical problems?

For example, Mitchelmore and White (1995) write "She wrote, 'Andrew understands area in the abstract but cannot yet apply it in practice' " (p. 55). Similarly, Ottemo, Berge, and Silfver (2020) write "However, in its 'personalized' rooting of knowledge and privileging of concrete and immediate usefulness of, for example, knowledge of engines, it does correspond to a relative emphasis of horizontal over vertical discourse" (p. 707).

Sensoriality

Sensoriality refers to how the representation can be accessed by the learner through their senses, a more concrete representation thus leverages more senses or the same sense with more stimulation. Importantly, this involves both design choices at the representation level, and how these choices are leveraged by learners, either due to their sensorimotor abilities or due to their interaction behavior. For example, one design choice would be to integrate audio feedback in the representations. However, this audio feedback would not be perceived by a learner with deafness or by a learner who never performs the action that triggers said audio feedback.

For example, Gravemeijer (2011) writes "In reflection, we may conclude that trying to make abstract mathematics concrete by representing the mathematics with tactile or visual models, is highly problematic" (p. 7). In contrast, Durmus and Karakirik (2006) write "Manipulative materials are concrete models that involve mathematical concepts, appealing to several senses including the socio-cultural needs that can be touched and moved around [. . .]" (p. 120).

4.1.2 Mapping level: Between the representation and the concept

Furthermore, there are six axes that describe the mapping between the concept and the representation: *situatedness* between idealized and real life, *experientiality* between symbolic and enacted, *specificity* between universal and specific, *informality* between formal and informal, *approximation* between rigorous and approximate, and *physicality* between virtual and physical. These axes are objective as they do not depend on the learners.

Situatedness

Situatedness refers to how much the representation situates the concept in a real-world context. For example, a representation of a mathematical graph as a water network would be more concrete (*situatedness* → *concrete*) as such networks exist in the real world (Chatain et al., 2023). However, representing the same graph as a set of circles (nodes) and lines (edges) would be more abstract as it is less context-specific (*situatedness* → *abstract*).

Already now, we may point out that while increased *situatedness* may result in higher *familiarity* as learners are more likely to be familiar with "real world" examples, these two axes should not be confused with one another. Indeed, some learners, e.g. high school students, may be familiar with non-situated representations, e.g. functions' graphs. Moreover, *situatedness* is an objective axis, i.e. depends only on how the underlying mathematical concept is mapped into the representation, while *familiarity* is subjective, i.e. depends on how learners perceive said representation.

For example, Bleazby (2015) writes "[. . .] our experience of the world is embodied, meaning it is subjective, concrete and situated" (p. 673). Similarly, Wolfmeyer (2018) writes "This means that mathematics abstracts the concrete real world objects into "ideal" shape and quantity" (p. 88).

Experientiality

Experientiality refers to the extent to which the representation is an experiential mapping of the concept as opposed to a symbolic mapping. Typically, symbols are arbitrary, essentially meaningless shapes that are only connected through the concepts they represent via grounding (Harnad, 1990). In contrast, an experiential mapping is concerned with ensuring that the representation is meaningful in and of itself, that is that it is based on universal human experiences of the world in a way that conveys meaning. Simply put, while experiential representations (*experientiality* → *concrete*) intrinsically embed meaning, symbolic ones (*experientiality* → *abstract*) need to be explained.

Already now, we should point out that the design of experiential representations is often done by using multimodal cues, especially in the case of embodied design (Abrahamson et al., 2020), and may thus often result in increased *sensoriality*. However, the *experientiality* axis is not concerned with how the learner experiences this embedding but how the meaning of a certain concept is embedded in the given representation. For example, while the representation may include tangible elements to articulate meaning, the learner may not (or may not be able to) touch these elements. Moreover, this embedding of meaning does not necessarily need to be physical, or accessible to the senses of a learner, but it may further be imaginative. Therefore, while the axes *physicality* and *sensoriality* may often align with this axis, they are distinct.

For example, Ottemo et al. (2020) write "[. . .] the teacher repeatedly emphasizes the value of experiential and contextually rooted knowledge over the abstract and the mathematically modeled" (p. 705). Coles (2014) highlights the transition from abstract to concrete when he writes "[. . .] all students were able to make the transition between a visual or enactive representation of a movement/relationship and its symbolic or formal description and go on to use this symbolism with confidence" (p. 29).

Specificity

Specificity relates to how much specificity or exemplarity, as opposed to generality or universality, the representation supports. This is important to consider at the mapping level as some representations may be able to translate this specificity or not: For example, using graph representations for functions will only enable the description of specific functions, while symbolic representations can account for specific functions ($f : x \to x^2$) but also more general forms ($f : \mathbb{R} \to \mathbb{R}$).

For example, Sezgin Memnun, Aydin, Özbilen, and Erdogan (2017) write "[The abstraction process] takes place in the form of isolating a concept from its specific characteristics [...]" (p. 349). Similarly, Mildenhall and Sherriff (2018) write "[...] should lead to a rich generalised and abstract understanding of mathematical concepts" (p. 402).

Informality

Informality relates to how different the representation is from representations that are traditionally used in formal education contexts (Johnson and Majewska, 2022). For example, using manipulatives would be more concrete (*informality* → *concrete*) than using symbols (*informality* → *abstract*), as academic contexts often rely on symbols rather than symbolisms. This aspect varies across countries, as different formal education systems might leverage different representations, and may also change over time as new modalities are integrated in learning materials.

For example, Zubainur, Johar, Hayati, and Ikhsan (2020) write "Using models or symbols in solving problems to bridge the concrete level to a more formal level gets emphasis in [Realistic Mathematics Education]" (p. 460). Similarly, Quigley (2021) writes "They added there does not appear to be a clear advantage in using concrete materials compared to 'more traditional methods of instruction' (Uttal, Scudder, & DeLoache, 1997, p. 38)" (p. 62).

Approximation

Approximation refers to how approximate or inaccurate the representation is with respect to the underlying concept, rather than accurate and rigorous. It is important to note here that, while we present the axis from its *concrete* end to be consistent across the section, we only found codes for the *abstract* end of the axis, meaning that scholars insists on the rigor of abstract representations, and how they leverage logical or analytical processes. That being said, both ends of the axis may be interesting to discuss when designing representations. Indeed, to what extent are these design choices informed by the concept, and to what extent is there a direct mapping between the parameters defining the final representation and those of the concept? For example, using colored cubes to represent units would be an approximate representation as the color parameter does not relate to any aspect of the underlying concept: Indeed, while the cubes all represent the number 1, they are perceptually different and may foster misinterpretations or misconceptions.

For example, Jablonka, Ashjari, and Bergsten (2017) mention "Leviatan (2008) goes as far as to describe the shift in criteria as a cultural gap: 'Tertiary mathematics is more abstract and emphasizes the inquisitive as well as the rigorous nature of mathematics' (p. 105)" (p. 75). Similarly, Zhou and Peverly (2005) write "Throughout the curriculum, Chinese teachers emphasize the development of logical, analytical, and abstract mathematical thinking" (p. 271).

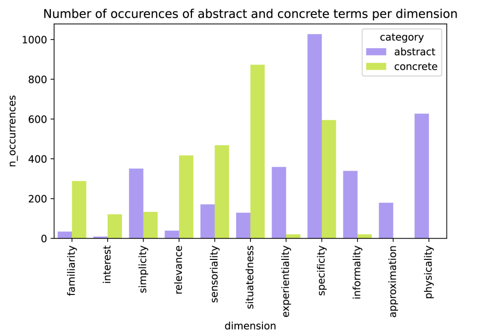

Fig. 5 Barplot of the number of occurrences of words morphologically related to *concrete* and *abstract* against each dimension of our taxonomy.

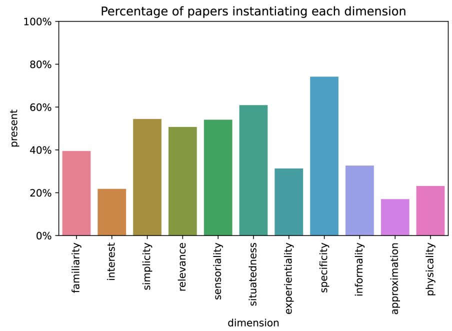

Fig. 6 Percentage of papers in our dataset instantiating each dimension of our taxonomy.

Physicality

Physicality refers to whether or not the representation is implemented using a physical object as opposed to a virtual one. Hybrid forms, such as physical objects overlaid by virtual ones using Augmented Reality would land at the center of the spectrum. Conceptually, this axis could be aligned with Milgram, Takemura, Utsumi, and Kishino (1995)'s (reversed) Reality-Virtuality continuum, including both Augmented Reality and Augmented Virtuality at the center of the axis.

For example, Pirasa (2016) writes "Examples given for a circular arc have been divided into two groups as concrete objects seen in daily life such as a piece of bagel and virtual models thought as arc that the swing makes or the clock" (p. 2847). Similarly, Satsangi, Bouck, Taber-Doughty, Bofferding, and Roberts (2016) write "In contrast to the physical form of concrete manipulatives, virtual manipulatives are generally associated with computer-based technology [...]" (p. 2).

4.2 Outlook on past research

One of the goals for our taxonomy is to help classifying past research, and as such getting a better understanding of the role of concreteness and abstraction in mathematics education, but also to identify potential gaps. While the in-depth analysis is out of scope for this paper, we present here an initial exploration in this direction in Figures 5, 6, 7, and 8. Regarding the distribution of papers according to the dimensions of our taxonomy, Figure 5 shows that some dimensions are mostly discussed as abstract (e.g. *physicality* or *approximation*) or concrete (e.g. *situatedness* or *sensoriality*). In Figure 6, we can see that the most discussed dimension is *specificity*, mostly on the abstract side, while *approximation* is less commonly mentioned. With regards to the evolution over time visualized in Figures 7 and 8, no clear trend is identifiable from this initial exploration.

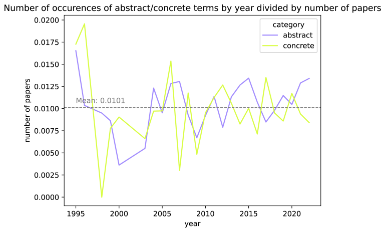

Fig. 7 Lineplot of the number of occurrences of words included in our taxonomy, over the years, corrected by the number of papers in our corpus published that year.

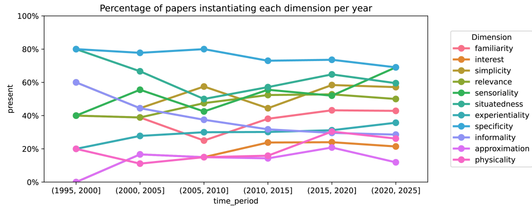

Fig. 8 Lineplot of the percentage of papers in our corpus instantiating each dimension of our taxonomy over the years.

5 Discussion

This work presents the analysis process and subsequent taxonomy generation, describing the semantic landscape of concreteness and abstraction in mathematics education. We now expand on this to illustrate how this taxonomy can be used to support scientific inquiry: (1) we describe how the axes of the taxonomy may relate to each other, and the potential of exploring these relationships, (2) we illustrate how the taxonomy can be used to frame past research. We then discuss the relevance of our taxonomy in light of previous frameworks, and finally, the potential for future work, and the limitations of the current work.

5.1 Relationship between the axes of the taxonomy

Our goal when building this taxonomy was to identify the different meanings of concreteness and abstraction used in the literature, and therefore, the resulting axes in the taxonomy differ in *meaning*. However, they do not necessarily differ in their impact on design and learning. For example, *sensoriality*, i.e. the quantity and amount of senses leveraged by the representation, differs in meaning from *interest*, i.e. how interesting and engaging the representation is for the learner. For learners, however, high sensoriality may result in increased engagement (Closser, Chan, Smith, and Ottmar, 2022).

We believe that exploring these relationships, as a researcher, but also as a field, can help identify trends, gaps, but also highlight critical questions and foster interesting conversations to grow as a field. In this section, we present a non-exhaustive list of examples showing how the axes of our taxonomy can relate to each other. In particular, relationships between objective axes, at the mapping level, and subjective axes, at the interaction level, are quite interesting because they tell us how design choices impact learners' perception and subjective experiences. These examples of relationships between axes are further visualized in Figure 9.

Physicality refers to the use of a physical representation over a virtual one. As such, it is related to *sensoriality*, as a physical object will leverage the entirety of the learners' senses, and thus necessarily more senses than a virtual one.

Situatedness refers to whether the representation is situated in the real world. As such, this is likely to also increase *familiarity*, as learners are more likely to be familiar with real-world objects. For example, choosing to represent groups as pizzas (Kaminski et al., 2008; Trninic et al., 2020; De Bock et al., 2011), or equations as gears (Papert, 1980). However, one should note that something existing in the real world does not necessarily mean that students are familiar with it, and the demographics and the geographical and cultural contexts should be considered.

Moreover, *situatedness* is likely to increase perceived *relevance* as it directly shows the concept in a potential context of use. The latter was also identified in a study in which using a situated representation for graphs by representing them as pipe systems significantly increased the perceived relevance of learners (Chatain et al., 2023).

Experientiality refers to the use of an experiential mapping of a concept, rather than a symbolic one. Often, especially in embodiment research (Ottmar et al., 2019; Nathan, 2021), this is done by leveraging bodily movements or object manipulations. Therefore, it is also likely to increase *sensoriality* and *familiarity*. Moreover, several studies showed that such designs increase learners' perceived *interest* (Cockett, Kilgour, Cockett, and Kilgour, 2015; Samur, 2012; Chatain et al., 2023; Sedig, 2008). Importantly, while increased *experientiality* may increase *sensoriality*, it is important to note that these axes are distinct: While the representations may be perceived through various senses (*sensoriality*, interaction level), for example by having visual and tangible features, the underlying meaning may not be expressed through these features and thus not be highlighted by experiencing these features (*experientiality*, mapping level).

Specificity can be connected to both *situatedness* and *experientiality*, in a bi-directional manner. Indeed, when looking for specificity, one will add details to the representation that anchor it in a certain context (*situatedness*), or that convey meaning through an experiential mapping (*experientiality*). And vice versa, as *situatedness* leverages a specific context, and as *experientiality* makes the representation more specific to a certain concept, they both increase *specificity*.

Informality is defined in contrast to formal and traditional representations such as the ones used in academic contexts. Research on math anxiety suggests that stepping away from these representations may alleviate the negative effects of math-anxiety (Jamaludin, Jabir, Wang, and Tan, 2024) and thus increase learners' perceived *interest*, but also even *simplicity* as they may feel more capable with this representations, and thus find them simpler.

Finally, adding contextual features to the representation (*situatedness*) or embedding meaning through an experiential mapping (*experientiality*), can potentially result in a lack of rigor (*approximation*).

In the future, we invite further reflections on the relationships between these meanings. For example, we believe that a series of meta-analyses exploring these relationships in depth can lead to an even more refined understanding of how to design concrete representations, and how these design choices impact the activated learning mechanisms. Moreover, these considerations can also be used to simplify the taxonomy and thus improve its usability for practitioners. For example, if a meta-analysis concludes that an experiential representation (i.e. *experientiality*) always results in

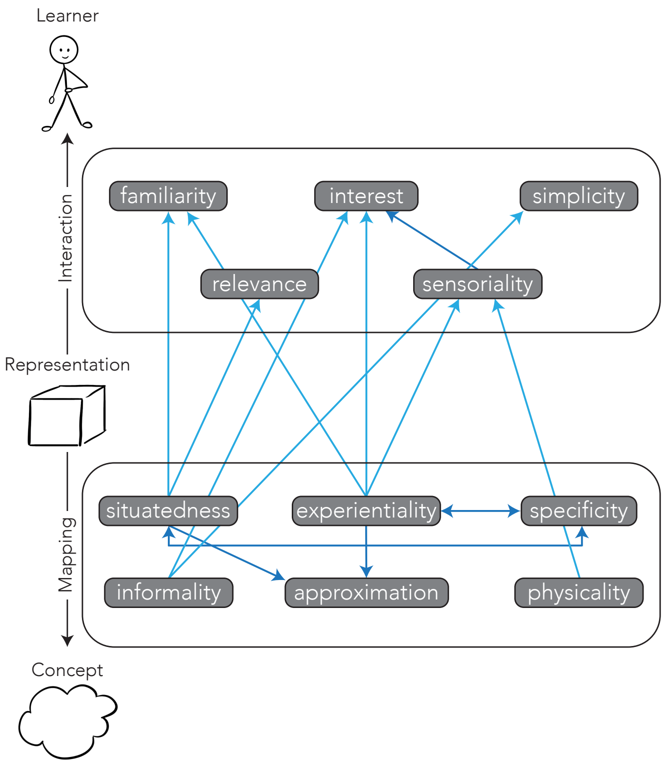

Fig. 9 Examples of relationships between axes in our taxonomy. Arrows in light blue, from the mapping categories to the interaction categories are particularly interesting as they reflect how objective design choices may impact subjective experience.

increased *interest*, the two axes can be combined by making the relationship explicit. At this stage, we refrained from making these simplifications as we solely focused on meaning.

5.2 Application on use cases

In this section, we illustrate how our taxonomy enables distinctions that are relevant for research. Indeed, it is possible to both situate representations as well as findings in the present taxonomy. In the following, we present three examples. First, Trninic et al. (2020) as well as De Bock et al. (2011) both replicated a study originally conducted by Kaminski et al. (2008) on concrete or abstract examples in preparation for transfer with minor changes to the concrete representation. Second, Chatain et al. (2023) conducted a single study with one abstract and two concrete conditions but with the concrete conditions differing in how the learning content was concretized. Finally, both Burns and Hamm (2011) and Shurr et al. (2021) compared virtual and physical manipulatives, but found different results for different populations. We organize the conditions of these use cases in our taxonomy, thus illustrating how it can be used as a tool for scientific inquiry, enabling nuanced and comprehensive comparisons of past (and future) work. For space and clarity, we only mention the main relevant axes, i.e. related to the design choices discussed in the referenced papers.

5.2.1 Familiarity versus situatedness

Kaminski et al. (2008) in the following K, De Bock et al. (2011) in the following D as well as Trninic et al. (2020) in the following T presented the students with either an abstract or a concrete example of a mathematical group (including addition, inverse, and identity element, see Figure 10). In the concrete condition, K and D both introduced an example including a group of three cups (which exist in the real world), one full, one filled 1/3 and one filled 2/3 (*situatedness* → *concrete*). The full cup was the identity element. However, T argued that as adding a full cup to a cup would usually change the volume (*familiar* → *abstract*), it is more concrete if the identity element is an empty cup (*familiar* → *concrete*). Looking at these studies in our taxonomy, the representations chosen by K, D and T are all situated (*situatedness* → *concrete*). However, the representation chosen by K and D is less familiar (*familiarity* → *abstract*) than the one chosen by T (*familiarity* → *concrete*).

K and D both found that the abstract group outperformed the concrete group in a symbolic (*experientiality* → *abstract*) transfer task involving an ancient game which employs the same rules as the preparatory example task. However, D further assessed a concrete (*situatedness* → *concrete*) transfer task. D was able to replicate K's findings. However, they further found that the concrete group performed better in the concrete transfer. This goes to show that it is not sufficient to ask which representation is best suited for learning, but that the content of the assessment and the design of the representations employed in the assessment is also important.

The minor change in representation, which made the representation more familiar, lead to T finding no difference between the groups in this task. However, as D, they found that for a more concrete (*familiar* → *concrete*) transfer task involving partially filled buckets of tennis balls, the concrete condition significantly outperformed the abstract group. Applying the taxonomy, we could therefore argue that it is not sufficient for a concrete example to include objects that exist in the real world, but that additionally, the situation needs to be familiar, or realistic to the students.

This use case also illustrates the importance of the multi-dimensionality of our taxonomy. For example, one may assume that a representation that is *concrete* in terms of *situatedness*, that is, a representation that exists in the real world, will most often also be more *concrete* on the *familiarity* dimension. While it is often the case, failing to explicitly account for *familiarity* may result in unproductive representations, such as the one presented in K's study.

5.2.2 Experientiality versus situatedness

The study conducted by Chatain et al. (2023) investigated how learning with abstract and concrete (*situatedness* → *concrete* or *experientiality* → *concrete*, see Figure 11) representations impacts learning about graph theory, specifically the maximum flow problem. The condition they denote as MNPL used an interactive representation (*experientiality* → *concrete*) with the mathematical graphs represented with circles

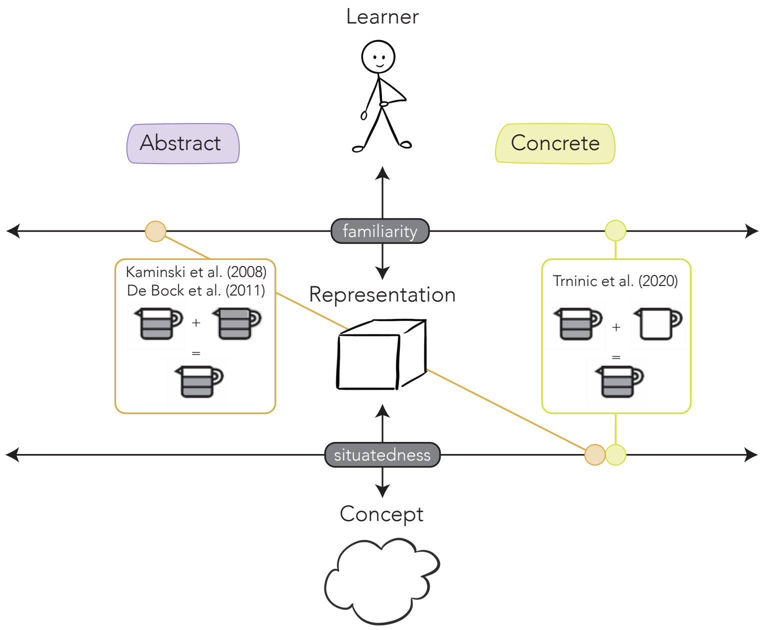

Fig. 10 Comparison of the concrete representations used by Kaminski et al. (2008) and De Bock et al. (2011) and the one used by Trninic et al. (2020) in our taxonomy.

and lines (*situatedness* → *abstract*). The other condition (denoted as EMBD) also used an interactive representation (*experientiality* → *concrete*). However, the graphs were represented as pipe systems (*situatedness* → *concrete*). In both cases, the goal was to maximize the flow that travels through the network, either in the circles and lines representation (MNPL) or by maximizing the water flow between a lake and a town (EMBD).

Chatain et al. (2023) found no difference between groups in engagement (*interest* → *concrete*) but found EMBD to be perceived as more relevant by the students (*relevance* → *concrete*) than MNPL (*relevance* → *abstract*). Applying the taxonomy, we could argue that simply allowing students to interact with a representation is not sufficient to allow them to relate to the problem, or to perceive it as relevant to them. In addition, the problem needs to be situated in the real world.

With this use case, we illustrate that our taxonomy can also be further explored to identify relationship between axes, for example the impact of *situatedness* on *relevance* as demonstrated by Chatain et al. (2023).

5.2.3 Physicality and individual differences

As a final use case, we want to illustrate a situation, in which the same representation might result in different outcomes based on individual differences. Here, we explore the difference between physical manipulatives, for example physical blocks to perform additions, and virtual manipulatives, for example the same blocks, but on iPad

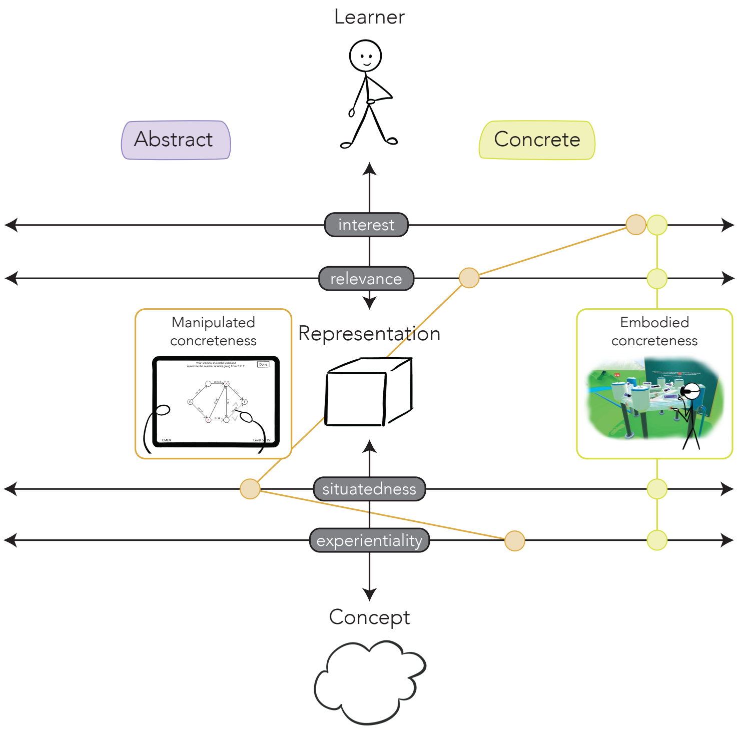

Fig. 11 Comparison of the concrete representations used by Chatain et al. (2023) in our taxonomy.

(See Figure 12). Both conditions support experiential sense-making (*experientiality* → *concrete*). However, *virtual* manipulatives are more abstract in terms of *physicality*.

When looking at neurotypical learners, Burns and Hamm (2011) found a positive yet not significant advantage for physical manipulatives. However, through use case exploration with 3 learners who have Autism Spectrum Disorder (ASD), Shurr et al. (2021) found an advantage for virtual manipulatives. They argued that one of the main advantages were that the virtual manipulatives resulted in more stable and accurate manipulations.

Here again, looking at this within our proposed taxonomy, we can say that in these studies, the increased *physicality* also resulted in increased *sensoriality* for the learners. Indeed, although both types of manipulatives leverage touch, the physical manipulatives leverage it more (*sensoriality* → *concrete*). This distinction could potentially explain why learners with ASD favored the more *abstract* representation.

5.3 Relations to previous frameworks

With our taxonomy, we offer a data-driven exploration of the meanings of the words "abstract" and "concrete" in mathematics education. However, other works have attempted at defining these terms before, and offered conceptual multi-dimensional frameworks to describe concreteness and abstraction. In this section, we show examples

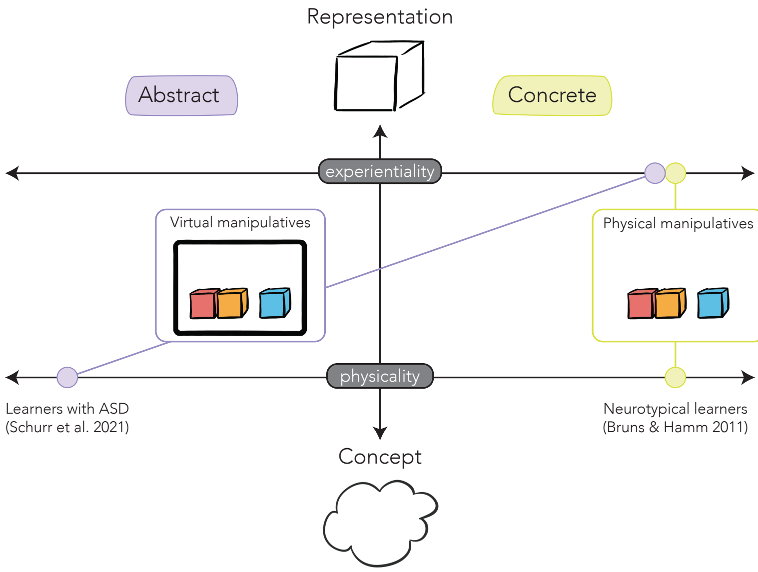

Fig. 12 Comparison of virtual and physical manipulatives for learners with ASD (Shurr et al., 2021) and neurotypical learners (Burns and Hamm, 2011) in our taxonomy.

of how our taxonomy relates to these existing frameworks, and how, while capturing some previously mentioned aspects, it also offers a more complete picture of the space that can be explored when designing representations of mathematical concepts.

5.3.1 Organizing framework for categorizing external knowledge representations by Belenky and Schalk (2014)

Our taxonomy aligns with the framework suggested by Belenky and Schalk (2014). However, the two frameworks may be useful in different contexts.

In their framework, Belenky and Schalk (2014) organized external representations into two dimensions, groundedness and relevance. Groundedness here means to be instantiated in some context while relevance refers to the relevance of individual aspects of the external representation to the problem to be solved. Therefore, the groundedness dimension may be part of quite a few of the axes found by us, such as *familiarity*, *relevance* (in the taxonomy presented here, this axis refers to how relevant the learner perceives a certain concept in a given representation), *specificity*, or - most importantly - *situatedness*.

However, while their relevance dimension may be included in our situatedness axis, we focus less on what the learner needs to solve a certain problem, but more on what a representation will activate in the learner. We suggest that for practitioners who want to design effective visual external representations for a specific problem, the taxonomy by Belenky and Schalk (2014) may be suited. For research design, specifically if more complex representations over several modalities are involved, our taxonomy will help to organize cognitive processes found in connection with a wider set of axes.

5.3.2 Concreteness fading

Looking at instructional design, a common approach discussed in mathematics education is *concreteness fading* (Bruner, 1974; Fyfe et al., 2014), relying on a progression of representations from enactive to iconic to symbolic.

Commonly, the enactive representation is a physical object (*physicality* → *concrete*) that can be manipulated (*sensoriality* → *concrete*). Notably, however, Suh et al. (2020) explained that related frameworks, such as the Virtual-Representational-Abstract one (Bouck et al., 2017; Cooper, 2012), leverage a virtual manipulative (*sensoriality* → *concrete* still, but *physicality* → *abstract*) in the first stage. The iconic form is usually a visual and pictorial representation, meaning that it moves towards the abstract end of the spectrum on the *sensoriality* axis, but also *situatedness* axis as the representations gets further detached from the real world. Finally, the last form, the symbolic one, is abstract along the *experientiality* axis.

We present this evolution on Figure 13, focusing on the main relevant axes for clarity.

Interestingly, we can see that concreteness fading includes several axes at the mapping level, between the representation and the concept. This level is necessarily domain dependent, as it depends on the underlying concepts, justifying why concreteness fading is difficult to generalize to other domains outside of mathematics (Kokkonen and Schalk, 2021).

5.3.3 External representations in concreteness fading by Fyfe and Nathan (2019)

The observation that representations involved in concreteness fading may involve different axes has been made before. Fyfe and Nathan (2019) explained that concreteness fading is an under-specified theory, and suggested to (1) use the term "idealized" representations rather than "abstract" representations, (2) consider "concrete" and "idealized" on a spectrum rather than a dichotomy, and (3) consider several perceptual and conceptual axes along which a representation can be more or less concrete.

They suggest four axes, which we compare to ours in Table 4.

Here, the main distinction is that our taxonomy separates the perceptual features of the representation from its situatedness, as a representation can be perceptually rich (e.g. leveraging various colors and patterns), while not looking real. Moreover, we consider physicality slightly differently, as we consider the *sensoriality*, which is learner-dependent and relates to how learners senses are leveraged by the representations, as well as whether the representation is virtual or physical (our definition for *physicality*).

Importantly, their framework focuses on serving the design of representations within the concreteness fading paradigm, and is well suited for this goal. In contrast, our taxonomy is focused on a more general purpose, as illustrated by the extra nuances it accommodates for, and supports the comparison of more types of concrete representations. Because of these diverging goals, the approaches are also different: While their framework is conceptual in nature, we offer a data-driven perspective of the semantical landscape covered by the literature.

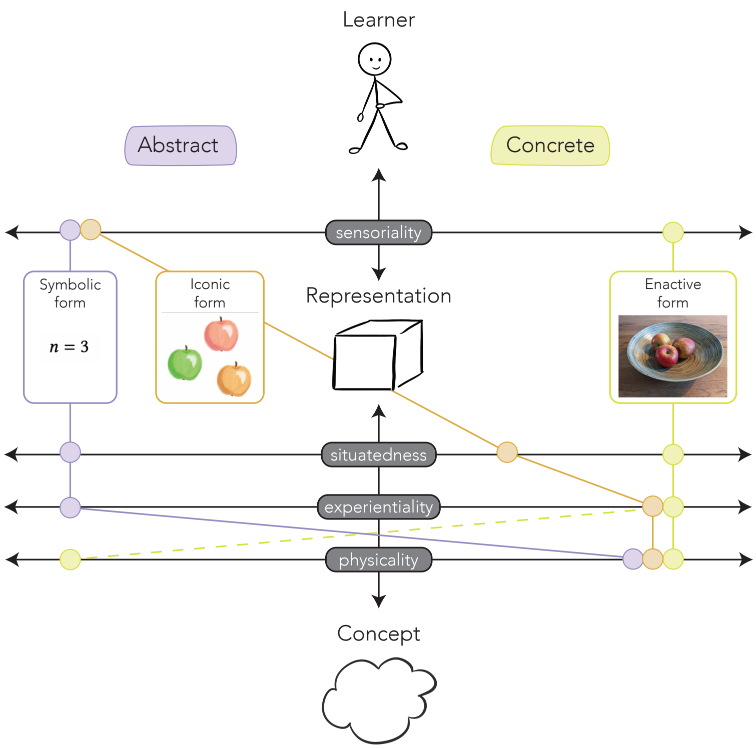

Fig. 13 Concreteness fading (Bruner, 1974; Fyfe et al., 2014) represented in our taxonomy. Only main relevant axes are included, and adjacent points should be considered as overlapping.

5.4 Gaps for future work

In the future, we hope this taxonomy helps getting a better understanding of which *concrete* and *abstract* aspects of a certain representation support meaning-making in mathematics education, in four ways:

The first step towards this goal will be to conduct further analyzes according to the axes of the taxonomy to get a fine-grained understanding of which concrete and abstract aspects of representations truly support learning. For example, when arguing that a concept is too abstract and therefore difficult to learn, being able to understand exactly what makes the concept abstract may result in better design recommendations, informed by such fine-grained meta-analysis. Is the concept abstract because it cannot be seen? Then maybe we should design visible representations. Is it abstract because it is universal? Then we may want to highlight the power of generalization by designing several problems relying on the same concept. Is it abstract because it is not familiar enough? Then we may want to choose a representation that relates to the learner's prior experiences.

Second, the relationships between the axes should be further discussed and explored to identify trends and gaps, but also critical questions. As a field, this is an opportunity to further reflect on how we design concrete representations, and what learning

Axes from Fyfe and Nathan (2019)	**Comparison with our taxonomy**
Physicality distinguishes two-dimensional representations (e.g. drawing on paper) from three-dimensional ones (e.g. manipulative).	This axis is not directly represented in our taxonomy. However, it is related to our *sensoriality* axis, reflecting how many and how much senses are leveraged by the representation. Our own *physicality* axis, in contrast, refers to whether the representation is virtual or physical, based on the use of technology.
Perceptual richness refers to the visual features of the representation, such as colors and patterns, and how they are leveraged to make the representation look real.	This axis included both our *sensoriality* axis, regarding the visual features, and our *situatedness* axis, regarding how real the representation looks like.
Familiarity refers to how close the representation is to learners' prior experiences.	This axis is directly equivalent to our *familiarity* axis.
Context refers to how much the representation in embedded in a narrative context.	This axis is contained in our *situatedness* axis, which relates to how much the representation exists in the real world. While narrative is a way of doing this, it is not the only one. For example, in our taxonomy, using a depiction of a mountain to discuss the concept of derivatives would be more *concrete* along the *situatedness* axis than using a function graph.

Table 4 Comparison of the axes described by Fyfe and Nathan (2019) (left) with ours (right).

mechanisms they leverage. This includes exploring relationships between the axes in the mapping level and those in the interaction level, and these relationships would give us further understanding in how certain objective design decisions impact learners' subjective experiences. This would enable further simplification of the taxonomy, converging towards a data-informed design tool for practitioners.

Third, a thorough evaluation of previous research would help identify insights as well as research gaps over the years, to formulate both design insights and future research questions. We presented preliminary insights in Section 4.2, but future work would include an in-depth meta-analysis of the role of concreteness and abstraction in mathematics education, discriminating design parameters described in our taxonomy.

Finally, as discussed in the following limitations section, we believe this taxonomy, both in its definition and in its usage, can be further refined. Future work would include conducting interviews with experts, either researchers or practitioners, asking them to organize their work using the taxonomy.

5.5 Limitations

This work comes with three main limitations.

First, while the coding of the data was done by four researchers with highly different backgrounds and relevant expertise, the analysis process and most of the interpretations of the generated themes was performed by two educational researchers (R3, R4) who shared a common doctoral supervisor. We tried to be highly transparent about the positionality of the researchers involved (Table 2) and documented the analysis process intensively (in Section 3.4 as well as in Appendix A). However, it is clear that the taxonomy at the present state is a first suggestion that should be used, criticized, and further improved together with researchers and practitioners in the educational field.

Second, we have chosen the ERIC database to obtain the corpus of this analysis. While this choice provided us with a corpus that was highly education-specific, we might have missed works by adjacent communities such as educational researchers situated in computer science.

Third, as it is not possible to know exactly whether a list of adjectives is a stylistic choice by the authors, listing several words of similar meaning to create emphasis or whether it is a list of adjectives of different meanings, we have coded words "more related to concrete and abstract". For example, Bleazby (2015) writes that experience is "subjective, concrete, and situated" (p. 673). In this case, we coded *subjective* and *situated* as more related to concrete. However, Bleazby (2015) might have had a different meaning of concrete in mind. Therefore, again, it is crucial to revise the taxonomy based on user feedback and in collaboration with experts in the field.

Finally, we acknowledge that our taxonomy, in its current state, is quite complex as it contains a large number of axes, making it potentially difficult to apply to use cases. This is due to the fact that we focused first on building a taxonomy that reflects our underlying data, to get the complete picture of the semantic landscape. However, in the future, we hope that this taxonomy can be simplified through usage as well as further meta-analyses and studies evaluating the impact of each axis on learning outcomes and other relevant metrics.

6 Conclusion

We analyzed 368 articles using a qualitative data-driven approach to shed light on the semantic landscape on abstraction and concreteness in mathematics education, and identified eleven axes. We found that five of these axes focused on the interaction level between the learner and the representation (i.e. how is the representation perceived by this specific learner, in a subjective manner) and the remaining six axes focused on the mapping level between the representation and the concept (i.e. what aspects of the concept are represented and how, in an objective manner). Here, we presented the taxonomy, illustrated how it can be leveraged to reflect on how we consider concrete representations, and presented three use cases to show how it can be used to organize existing research.

Leveraging this tool, we would like to encourage the following future endeavors. First, with our taxonomy, we focused on providing a complete overview of the *meanings* of the terms "concrete" and "abstract" in the mathematics education literature. Now, the relationships between these axes should be further investigated: What relationships have been explored, and validated or disproved? What gaps remain to be explored? We offer an initial discussion in this paper, but we believe that pursuing this approach thoroughly and with the wider community can raise critical questions and inquiries in the field, and eventually, through empirical evidence, help refine the tool towards a simplified version that can be widely used in practice.

Second, we encourage scholars to use the tool to better situate their studies and designs in this landscape. As illustrated in our paper, we believe that this opens the floor to nuanced comparisons of existing studies, and further conversations about both design and learning processes leveraging concreteness and abstraction. We believe that our comprehensive taxonomy as well as these explorations can further facilitate future meta-analyses and reviews of the field.

Third, we believe that this tool should evolve through practice and discussions in the community. Therefore, we would like to encourage researchers and practitioners alike to use the taxonomy to reflect on their own studies and design processes, and share their experiences and suggestions. While our approach is anchored in data, this tool is designed to evolve and improve through continued use.

To support these endeavors, we provide three dedicated templates in Appendix B of this manuscript for interested readers who would like to apply the framework to their own work.

Finally, in addition to the future avenues discussed in Section 5.4, we note that this taxonomy might be applicable to other STEM disciplines as well. Specifically, many scientific concepts are physical in the sense that they exist in the real-world. However, they are often not perceptible (*sensoriality* → *abstract*), either because they are too small, for example molecules, or too big, for example the atmosphere (Niebert and Gropengiesser, 2015). This taxonomy might support organizing findings related to targeting and representing imperceptible concepts.

Acknowledgements. We would like to thank the Future Learning Initiative of ETH Zurich for supporting this project, as well as Dragan Trninić, Sofia Tancredi and Tomohiro Nagashima for inspiring conversations about concreteness in math education.

Statements and Declarations

6.1 Competing interests

The authors declare **NO** financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

References

Abrahamson, D., Nathan, M.J., Williams-Pierce, C., Walkington, C., Ottmar, E.R., Soto, H., Alibali, M.W. (2020). The future of embodied design for mathematics teaching and learning. *Frontiers in education* (Vol. 5, p. 147).

Ainsworth, S. (2006). Deft: A conceptual framework for considering learning with multiple representations. *Learning and instruction*, *16*(3), 183–198.

Altintas, E. (2018). Analyzing students' views about mathematics teaching through stories and story generation process. *Educational Research and Reviews*, *13*(7), 249–259.

Altintas, E., & Ilgün, S. (2017). Exploring the opinions about the concepts of "formula" and "rule" in mathematics. *Educational Research and Reviews*, *12*(19), 956–966.

Artigue, M. (2009). The teaching and learning of mathematics at the university level. *Coleccion Digital Eudoxus*(7), 1377–1385.

Balcerak Jackson, B. (2014). Verbal disputes and substantiveness. *Erkenntnis*, *79*(1), 31–54.

Becker, N. (2022). *The verbal dispute about verbal disputes* (Unpublished doctoral dissertation). University of St Andrews.

Belenky, D.M., & Schalk, L. (2014). The effects of idealized and grounded materials on learning, transfer, and interest: An organizing framework for categorizing external knowledge representations. *Educational Psychology Review*, *26*, 27–50.

Bird, S., Klein, E., Loper, E. (2009). *Natural language processing with python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.".

Bishop, A., Mellin-Olsen, S., van Dormolen, J. (1991). *Mathematical knowledge: Its growth through teaching* (Vol. 10). Springer Science & Business Media.

Bleazby, J. (2015). Why some school subjects have a higher status than others: The epistemology of the traditional curriculum hierarchy. *Oxford Review of Education*, *41*(5), 671–689.

Bossard, C., Kermarrec, G., Buche, C., Tisseau, J. (2008). Transfer of learning in virtual environments: a new challenge? *Virtual Reality*, *12*(3), 151–161.

Bouck, E.C., Park, J., Sprick, J., Shurr, J., Bassette, L., Whorley, A. (2017). Using the virtual-abstract instructional sequence to teach addition of fractions. *Research in developmental disabilities*, *70*, 163–174.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77–101.

Bruner, J. (1974). *Toward a theory of instruction.* Harvard university press.

Burns, B.A., & Hamm, E.M. (2011). A comparison of concrete and virtual manipulative use in third-and fourth-grade mathematics. *School Science and Mathematics*, *111*(6), 256–261.

Chalmers, D.J. (2011). Verbal disputes. *Philosophical Review*, *120*(4), 515–566.

Chatain, J., Varga, R., Fayolle, V., Kapur, M., Sumner, B. (2023). Grounding graph theory in embodied concreteness with virtual reality. *Proceedings of the seventeenth international conference on tangible, embedded, and embodied interaction.*

Clarke, V., & Braun, V. (2021). Thematic analysis: a practical guide. *Thematic Analysis*, 1–100.

Closser, A.H., Chan, J.Y.-C., Smith, H., Ottmar, E.R. (2022). Perceptual learning in math: Implications for educational research, practice, and technology.

Cockett, A., Kilgour, P.W., Cockett, A., Kilgour, P. (2015). Mathematical manipulatives: Creating an environment for understanding, efficiency, engagement, and enjoyment. *Teach Collection of Christian Education*, *1*(1), 47–54.

Coles, A. (2014). Transitional devices. *For the learning of mathematics*, *34*(2), 24–30.

Coles, A., & Sinclair, N. (2019). Re-thinking 'concrete to abstract'in mathematics education: Towards the use of symbolically structured environments. *Canadian Journal of Science, Mathematics and Technology Education*, *19*, 465–480.

Cooper, T.E. (2012). Using virtual manipulatives with pre-service mathematics teachers to create representational models. *International Journal for Technology in Mathematics Education*, *19*(3), 105-115.

Courant, R., & Robbins, H. (1996). *What is mathematics?: an elementary approach to ideas and methods.* Oxford University Press, USA.

Crosswhite, F.J., et al. (1986). *Second international mathematics study detailed report for the united states.*

Davis, P., Hersh, R., Rota, G. (1981). *The mathematical experience.* Harvester Press.

De Bock, D., Deprez, J., Van Dooren, W., Roelens, M., Verschaffel, L. (2011). Abstract or concrete examples in learning mathematics? a replication and elaboration of kaminski, sloutsky, and heckler's study. *Journal for research in Mathematics Education*, *42*(2), 109–126.

Dewey, J. (1910). *How we think.* D. C. HEATH & CO.

DiSessa, A.A. (2018). Computational literacy and "the big picture" concerning computers in mathematics education. *Mathematical thinking and learning*, *20*(1), 3–31.

Dörfler, W. (1991). Forms and means of generalization in mathematics. *Mathematical knowledge: Its growth through teaching* (pp. 61–85). Springer.

Dossey, J.A. (1992). The nature of mathematics: Its role and its influence. *Handbook of research on mathematics teaching and learning*, *39*, 48.

Dreyfus, T. (2020). Abstraction in mathematics education. *Encyclopedia of mathematics education*, 13–16.

Durmus, S., & Karakirik, E. (2006). Virtual manipulatives in mathematics education: A theoretical framework. *Turkish Online Journal of Educational Technology-TOJET*, *5*(1), 117–123.

Ernest, P. (1985). The philosophy of mathematics and mathematics education. *International Journal of Mathematical Education in Science and Technology*, *16*(5), 603–612.

Fisher, R.A. (1970). Statistical methods for research workers. *Breakthroughs in statistics: Methodology and distribution* (pp. 66–70). Springer.

Fyfe, E.R., McNeil, N.M., Son, J.Y., Goldstone, R.L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational psychology review*, *26*, 9–25.

Fyfe, E.R., & Nathan, M.J. (2019). Making "concreteness fading" more concrete as a theory of instruction for promoting transfer. *Educational Review*, *71*(4), 403–422.

Ganesh, T.G., & Middleton, J.A. (2006). Challenges in linguistically and culturally diverse elementary settings with math instruction using learning technologies. *The Urban Review*, *38*, 101–143.

Gravemeijer, K. (2011). How concrete is concrete? *Journal on Mathematics Education*, *2*(1), 1–14.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1-3), 335–346.

Hirsch, E. (2005). Physical-object ontology, verbal disputes, and common sense. , *70*(1), 67–97, https://doi.org/10.1111/j.1933-1592.2005.tb00506.x

Jablonka, E., Ashjari, H., Bergsten, C. (2017). "much palaver about greater than zero and such stuff"–first year engineering students' recognition of university mathematics. *International Journal of Research in Undergraduate Mathematics Education*, *3*, 69–107.

Jamaludin, A., Jabir, A.I., Wang, F., Tan, A.L. (2024). Low-progress math in a high-performing system: The role of math anxiety in singapore's elementary learners. *The Asia-Pacific Education Researcher*, *33*(5), 1071–1082.

Jenkins, C.S.I. (2014). Merely verbal disputes. , *79*(1), 11–30, https://doi.org/10.1007/s10670-013-9443-6

Johnson, M., & Majewska, D. (2022). Formal, non-formal, and informal learning: What are they, and how can we research them? research report. *Cambridge University Press & Assessment*, 0–36.

Kaminski, J.A., Sloutsky, V.M., Heckler, A.F. (2008). The advantage of abstract examples in learning math. *Science*, *320*(5875), 454–455.

Kokkonen, T., & Schalk, L. (2021). One instructional sequence fits all? a conceptual analysis of the applicability of concreteness fading in mathematics, physics, chemistry, and biology education. *Educational Psychology Review*, *33*(3), 797–821.

Kövecses, Z. (2010). *Metaphor: A practical introduction.* Oxford university press.

Lakatos, I. (1976). A renaissance of empiricism in the recent philosophy of mathematics. *The British Journal for the Philosophy of Science*, *27*(3), 201–223.

Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from* (Vol. 6). New York: Basic Books.

Lang, H., & Pagliaro, C. (2007). Factors predicting recall of mathematics terms by deaf students: Implications for teaching. *Journal of deaf studies and deaf education*, *12*(4), 449–460.

Leviatan, T. (2008). Bridging a cultural gap. *Mathematics Education Research Journal*, *20*(2), 105–116.

Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic inquiry.* sage.

Löhr, G. (2023). Does the mind care about whether a word is abstract or concrete? why concreteness is probably not a natural kind. *Mind & Language*, 1–20.

McElhaney, K.W., Chang, H.-Y., Chiu, J.L., Linn, M.C. (2015). Evidence for effective uses of dynamic visualisations in science curriculum materials. *Studies in Science Education*, *51*(1), 49–85.

Mildenhall, P., & Sherriff, B. (2018). Using multiple metaphors and multimodalities as a semiotic resource when teaching year 2 students computational strategies. *Mathematics Education Research Journal*, *30*(4), 383–406.

Milgram, P., Takemura, H., Utsumi, A., Kishino, F. (1995). Augmented reality: A class of displays on the reality-virtuality continuum. *Telemanipulator and telepresence technologies* (Vol. 2351, pp. 282–292).

Mitchelmore, M.C., & White, P. (1995). Abstraction in mathematics: Conflict, resolution and application. *Mathematics Education Research Journal*, *7*(1), 50–68.

Muma, J.R. (1991). Experiential realism: Clinical implications. *Pragmatics of language* (pp. 229–247). Springer US.

Nathan, M.J. (2021). *Foundations of embodied learning: A paradigm for education.* Routledge.

Niebert, K., & Gropengiesser, H. (2015). Understanding starts in the mesocosm: Conceptual metaphor as a framework for external representations in science teaching. *International Journal of Science Education*, *37*(5-6), 903–933, https://doi.org/10.1080/09500693.2015.1025310

Noonburg, D., & Cid, A.A. (2005). Poppler [Computer software manual]. (Poppler is a fork of the xpdf PDF viewer developed by Derek Noonburg of Glyph and Cog, LLC. xpdf is written by Derek Noonburg. libpoppler is a fork of xpdf-3.00. Current Maintainer: Albert Astals Cid.)

Ottemo, A., Berge, M., Silfver, E. (2020). Contextualizing technology: Between gender pluralization and class reproduction. *Science Education*, *104*(4), 693–713.

Ottmar, E.R., Walkington, C., Abrahamson, D., Nathan, M.J., Harrison, A., Smith, C. (2019). Embodied mathematical imagination and cognition (emic) working group. *North American Chapter of the International Group for the Psychology of Mathematics Education*, 1913–1923.

Papert, S. (1980). The gears of my childhood. *Mindstorms: Children, Computers, and Powerful Ideas. Basic Books. USA*, vi–viii.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pennington, J., Socher, R., Manning, C.D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1532–1543).

Penrose, R. (1991). The emperor's new mind. *RSA Journal*, *139*(5420), 506–514.

Petersen, L.A., & McNeil, N.M. (2013). Effects of perceptually rich manipulatives on preschoolers' counting performance: Established knowledge counts. *Child development*, *84*(3), 1020–1033.

Pirasa, N. (2016). The connection competencies of pre-service mathematics teachers about geometric concepts to daily-life. *Universal Journal of Educational*

Research, *4*(12), 2840–2851.

Quigley, M.T. (2021). Concrete materials in primary classrooms: Teachers' beliefs and practices about how and why they are used. *Mathematics Teacher Education and Development*, *23*(2), 59–78.

Rau, M.A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing stem learning. *Educational Psychology Review*, *29*, 717–761.

Rich, K.M., & Yadav, A. (2020, May). Applying Levels of Abstraction to Mathematics Word Problems. *TechTrends: Linking Research and Practice to Improve Learning*, *64*(3), 395–403.

Samur, Y. (2012). Measuring engagement effects of educational games and virtual manipulatives on mathematics.

Satsangi, R., Bouck, E.C., Taber-Doughty, T., Bofferding, L., Roberts, C.A. (2016). Comparing the effectiveness of virtual and concrete manipulatives to teach algebra to secondary students with learning disabilities. *Learning Disability Quarterly*, *39*(4), 240–253.

Schalk, L., Saalbach, H., Stern, E. (2016). Approaches to foster transfer of formal principles: which route to take? *PloS one*, *11*(2), e0148787.

Schoenfeld, A.H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. *Handbook of research on mathematics teaching and learning*, *334370*, 1–38.

Sedig, K. (2008). From play to thoughtful learning: A design strategy to engage children with mathematical representations. *Journal of Computers in Mathematics and Science Teaching*, *27*(1), 65–101.

Sezgin Memnun, D., Aydin, B., Özbilen, Ö., Erdogan, G. (2017). The abstraction process of limit knowledge. *Educational Sciences: Theory and Practice*, *17*(2), 345–371.

Shurr, J., Bouck, E.C., Bassette, L., Park, J. (2021). Virtual versus concrete: A comparison of mathematics manipulatives for three elementary students with autism. *Focus on Autism and Other Developmental Disabilities*, *36*(2), 71–82.

Sinclair, N. (2001). The aesthetic" is" relevant. *For the learning of mathematics*, *21*(1), 25–32.

Sinclair, N. (2011). Aesthetic considerations in mathematics. *Journal of Humanistic Mathematics*, *1*(1), 2–32.

Stahl, N.A., & King, J.R. (2020). Expanding approaches for research: Understanding and using trustworthiness in qualitative research. *Journal of developmental education*, *44*(1), 26–28.

Suh, S., Lee, M., Law, E. (2020). How do we design for concreteness fading? survey, general framework, and design dimensions. *Proceedings of the 19th acm international conference on interaction design and children* (pp. 581–588).

Trninic, D., Kapur, M., Sinha, T. (2020). The disappearing "advantage of abstract examples in learning math". *Cognitive Science*, *44*(7), e12851.

Uttal, D.H., Scudder, K.V., DeLoache, J.S. (1997). Manipulatives as symbols: A new perspective on the use of concrete objects to teach mathematics. *Journal of applied developmental psychology*, *18*(1), 37–54.

Vermeulen, I. (2018). Verbal disputes and the varieties of verbalness. *Erkenntnis*, *83*(2), 331–348, https://doi.org/10.1007/s10670-017-9892-4

Wagner, R. (2019). Mathematical abstraction as unstable translation between concrete presentations. *Philosophy of Mathematics Education Journal*, *35*, (No page numbers).

Wilensky, U. (1991). *Abstract meditations on the concrete and concrete implications for mathematics education.* Epistemology and Learning Group, MIT Media Laboratory Cambridge, MA.

Wilkinson, A. (2021). What is mathematics? *The New Yorker*, (No page numbers).

Wolfmeyer, M. (2018). Teaching ecojustice in stem methods courses. *Issues in Teacher Education*, *27*(2), 83–95.

Woodhouse, H. (2012). Mathematics as liberal education: Whitehead and the rhythm of life. *Interchange*, *43*, 1–23.

Zevenbergen, R. (2001). Changing contexts in tertiary mathematics: Implications for diversity and equity. *The teaching and learning of mathematics at university level: An icmi study* (pp. 13–26). Springer.

Zhou, Z., & Peverly, S.T. (2005). Teaching addition and subtraction to first graders: A chinese perspective. *Psychology in the Schools*, *42*(3), 259–272.

Zubainur, C.M., Johar, R., Hayati, R., Ikhsan, M. (2020). Teachers' understanding about the characteristics of realistic mathematics education. *Journal of Education and Learning (EduLearn)*, *14*(3), 456–462.

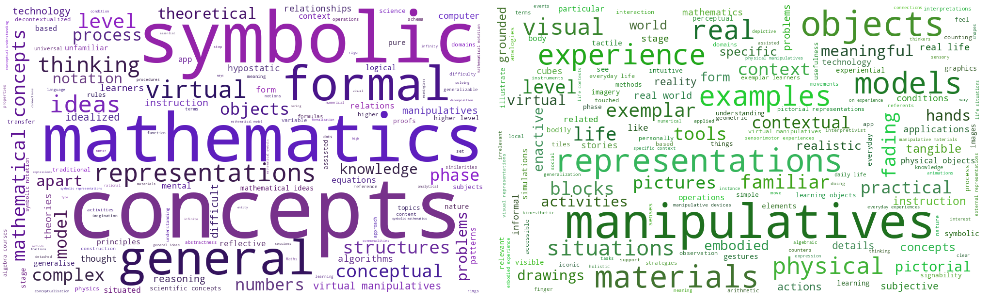

Fig. A1 Word clouds of the words selected for abstraction (left) in concreteness (right) at the end of the pre-processing phase.

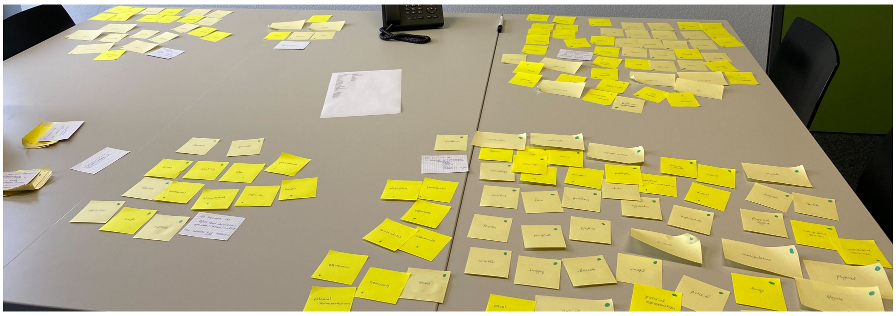

Fig. A2 The codes after a first familiarization and organization into codes related to external representations, overarching characteristics, subjective characteristics and contextual codes.

Appendix A Thematic analysis process

After an initial pre-processing phase, the remaining words were visualized in word clouds (see Figure A1).

The detailed thematic analysis process based on these words is described in the following. For a full list of the codes used in this process together with data examples, please see the corresponding table in the supplementary material.

Iteration 1

The first list including all occurrences higher than 4 and excluding *algebra*, *on*, *the*, *non* and *mathematics* were transferred to post-its. R4 sorted the post-its a first time according to thematic relations (see Figure A2).

In summary, the following **concrete** categories were generated:

- Kind of external representations (visual, pictorial, tactile, stories)
- Descriptions of external representations depending on <u>context</u> (real, informal, particular)
- Subjective characteristics (familiar, everyday experience, move, interest)
- Higher level descriptions (process, (ir)relevant, simple)
- Concreteness fading

Discussing these first organization, the researchers R3 and R4 found that some **concrete** codes seemed to describe the concept, some the external representation of it (the physical object), some the interaction between the representation and the learner and some the subjective experience of the learner (i.e. their conceptualizations and experiences). They organized the codes accordingly into a first graph structure presented in Figures A3 and A4.

At this point, the following list of words were removed either due to relating to terms specific to concreteness fading and therefore having an analog in the abstract codes (*e.g.* level, stage, phase, conditions) or due to them being clearly coded related to concrete as synonym for specific (*e.g.* strategies, methods, operations, way): strategies; methods; operations; level; stage; phase; conditions; form; problems; fading; concepts; shapes; way; tasks; nature; instruction.

For the **abstract** terms, sorting was more diverse as the term was used in many different ways to assign a certain value or importance to a concept. Generally, we found that abstract often related to either being fundamental or being universal and hence being conceptualized vertically. Alternatively, abstract was conceptualized horizontally by indicating relationships, patterns and similarities. We also found codes connected to the learner (i.e. thinking, conceptualizing) and to prejudice (difficult, boring) (see Figures A3 and A4). Finally, some codes related to tools for abstracting such as processes, algorithms and procedures.

We removed the following words due to the same reasons as above: activities; fractions; rings; set; form; phase; dots; topics; stage; nature; objects; sessions; manner; ways; subjects; step; condition; content; approach; problems; domains.

Iteration 2

After the first iteration, the authors decided to distinguish between codes associated to the learner, the representation and the concept. Hence, separately, R3 and R4 reorganized the codes according to these three poles as well as the interactions between them (see Figures A5, A6, A7, and A8). Initially, they also assumed there to be codes describing the interaction between the learner and the concept. However, they later found that any understanding of the concept by the learner is mediated by the representation.

Iteration 3

In a third iteration, R3 and R4 re-arranged the triangle representation to a linear taxonomy learner-representation-concept, resulting in the final taxonomy described in Section 4.1.

Iteration 4

After the first revision round on February 07, 2025, we expanded our corpus to include literature of younger ages as well, as described in the methods section. This resulted in several new words for **concrete**: exploration, experimentation, empirical, manipulative objects, **metaphors**, **fun**, real objects, perception, **app-based** manipulative, operational, rods, physical materials, personalization, discovery, operational

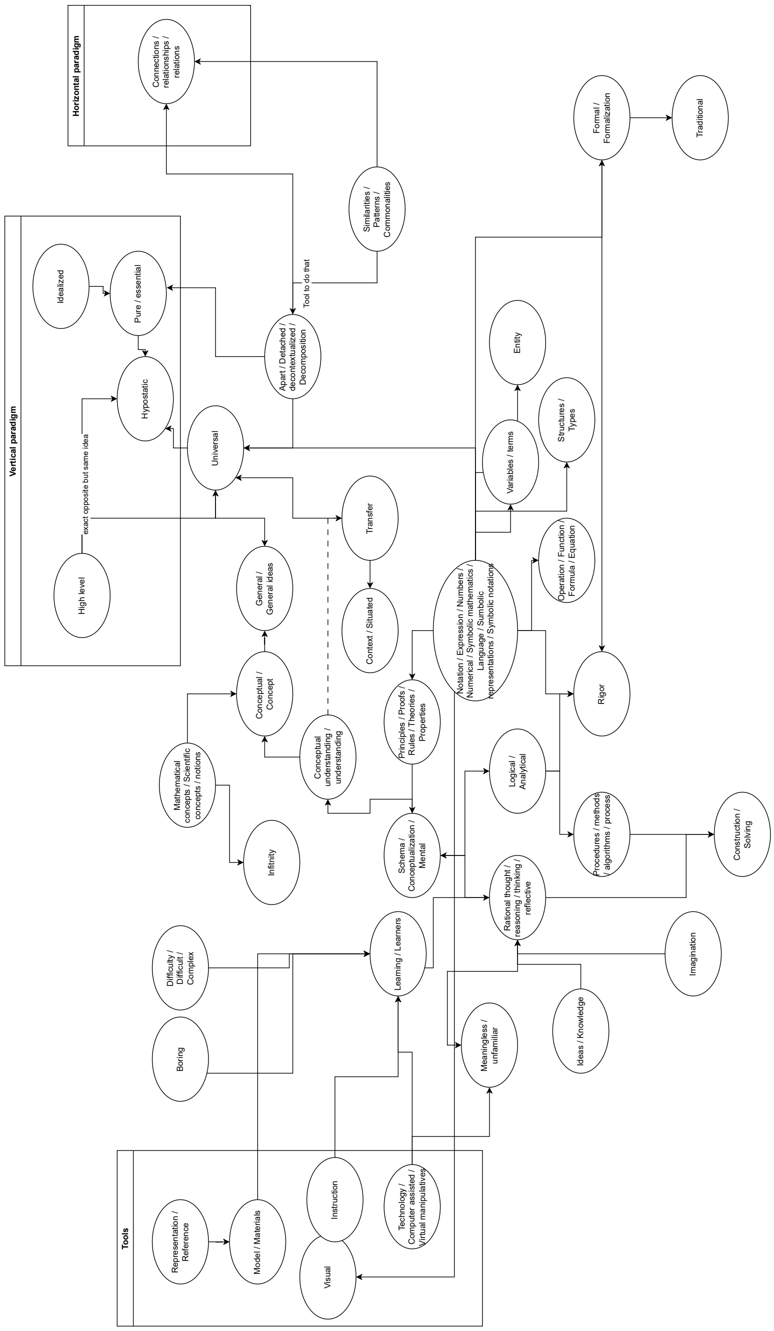

Fig. A3 Digitized mind map for “abstract” after first iteration.

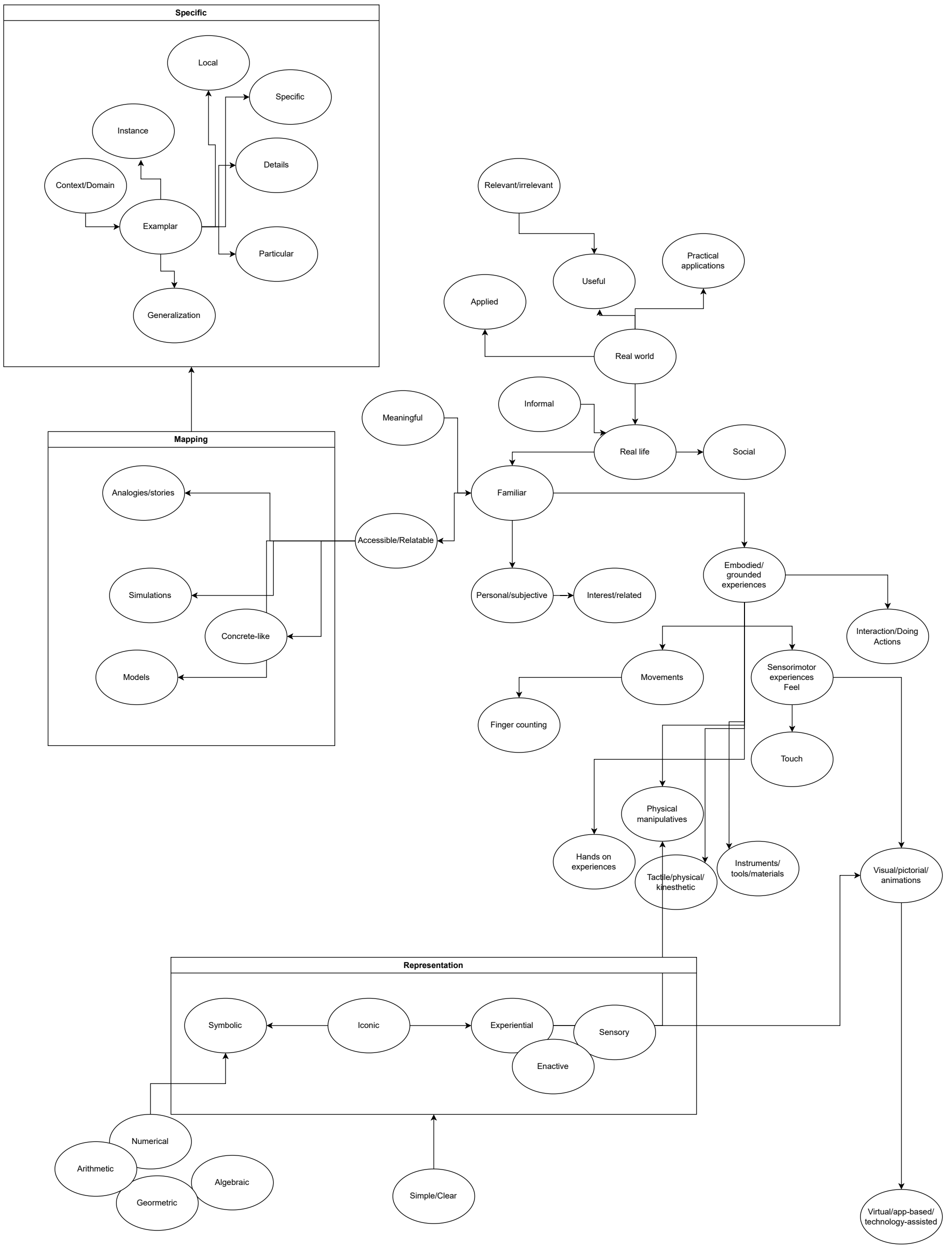

Fig. A4 Digitized mind map for "concrete" after first iteration.

stage, features, diagrams, computations, **bodily experiences**, items, procedure, **understandable**, media. And several new words for **abstract**: geometry, geometrical, **advanced**, **compound**, teacher, tasks, algebraic structures, systems, **online**, **app-based** manipulatives.

R3 and R4 identified the words related to representations that are not yet represented in the taxonomy (in bold above), and integrated them. This step did not change the categories of the taxonomy, but the new words are integrated in the final representation of the taxonomy.

Appendix B Templates

In alignment with our goals, we provide three templates:

- A template to reflect on the relationships between the categories of the taxonomy, to identify trends and gaps in existing research and generate critical question (Figure B9),
- a template to organize existing studies and conditions within our taxonomy (Figure B10),
- and a template to reflect about the concrete and abstract nature of learning materials (Figure B11).

We also attach these templates in full resolution in the Supplementary Materials.

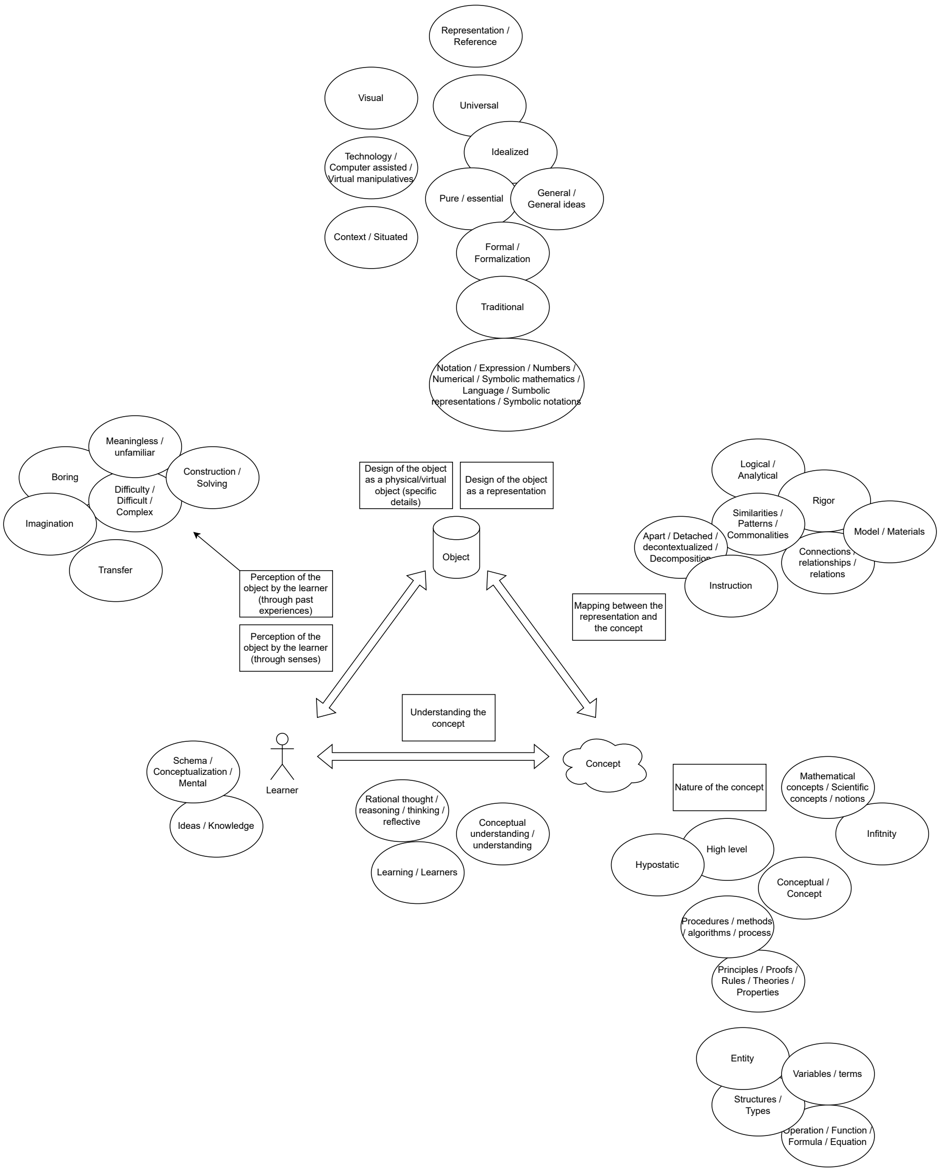

Fig. A5 R3's learner-representation-concept triangle for "abstract" after second iteration.

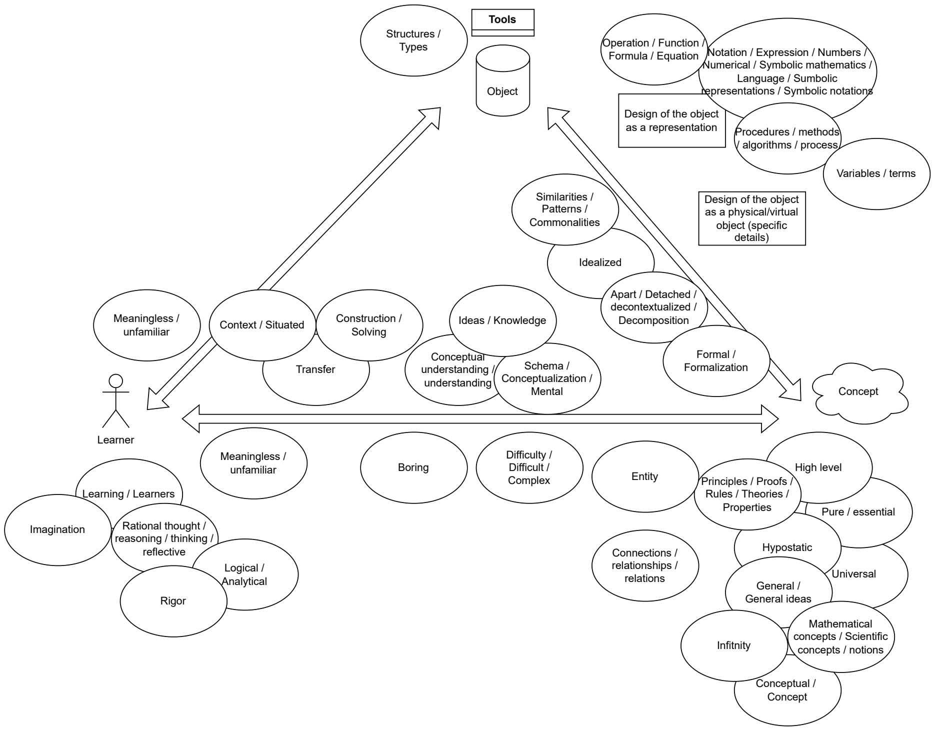

Fig. A6 R4's learner-representation-concept triangle for "abstract" after second iteration.

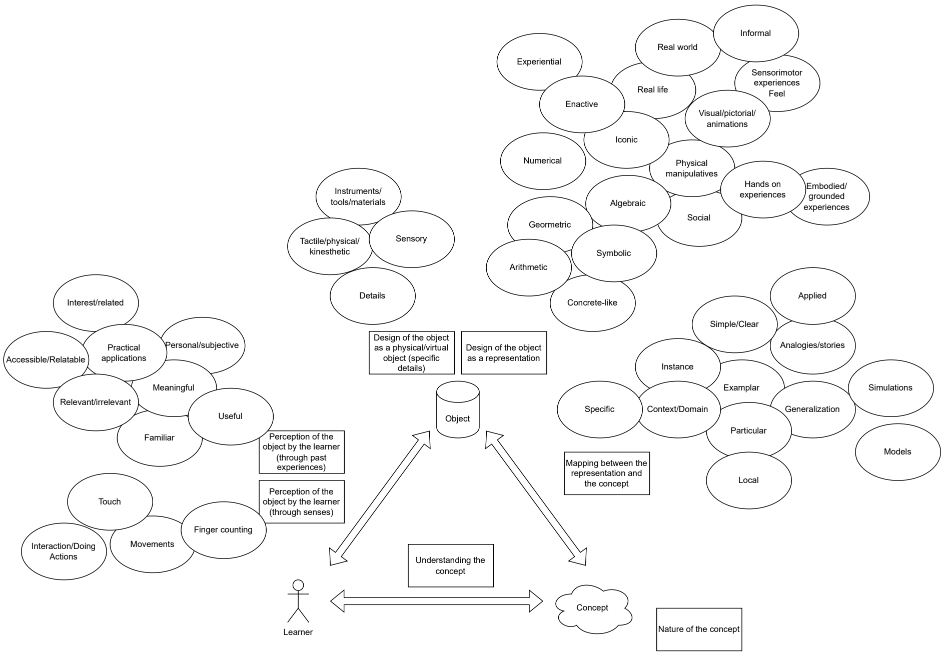

Fig. A7 R3's learner-representation-concept triangle for "concrete" after second iteration.

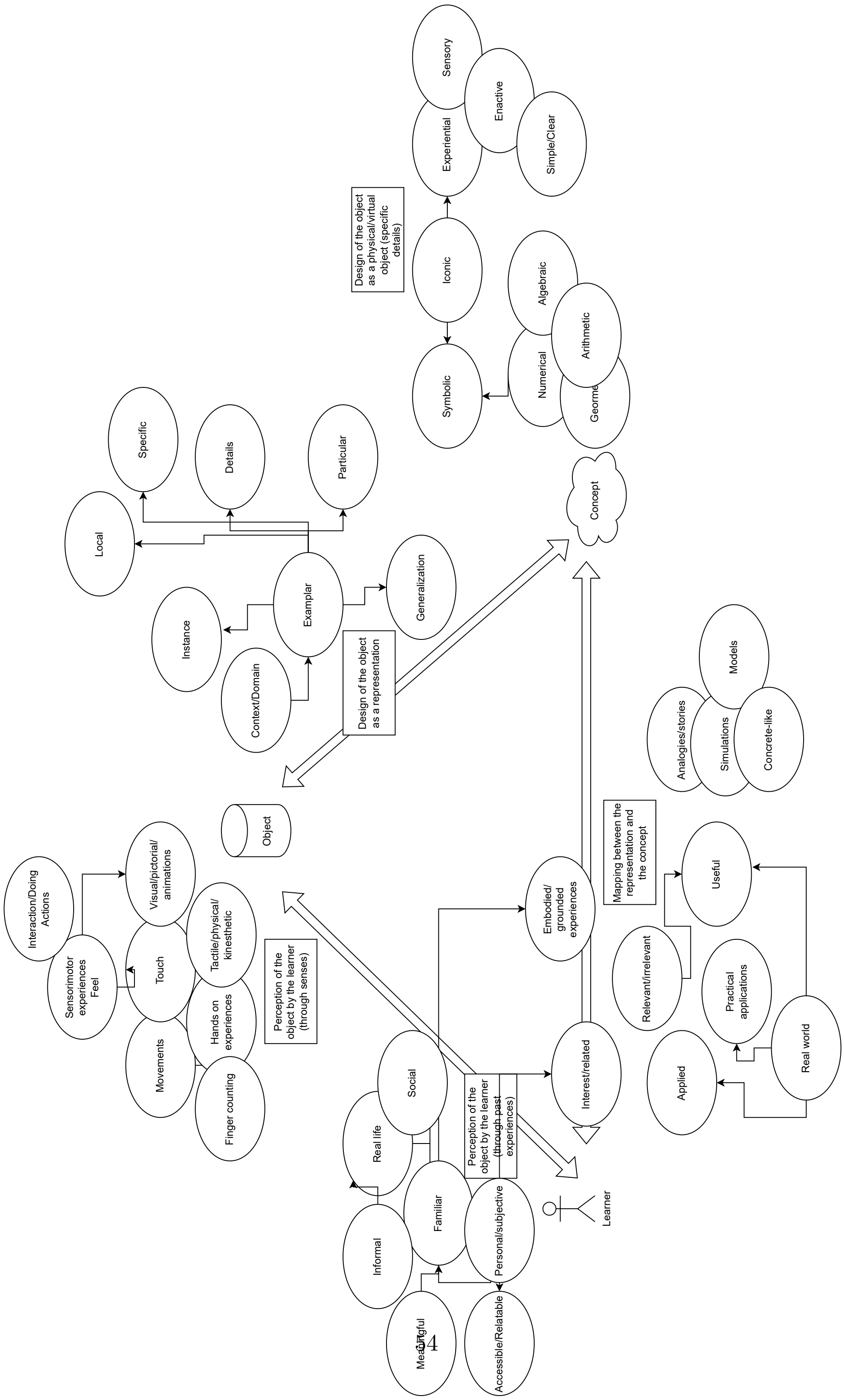

Fig. A8 R4's learner-representation-concept triangle for "concrete" after second iteration.

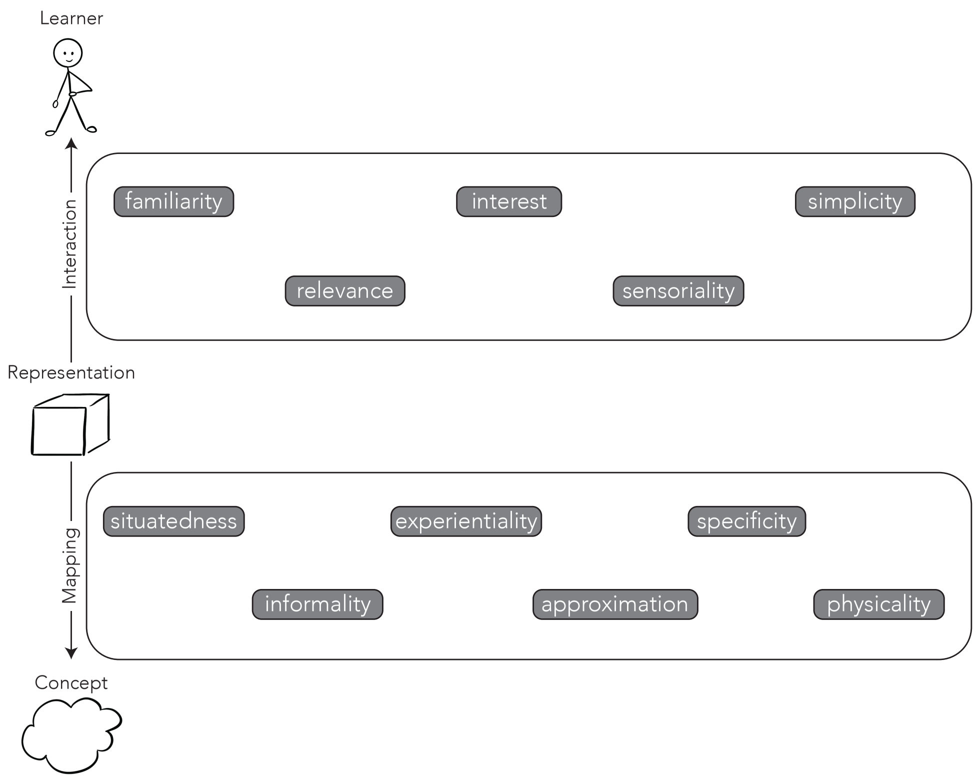

Fig. B9 Template to map relationships between the categories and identify trends, gaps, or critical questions.

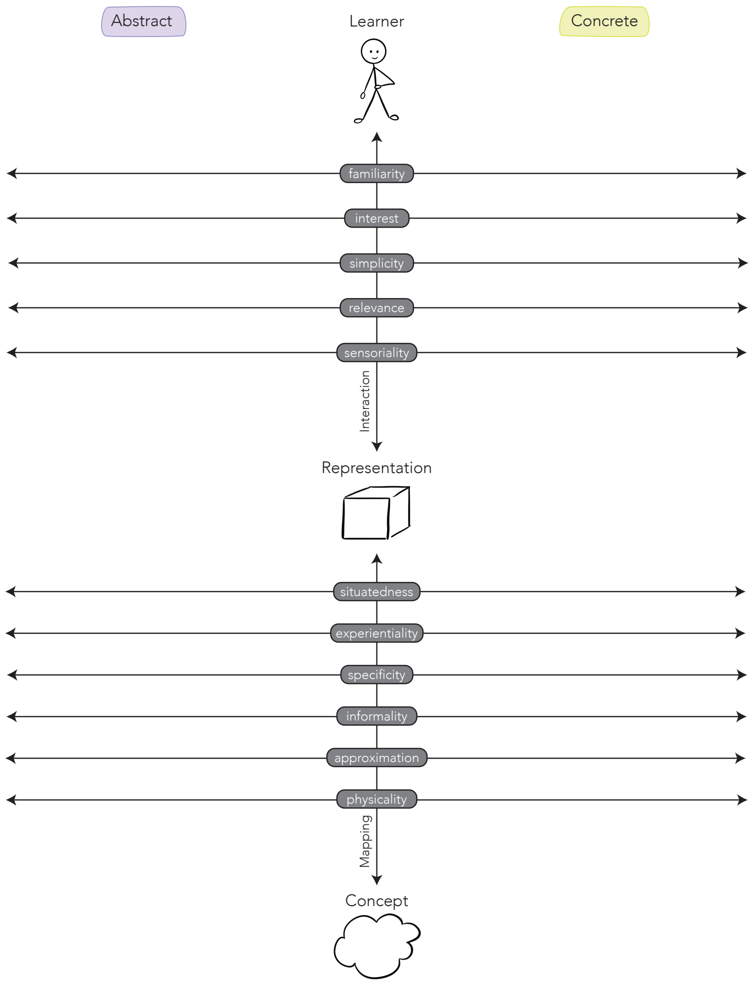

Fig. B10 Template to categorize learning materials according to the taxonomy of abstraction and concreteness.

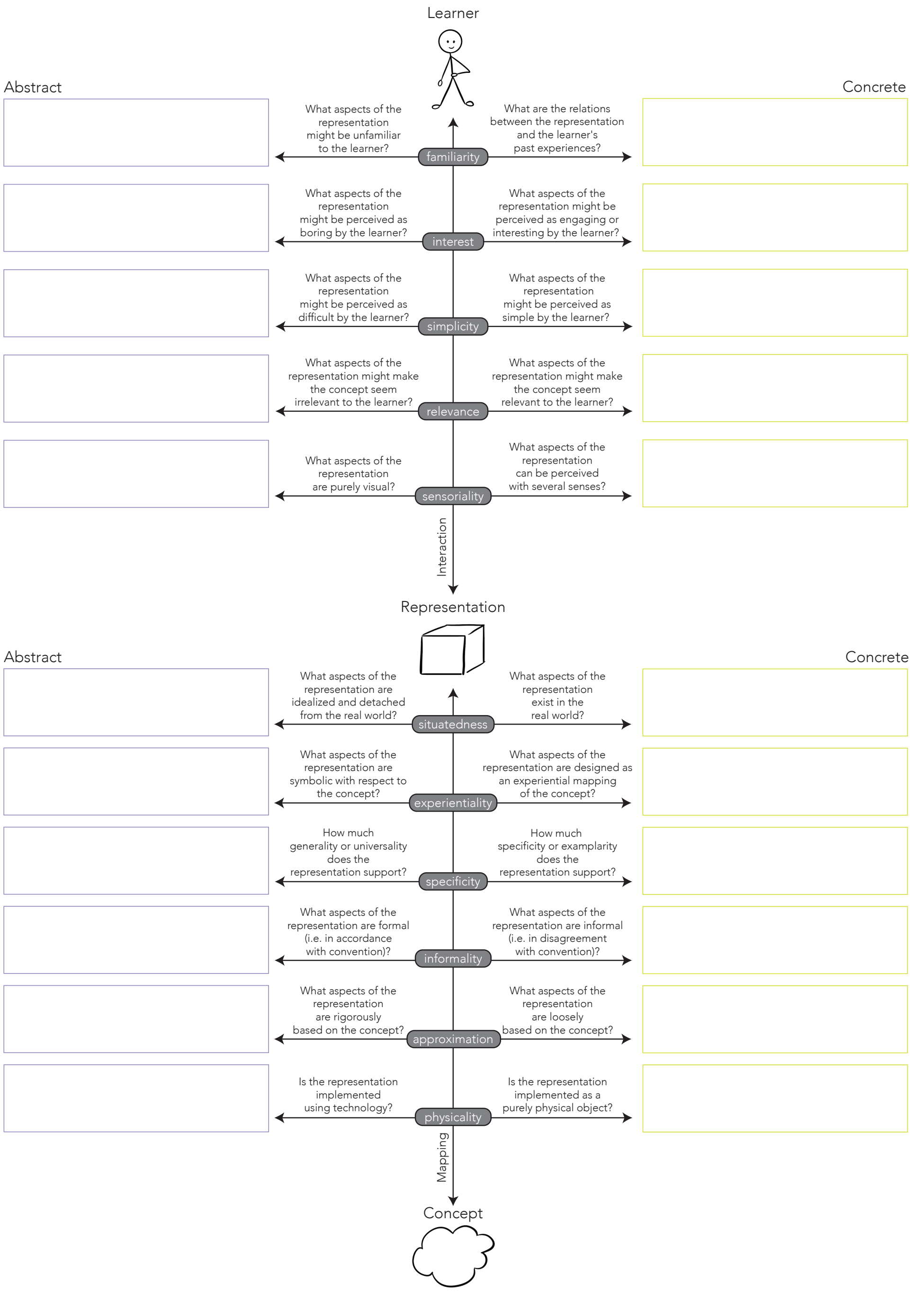

Fig. B11 Template to reflect on learning materials according to the taxonomy of abstraction and concreteness.