The Disappearing "Advantages of Abstract Examples in Learning Math"

Dragan Trninic, Manu Kapur, Tanmay Sinha

ETH Zürich

Abstract

When introducing a novel mathematical idea, should we present learners with abstract or concrete examples of this idea? Considerable efforts have been made over the last decade to settle this question in favor of either abstract or concrete representations. We contribute to this discussion through a critical replication and extension of a well-known study in this area. Whereas the target article argues for the general superiority of abstract representations, we demonstrate that seemingly minor modifications of the study design indicate otherwise. Our results suggest that the previously reported "advantage of abstract examples" manifested not because abstract examples are advantageous in general, but because the earlier studies utilized concrete examples that are pedagogically suboptimal.

Keywords: mathematics education; examples; abstract versus concrete; transfer of learning; replication

Introduction

The use of abstract or concrete representations during mathematics and science instruction has been called a "longstanding controversy" (Fyfe, McNeil, Son, & Goldstone, 2014), and with good reason. Conceptually, we might argue that concrete representations have the advantage of connecting to students' existing knowledge. On the other hand, abstract representations have the advantage of eliminating potentially extraneous perceptual elements. But the elimination of these "extraneous" elements may also reduce the degree to which students can ground a particular representation in their prior knowledge. An advantage of abstract representations is thus at odds with an advantage of concrete representations. Which is better? A review of literature suggests that the answer depends on who asks the question: both pro-concrete and pro-abstract advocates are able to cite research where concrete or abstract representations are more, or less, effective (see, e.g., Koedinger, Alibali, & Nathan, 2008; Schalk, Saalbach, & Stern 2016, for examples).

In this paper, we contribute to the debate through a replication and extension of a well-known and unique study in this area, in particular the central experiment discussed in Kaminski, Sloutsky, and Heckler (2008), "The Advantage of Abstract Examples in Learning Math."

Compared to other papers on the topic, Kaminski et al. is unique in that it makes a universal argument in favor of abstract representations. In particular, the authors argue that "Instantiating an abstract concept in a concrete, contextualized manner… obstructs knowledge transfer. At the same time, learning a generic instantiation allows for transfer" (p. 455). In their study, abstract representations are *in general* superior to concrete representations.

Being a rare mathematics education article published in *Science*, the study caught the attention of not only other scholars, but found recognition in the popular media circuit as well. In a *New York Times* science column, Chang (2008) praised the article, criticized other education researchers for failing to conduct proper research (i.e., "randomized, controlled experiments") and made an even stronger recommendation: "let the apples, oranges and locomotives stay in the real world and… focus on abstract equations." Similar articles appeared in *Le Monde*, *De Standaard*, and elsewhere.

Various elements of the study were criticized over the next few years (see De Bock, Deprez, Dooren, Roelens, & Verschaffel, 2011, for a summary). These criticisms frequently took the form of conceptual disagreements published in math education journals. Despite these conceptual critiques, or perhaps because of them, Kaminski et al. remains steadily cited over the last decade.

The core of the present text is an empirical argument for a more critical re-interpretation of Kaminski et al. In our critical iteration of the experiment, we made relatively minor modifications to the design that nonetheless appear to have had a large impact on the results. We also extended the design to include additional transfer domains, including transfer to a formal mathematical context. The accelerated development of formal knowledge is, after all, a key motivation behind using examples in a math classroom. Our results do not support the hypothesized advantage of abstract examples; on the contrary, they favor the concrete example.

In order to contextualize our own critical replication and extension, we first discuss the central experiment reported by Kaminski et al. (2008), as well as De Bock et al.'s (2011) replication, the first to empirically challenge the original.

Kaminski, Sloutsky, and Heckler (2008)

Kaminski et al. reported a number of experiments drawn from Kaminski's (2006) dissertation. In this paper, we focus on the central experiment, as it forms the foundation of their argument. Here we summarize this experiment, and refer the reader to Kaminski et al. online supplemental materials for a more extensive description.[1]

The experiment consisted of two phases. In the learning phase, undergraduate students (Ohio, USA) were introduced to a mathematical concept, that of an abstract group of order 3 via rules and examples of these rules. (Briefly, an abstract group of order 3 is a set of three elements and a binary operation that satisfies certain abstract rules—closure, associativity, identity, and inverse. As a consequence of these rules, all mathematical groups of order 3 are isomorphic to each other.) The manipulated variable in the study was whether students were introduced to this mathematical concept via more concrete or more abstract representations.

In the concrete representations condition, participants were provided with three icons of a cup—1/3 full, 2/3 full, and 3/3 full—and rules for combining these cups. In the abstract representations condition, participants were provided with three generic shapes—a flag, a square, and a circle—and rules for combining these shapes. Unbeknownst to the participants, adherence to these rules (in either condition) is mathematically equivalent to operating in an abstract group of order 3.

At the end of the learning phase, a multiple-choice test was administered. The second phase—transfer phase—began immediately after completing this test. There, participants were presented with new, seemingly arbitrary icons of real-world objects (e.g., a vase). Unlike in the learning phase, participants received no explicit training in the transfer domain; they were, however, told that these icons combine in ways structurally identical to the rules they just learned, and provided four examples. Then they answered a series of questions structurally identical to the ones they encountered in the learning phase. The training and the tests were accomplished individually via a computer terminal.

Table 1: Average scores (SD), as a percentage. **A** indicates that the learning phase was conducted with abstract instantiations, **C** with concrete ones.

Condition	Learning	Transfer
A (*N* = 18)	80 (13.7)	76 (21.6)
C (*N* = 20)	76 (17.8)	44 (16.0)

See Table 1, above, for a descriptive summary of their results. For now, we note that the "abstract representations" learning condition drastically outperformed the concrete condition on the transfer test: 76% to 44%. This difference is remarkable, all the more so as there were apparently no differences in learning scores or learning times.

[1] The experiments presented are easier to grasp visually. See http://www.sciencemag.org/cgi/content/full/320/5875/454/DC1

De Bock et al. (2011)

In their replication of Kaminski et al., De Bock et al. argued that the transfer domain used by Kaminski et al. is better interpreted as an "abstract transfer" (a terminology we will also use), because it satisfies Kaminski et al.'s own definition of an abstract instantiation. De Bock et al. made the reasonable prediction that, while learning with abstract instantiations may transfer better to an abstract domain, concrete instantiations may transfer better to a concrete domain.

To test this hypothesis, undergraduate students (Belgium) were randomly assigned to one of four conditions:

- AA, abstract learning then abstract transfer
- AC, abstract learning then concrete transfer
- CA, concrete learning then abstract transfer
- CC, concrete learning then concrete transfer.

That is, De Bock et al. kept the two-phase format of the original study, but expanded it to include a transfer to a more concrete domain.

For abstract and concrete learning, and abstract transfer, De Bock et al. used identical materials to Kaminski et al. For concrete transfer, they repurposed one of the alternate concrete learning conditions in the original study—that of a pizza divided in thirds.

Table 2: Average scores (SD), as a percentage. See text, above, for a description of the four conditions.

Condition	Learning	Transfer
AA (*N* = 23)	71 (16.3)	75 (15.8)
AC (*N* = 30)	64 (14.6)	73 (17.5)
CA (*N* = 28)	77 (12.1)	50 (17.9)
CC (*N* = 24)	76 (14.6)	84 (10.0)

See Table 2, above, for a descriptive summary. In brief, the results confirmed the original findings, as well as De Bock's own hypothesis. In their words: "if transfer to a new abstract domain is targeted, abstract instantiations are indeed more advantageous than concrete instantiations" (p. 120). They continue, "However… the opposite holds as well: Transfer to a new concrete domain is more enhanced by a concrete learning domain than by an abstract one" (p. 120). While not contradicting the original study, De Bock et al. demonstrated that there is more there than meets the eye.

Present Study

We questioned whether the observed "advantage of abstract examples" was due—at least in part—to certain pedagogically suboptimal aspects of the design, which we detail below. To the best of our knowledge, neither De Bock et al., nor anyone else using Kaminski et al.'s materials (e.g., Kaminski, Sloutsky, & Hecker, 2013; McNeil & Fyfe, 2012), attempted to *improve* the materials (as a teacher might). In addition to these pedagogical modification, we also extended the study beyond the original transfer task, as detailed below.

Design modifications and justifications

We identified two aspects of the original materials for improvement. First, the concrete representations used in the main study—those of 1/3, 2/3, and 3/3 liquid-filled cups[2]—caught our attention. The cover story for this instantiation involved combining two or more cups, and trying to determine the "left-over." For instance, 2/3 + 2/3 = 1/3 left-over. Why did Kaminski et al. use the full cup (3/3) as the identity element? The authors presumably used this scheme because it matches our everyday intuition that 1/3 + 2/3 = 3/3. There are at least two issues with this. First, it leads to unintuitive calculations, such as 3/3 + 3/3 = 3/3. More critically, 1/3 + 2/3 = 3/3 is precisely the wrong intuition for mod 3 arithmetic (arithmetic of groups of order 3), because *there* 1 + 2 does not equal 3, but 0.

To us, this suggested that Kaminski's study was not optimized for learning in the concrete condition. An introductory example, we hold, should align not mismatch the superficial concrete elements with the target mathematical structure. Consequently, the concrete representations of cups filled with varying quantities of liquid were modified from 1/3, 2/3, and a full cup (3/3) to 1/3, 2/3 and an empty cup (0/3). This leads to initially surprising but more structurally appropriate 1/3 + 2/3 = 0/3.

Our second concern had to do with the "cover stories" for each of the instantiations. Across these, participants were put into drastically different roles, some believable, others not. These cover stories are as follows (drawn from Kaminski et al. supplementary materials, and Kaminski, 2006):

Abstract instantiation: an archeologist trying to make sense of symbolic combinations left by an ancient civilization.

Concrete (main): an employee at a detergent company calculating the left-over after quantities of liquid are combined.

Concrete (alternative): a pizzeria owner discussing the chef who systemically and persistently burns predetermined portions of every pizza.

Concrete (alternative): an employee at a tennis ball factory dealing with malfunctioning machines producing incorrect quantities of balls.

Transfer: an anthropologist trying to understand a "children's game from another country."

While university students are surely capable of handling nonsense cover stories, such as the one where "the cook systematically burns a portion of each group order," we had concerns about their uneven, varying quality. Specifically, we felt that—pedagogically speaking—the concrete instantiations cover stories were poor in quality, while the generic and transfer narratives impressed us as reasonable. We conjectured that this matters, because a "reasonable" story may be more likely to connect to and activate relevant prior knowledge *without* also being overly distracting. In contrast, a cover story concerning a pizzeria where "the cook systematically burns a portion of each group order" is at odds with any prior knowledge one might have concerning pizzerias, cooking, or business profitability.

A closely related concern has to do with our general sense that the framing of the generic instantiation (an archeological discovery) and the transfer instantiation (a game from another country) had more to do with each other than the concrete instantiations (all of which had to do with odd work).

In response, we made the following modification to the study: every cover story was changed to "a children's game from another country." We generally accept that children play all kinds of games, and recognize that games can involve more concrete instantiations (e.g., combining cups of liquid), or more abstract instantiations (e.g., combining symbols). In other words, this particular cover story was chosen because it naturally accommodates concrete as well as abstract representations.

Because the students in our study would be asked to solve multiple transfer tests rather than one, a compromise was made to remove 4 items from the multiple-choice tests (same 4 from each test); this reduced the number of items on each of the tests from 24 to 20. Specifically, the items removed were 5, 8, 13, and 17 from the original abstract learning instantiation, and all the corresponding items from the other tests. (Of those, items 5 and 8 were chosen for elimination because they were basic and replicated across other questions. Items 13 and 17 were chosen because they used noticeably more text than the other items, a pattern we worried would become apparent across the phases.)

In addition to these modifications to the original study, we extended the study by introducing two additional transfer phases. Similar to De Bock et al.'s study, and for the same reason, we employed a concrete transfer task structurally identical to the original abstract transfer task. While De Bock et al. repurposed the alternative pizza concrete instantiation for this phase, we repurposed the tennis ball factory concrete instantiation.

Finally, we introduced a formal transfer phase, a group of order 5 and consisting of 0, 1, 2, 3, and 4. That is, addition mod 5, formally presented, where 2 + 2 = 4, yet 4 + 2 = 1, 4 + 3 = 2, and so on. We introduced this transfer test to evaluate a particular claim by Kaminski et al., namely that abstract representations lead to superior transfer because they support

[2] Again, we invite the reader to consult the supplementary online materials from the original study. These can be found at: http://www.sciencemag.org/cgi/content/full/320/5875/454/DC1

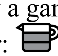
a deeper understanding of underlying mathematics. But, if students indeed developed a "deep understanding" of groups (or, at least, modular arithmetic), then it stands to reason that they should be able to transfer this knowledge to a formal instantiation of a group of order 5, which shares many similarities with a group of order 3. In this phase, just as in the other transfer phases, participants were not explicitly instructed on the rules, but provided with a few examples and told that the rules of this system are similar to the rules of the previous systems. This phase contained only 11 multiple choice items, focusing on deeper understanding of underlying principles, for example each element having an inverse.

Method

Undergraduate students attending a public university in Switzerland were randomly assigned to one of the (concrete or abstract) learning conditions (38 each; a priori power analysis informed by the original studies indicated that this number was sufficient).[3] Students majoring in mathematics or a computer science field were excluded. Training included explicit training in the rules, with accompanying examples (as in the original study). After this learning phase was completed, participants completed three more transfer phases in the following order: abstract, concrete, and formal. Transfer phases did not include explicit training, but did provide a few examples and inform participants that the rules are "the same" as in the previous tasks (again, as in the original study). To (partially) account for order effects, half the participants in each condition instead completed the phases in the following order: learning, concrete, abstract, formal (formal transfer was always last). No order effects were observed, and the orders are combined for this analysis.

As an illustration, this is what the cover stories and representations looked like for each phase:

Learning, abstract:
In another country, children play a game that involves three symbols: , , and .

Learning, concrete:
In another country, children play a game by combining cups with different quantities of water: , , and .

Transfer, abstract:
In another country, children play a game that involves these three objects:

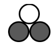
 (a ladybug)

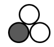
 (a vase)

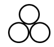
 (and a book).

[3] A third condition, a modification of the abstract learning instantiation, was also investigated in the study. As it has no bearing on our current discussion, it is omitted from the analysis.

Transfer, concrete:
In another country, children play a game that involves these three objects:

 - a container with two tennis balls

 - a container with one tennis ball

 - a container with zero tennis balls.

The final phase, formal transfer, did not use a cover story. There, participants were told that they will work with "a number system" and provided with examples of that system.

As in the original study and De Bock's replication, the study was completed individually, on a computer terminal, and there were no breaks during the study. The majority of participants completed the study within an hour, with no one taking more than 75 minutes. The study was conducted by assistants blind to the study expectations.

Reliability analysis for the learning, abstract transfer, concrete transfer, and formal transfer tests yielded McDonald's ω of 0.893, 0.857, 0.888, and 0.856, respectively.

Analysis

No participants were excluded from our analysis. The significance of this is addressed in the Discussion.

For inferential tests (JASP, 2018), Mann-Whitney U test was used as the data were not normally distributed. As is commonplace in education research, we report Cohen's *d*; however, we prioritize the rank-biserial correlation r_B as a more appropriate, unbiased effect size measure.

Results

Table 3, below, provides descriptive statistics for the present study. There were no significant differences in time for completion.

Table 3: Average scores (SD), as a percentage. **A** indicates that the learning phase was conducted with abstract instantiations, **C** with concrete ones. (Note that "Transfer Abstract" in this study corresponds to "Transfer" in previous studies reported in Table 1 and Table 2.)

	Learning	Transfer Abstract	Transfer Concrete	Transfer Formal
A (N = 38)	70 (24.8)	78 (18.3)	90 (14.3)	70 (24.9)
C (N = 38)	95 (12.1)	73 (25.9)	95 (10.7)	78 (27.7)

Comparing concrete to abstract learning conditions, we found a significant difference between the learning scores in favor of the concrete learning condition, Mann-Whitney U = 1167.5, $p < .001$, rank-biserial correlation $r_B = 0.617$ with 95% CI [.429, .754] (Cohen's d = 1.276). On abstract transfer, we found no difference on performance, U = 701.5, $p = 0.835$, and a very small effect size, $r_B = -0.028$ with 95% CI [-.282, .229] (Cohen's d = -0.176). On concrete transfer, we found evidence in favor of the concrete condition, U = 913.5, $p = .033$, and a small-to-moderate effect, $r_B = 0.265$ with 95% CI [.010, .488] (Cohen's d = 0.396). Finally, formal transfer favored the concrete condition, but this difference was not significant, U = 877, $p = .103$, $r_B = 0.215$ with 95% CI [-.043, .446] (Cohen's d = 0.304).

To check for the influence of outliers, we excluded all participants who scored more than two standard deviations from the mean on any of the tests (same criterion used by Kaminski et al. and de Bock et al.). Three participants were excluded from each condition. Two results were affected. First, the difference on concrete transfer changed from significant to trending in favor of the concrete learning condition, U = 758.5, $p = .064$. Second, the differences on formal transfer reached significance, U = 777, $p = .048$, and a small-to-moderate effect in favor of the concrete learning condition, $r_B = 0.269$ with 95% CI [0.003, 0.499] (Cohen's d = 0.408).

Discussion

We aimed to critically replicate and extend an influential study that argued for the advantage of abstract representations in learning mathematics. We made two modification to the original study: (1) using an icon of an empty cup rather than a full cup in the concrete learning condition, and (2) keeping the "cover stories" similar to each other across the tasks. These modifications were made with the intent of removing pedagogically suboptimal elements present in the original design. We also extended the study by included a more concrete transfer task and a formal transfer task. Overall, our results put into question the previously reported advantage of abstract examples.

Whereas Kaminski et al. found no difference in the learning scores, and De Bock's study found a small difference in favor of the concrete instantiation, we found a significant and very large effect in favor of the concrete instantiation. How is it that the concrete instantiation condition in our study performed much higher than participants in the original, and even De Bock's study, on *both* learning and abstract transfer? In the original study, concrete learning to abstract transfer showed 44%, compared to 76% for abstract learning to abstract transfer. In De Bock's study, students fared slightly better, at 50% vs. 75%. In the present study: 73% vs. 78%.

We briefly entertained the (surely self-satisfying) notion that our students are more capable. However, this explanation is unlikely, because our students scored comparatively similar on the other comparable tests, for example across the *abstract* learning condition to abstract transfer (Kaminski: 80%, De Bock: 75%, present study: 78%). This suggests that the concrete instantiation condition performed better because of the changes made to the original materials. But those changes, as detailed earlier, were minor. Of these, we conjecture that using an empty cup rather than a full one may have made the largest difference, as this modification better aligned the concrete representation with the underlying mathematical notion.

As with De Bock et al., we found evidence in favor of the concrete instantiation on the concrete transfer test, although in our case this evidence was not robust.

Furthermore, once outliers were removed, we found evidence in favor of concrete instantiations on the formal transfer test, as well.

An additional point on data analysis may be worth considering. When analyzing our data, we chose to conduct analysis on all the participants, and again after removing those participants scoring more than two standard deviations from the mean. In contrast, the results reported by Kaminski et al. and De Bock et al. (the later following the former), were performed *after* eliminating participants who scored below chance on the learning test, for "failing to learn" (as well as removing the outliers, as we did). This is an unusual method of removing participants in an educational study, and one not conceptually justified in previous articles. Note that it biases the results in favor of students who found the materials useful in the first place. This is an artificial restriction—imagine a mathematics professor evaluating her teaching but refusing to consider those students who "failed to learn" from her lectures, as determined by a learning test immediately following the lecture.

In our data, this "failure to learn" elimination favored the abstract learning condition, because only in that condition did the students score below chance on the learning test. It did not favor it enough to impact the results, but it suggests that this particular elimination introduces bias in favor of the abstract instantiation. This does not explain the drastic differences between our results and those of previous studies, but it raises a question as to why this particular method was employed in the first place. After all, we researchers are unlikely to eliminate data that favors our predictions.

Limitations

Because our design makes not one but multiple modifications to the original study, further work is needed to identify the impact of each modification, as well as to investigate the potential mechanisms through which these modifications influence the learning process.

Summary and Implications

We made a relatively minor change to the concrete learning instantiation in Kaminski et al., in addition to making the various "cover stories" similar to each other. In turn, we observed results that contradict Kaminski et al., and partially support De Bock et al.

Overall, our findings suggest that, if only one instantiation is to be used, and *for these types of tasks*:

Concrete representations facilitate the initial learning of a mathematical concept better than abstract representations of the same idea.

On an abstract transfer, there is no notable advantage between learning via concrete or abstract representations.

On a concrete transfer, learning via concrete representations is preferable, although this difference is relatively small.

When transferring to a formal domain, learning via concrete representations may be preferable, although this difference is, again, relatively small.

Can the current study make any pedagogical recommendations? De Bock et al. (2011) and Jones (2009) caution, and we concur, that brief interventions of this sort should not be applied directly and uncritically to mathematics classrooms. Seen from that perspective, this study claims no more than the following: concrete instantiations may be more or less useful, depending on their quality and context. To be clear, we do not advocate concrete examples as universally advantageous. We agree with Lampinen and McClelland (2018), who argue that it is not the static qualities of "abstractness" or "concreteness" that are likely to impact learning; rather, learning depends on the *interactive* aspects of the learning environment (see Abrahamson & Trninic, 2015). As such, the existence of universally "ideal" learning examples seems unlikely.

The scholarly value of this study lies instead in its contrast to previous work, which found a significant and large effect in favor of the abstract learning instantiation. Our results provide an alternative explanation for those earlier findings. The "advantages of abstract examples" of Kaminski et al. did not manifest because "abstract examples" are better in general. It was because, in that particular design, the concrete learning condition was suboptimal.

References

Abrahamson, D., & Trninic, D. (2015). Bringing forth mathematical concepts: Signifying sensorimotor enactment in fields of promoted action. In D. Reid, L. Brown, A. Coles, & M.-D. Lozano (Eds.), Enactivist methodology in mathematics education research [Special issue]. *ZDM Mathematics Education, 47*(2), 295–306.

Chang, K. (2008, April 25). Study suggests math teachers scrap balls and slices. *The New York Times.* Retrieved January 22, 2019, from http://www.nytimes.com/2008/04/25/science/25math.html

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.

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, 1*(17).

JASP Team (2018). JASP (Version 0.9)[Computer software].

Jones, M. G. (2009). Examining surface features in context. *Journal for Research in Mathematics Education, 40*, 94–96.

Kaminski, J. A. (2006). The effects of concreteness on learning, transfer, and representation of mathematical concepts. (Doctoral dissertation, Ohio State University, 2006). *Dissertation Abstracts International*.

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

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008b). Supporting online material for *The advantage of abstract examples in learning math*. Retrieved January 20, 2019, from http://www.sciencemag.org/cgi/content/full/320/5875/454

Kaminski, J. A., & Sloutsky, V. M. (2013). Extraneous perceptual information interferes with children's acquisition of mathematical knowledge. *Journal of Educational Psychology, 105*(2) 351–363.

Koedinger, K. R., Alibali, M. W., & Nathan, M. J. (2008). Trade-offs between grounded and abstract representations: Evidence from algebra problem solving. *Cognitive Science, 32*, 366–397.

Lampinen, A. K., & McClelland, J. L. (2018). Different presentations of a mathematical concept can support learning in complementary ways. *Journal of Educational Psychology, 110*(5), 664–682.

McNeil, N. M., & Fyfe, E. R. (2012). "Concreteness fading" promotes transfer of mathematical knowledge. *Learning and Instruction, 22*(6), 440–448.

Schalk L., Saalbach H., & Stern E. (2016). Approaches to Foster Transfer of Formal Principles: Which Route to Take? *PLoS ONE, 11*(2): e0148787. https://doi.org/10.1371/journal.pone.0148787