COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 44 (2020) e12851

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12851

The Disappearing "Advantage of Abstract Examples in Learning Math"

Dragan Trninic, Manu Kapur, Tanmay Sinha

Chair of Learning Sciences and Higher Education, ETH Zürich

Received 17 July 2019; received in revised form 7 March 2020; accepted 9 March 2020

Abstract

When teaching a novel mathematical concept, should we present learners with abstract or concrete examples? In this experiment, we conduct a critical replication and extension of a well-known study that argued for the general advantage of abstract examples (Kaminski, Sloutsky, & Heckler, 2008a). We demonstrate that theoretically motivated yet minor modifications of the learning design put this argument in question. A key finding from this study is that participants who trained with improved concrete examples performed as well as, or better than, participants who trained with abstract examples. We argue that the previously reported "advantage of abstract examples" manifested not because abstract examples are advantageous in general, but because the concrete condition employed suboptimal examples.

Keywords: Mathematics education; Abstract versus concrete; Replication; Examples; Transfer

1. Introduction

When introducing a mathematical idea, should we use abstract or concrete examples? The debate over abstract and concrete examples in mathematics and science education has been called a "longstanding controversy" (Fyfe, McNeil, Son, & Goldstone, 2014), with good reason. For one, the empirical evidence is inconclusive. Both sides are able to cite studies where abstract or concrete instantiations are more, or less, effective (e.g., Koedinger, Alibali, & Nathan, 2008; Schalk, Saalbach, & Stern, 2016).

Conceptual arguments can be made in favor of either approach. In this brief report, we focus on one apparent tradeoff identified in the literature—that is, *connection* and *distraction*. Concrete examples may support initial learning by activating real-world knowledge and *connecting* to prior experience (Fyfe et al., 2014; Gravemeijer, 1994; Kotovsky,

Correspondence should be sent to Dragan Trninic, Clausiusstrasse 59, RZ J5, 8092 Zurich, Switzerland. E-mail:trninicd@ethz.ch

Hayes, & Simon, 1985). However, they may also introduce features that *distract* from the targeted concepts (Belenky & Schalk, 2014; Kaminski & Sloutsky, 2013; Uttal, Scudder, & DeLoache, 1997). Of course, concrete examples do not equally connect or distract. Well-designed examples activate relevant connections while minimizing distractions, whereas suboptimal examples distract the learner without activating relevant connections. This issue of tradeoffs is relevant when considering transfer. While there are multiple accounts of knowledge transfer (e.g., Greeno, Moore, & Smith, 1993; Holyoak & Koh, 1987; Thorndike & Woodworth, 1901), all share an understanding that transfer is significantly influenced by similarities between training and transfer situations. This implies that there is no single, ideal training situation, and it supports the common approach to teaching mathematics by introducing multiple examples (e.g., Gravemeijer, 1994). By presenting learners with various instantiations of the same underlying concept, we increase their chance to transfer this knowledge to novel situations (also see Gick & Holyoak, 1983).

A series of articles by Kaminski and Sloutsky (2013) and Kaminski, Sloutsky, and Heckler (2008a, 2009, 2013) challenged the assumptions that (a) transfer depends on both learning and transfer domains, and that (b) example selection involves tradeoffs. In their central experiment, Kaminski and colleagues introduced a mathematical concept to students via either more concrete instantiations (measuring cups) or more abstract instantiations (geometric shapes). On a transfer test, the abstract condition significantly and substantially outperformed the concrete condition. Remarkably, this held true even when multiple concrete or mixed instantiations were compared to a single abstract one. The authors concluded that "Instantiating an abstract concept in a concrete, contextualized manner ... obstructs knowledge transfer. At the same time, learning a generic instantiation allows for transfer" (Kaminski et al., 2008a, p. 455). Their argument could be summarized as follows: (a) Transfer is primarily a function of the learning domain; in turn, (b) tradeoffs are overstated, because abstract examples are unambiguously advantageous.

This challenge was made most prominently in "The Advantage of Abstract Examples in Learning Math" (Kaminski et al., 2008a). Being a rare mathematics education article to appear in *Science*, it caught the attention of not only scholars, but also pundits across the popular media circuit. *The New York Times* praised the study, making an even stronger recommendation: "let the apples, oranges and locomotives stay in the real world ... focus on abstract equations" (Chang, 2008). Similar articles appeared in *De Standaard*, *Le Monde*, and other prominent newspapers, demonstrating Kaminski et al.'s extraordinary reach beyond academia. Over the next few years, various aspects of Kaminski et al. were criticized conceptually (see De Bock, Deprez, Van Dooren, Roelens, & Verschaffel, 2011, for a summary). To date, however, attempts to empirically interrogate the basic finding (i.e., "the advantage of abstract examples") have been rare.

The core of this paper is an empirical argument for a reevaluation of Kaminski et al. (2008a). A basic assumption of such experimental comparisons is that the conditions under comparison are similarly well-designed. We intend to argue that the original concrete condition employed suboptimal examples. To the best of our knowledge, neither De Bock et al. (2011) nor anyone else using Kaminski et al.'s materials (e.g., McNeil &

Fyfe, 2012; Siler & Willows, 2014) attempted to *improve* the concrete training examples in a way that strengthens connections and reduces distractions. We demonstrate that this minor but theoretically motivated modification of the learning design can substantially impact the results. Beyond this, the study was extended to include additional transfer domains. Our results do not support the hypothesized advantage of abstract examples. On the contrary, the participants in the modified concrete condition performed better overall than the participants in the abstract condition.

To contextualize our critical replication and extension, we first discuss the central experiment reported by Kaminski et al. (2008a), as well as De Bock et al.'s (2011) replication.

2. Target experiment

2.1. Kaminski, Sloutsky, and Heckler (2008)

The original experiment consisted of two phases. In the training phase, undergraduate students (U.S.) were introduced to certain procedural rules of an abstract group[1] of order 3. Random assignment determined whether the students were introduced to these rules via more concrete or more abstract examples. In the concrete condition, participants were provided with three cup icons—1/3 full, 2/3 full, and 3/3 full—and rules for combining the icons. In the abstract condition, participants were provided with three generic shapes—a flag, a square, and a circle—and rules for combining the shapes. Unbeknownst to the participants, adherence to these rules, in either condition, is mathematically equivalent to operating in a mathematical group of order 3.

At the end of the training phase, a multiple-choice initial learning test was administered. The transfer phase began immediately afterward. In the transfer phase, participants were presented with a mystery game utilizing three stylized instantiations of real-world objects (e.g., a vase). Participants received no training in the transfer domain, but were provided with four examples and explicitly told that "the rules of the last system are like the rules of this game." Finally, participants answered a series of multiple-choice questions similar to the ones they encountered in the learning phase. Transfer performance could be attributed to learning since, without training, the baseline performance on the multiple-choice transfer test was no better than chance. The entirety of the experiment was accomplished individually at a computer terminal. Participants proceeded at their own pace.

See Table 1 for a descriptive summary of the results. Before discussing the results, we note that this study only reported results after two levels of participant elimination. We consider this a significant point for discussion and return to it later in the report. The difference between conditions during initial learning was not significant. Despite this, the abstract condition significantly and substantially outperformed the concrete condition on the transfer test. In particular, note the large effect size, Cohen's $d > 1.6$, favoring the abstract training conditions.

Table 1
Average scores (and *SD*) as a percentage (Kaminski et al., 2008a)

Condition	Initial Learning	Transfer
Abstract (n = 18)	80 (13.7)	76 (21.6)
Concrete (n = 20)	76 (17.8)	44 (16.0)

2.2. De Bock, Deprez, Dooren, Roelens, and Verschaffel (2011) replication

Inspired by conceptual critiques of the original study (e.g., Jones, 2009a, 2009b), De Bock et al. (2011) argued that the mystery game transfer domain used by Kaminski et al. (2008a) is better interpreted as an *abstract* transfer domain, even by Kaminski et al.'s definition. De Bock et al. made the prediction that, while the abstract training condition may promote transfer to an *abstract* domain like the one used in the original study, the concrete condition may transfer better to a concrete domain.

To test this hypothesis, De Bock et al. (2011) kept the materials identical and used the two-phase format of the original study, but expanded it to include transfer to a more concrete domain. For concrete transfer, they repurposed one of the alternative concrete instantiations from the original study, that of a partitioned pizza. Undergraduate students (Belgium) were randomly assigned to one of four conditions: AA, abstract examples training then abstract domain transfer; AC, abstract training then concrete transfer; CA, concrete training then abstract transfer; and CC, concrete training then concrete transfer. See Table 2 for a descriptive summary of the results. Following Kaminski et al. (2008a), this study only reported results after two levels of participant elimination.

The results supported the original findings, but also the authors' prediction. In their words: "If transfer to a new abstract domain is targeted, abstract instantiations are indeed more advantageous than concrete instantiations. However … Transfer to a new concrete domain is more enhanced by a concrete learning domain than by an abstract one" (De Bock et al., 2011, p. 120).

Table 2
Average scores (and *SD*) as a percentage (De Bock et al., 2011)

Condition	Initial 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)

3. Present study

Our interest in replicating and extending Kaminski et al. (2008a) was driven by our analysis of their findings and learning design. We were surprised that, in that study, the students in either condition "successfully learned the material with no differences in learning scores" (p. 455). In turn, we wondered whether the observed advantage of abstract examples on transfer was due to, in part, suboptimal yet overlooked aspects of the concrete training condition. We identified two potentially suboptimal elements in the original design. First, the concrete example "cover stories" do not meaningfully connect to relevant prior knowledge and distract with unreasonable scenarios. Second, the concrete instantiation of the identity element distracts with nonintuitive use of everyday terms and, moreover, elicits inappropriate intuitions of mod 3 arithmetic[2]. In other words, regarding both connection and distraction, the setup of the concrete condition was suboptimal. Our study was designed to remove these suboptimalities, as we now discuss.

Our first concern had to do with the "cover stories." Across these, participants were put into drastically different situations, some believable, others not. These cover stories are as follows (from Kaminski, 2006; Kaminski et al., 2008b, supporting online materials):

1. Concrete instantiation (main): Chief engineer at a detergent company asks for help determining the amount of left-over liquid after mixing solutions.
2. Concrete (alternative): Pizza aficionado asks for help determining how much pizza will be burnt by a chef who burns predetermined portions of every order in a systematic way.
3. Concrete (alternative): Factory owner asks for help determining the quantity of extra balls after combining partially filled tennis ball containers.
4. Abstract instantiation: Archeologist explains symbolic combinations left by an ancient civilization.
5. Transfer: Cultural studies professor asks for help figuring out a children's game from another country.

It is known that difficulties arise when real-world scenarios are used in impractical, implausible ways to generate cover stories for math problems (e.g., Inoue, 2005; Lave, 1992; Palm, 2008).[3] We offer that the concrete cover stories used by Kaminski et al. (2008a) are inconsistent with prior experience, whereas the abstract and transfer ones appear more reasonable. We conjecture that this matters because, while no example is perfect, a "reasonable" example is more likely to activate relevant prior knowledge *without* also being unnecessarily distracting. In contrast, a cover story concerning a pizzeria where the cook systematically burns a calculated portion of each group order is likely to be at odds with prior knowledge concerning pizzerias, cooking, and business profitability.

Consequently, all the cover stories were changed to "a children's game from another country," the cover story used in the original transfer domain. We used play as an umbrella cover story because play naturally accommodates concrete as well as abstract instantiations.[4] We generally accept that children play all kinds of games, and we recognize that games can involve more concrete elements (e.g., combining cups of liquid) or more abstract ones (e.g., combining symbols). Our concrete training condition read “In another country, children play a game by combining cups filled with liquid,” whereas our abstract training condition read “In another country, children play a game by combining symbols.”

Our second concern had to do with the primary concrete training condition, where combining icons of two or more liquid-filled cups resulted in a “left-over” (i.e., a remainder). The icons used were those of 1/3, 2/3, and 3/3 (i.e., full) cups. Presumably, the authors used 3/3 because it matches the everyday intuition that 1/3 + 2/3 = 3/3. However, in their system, 3/3 acts as the identity element zero. But this means that, for example, 3/3 + 3/3 + 3/3 = 3/3 “left-over,” and therefore combining three full cups equal one full cup “left-over”—a rather nonintuitive conclusion. This particular issue was flagged by at least one scholar: Reed (2008) argued, and we agree, that “this use of the phrase ‘left over’ is inconsistent with its use in everyday language” and will “most likely [lead] to confusion” (p. 1633). Moreover, the equality 1/3 + 2/3 = 3/3 may lead to misleading intuitions about mod 3 arithmetic. In that system, 1 + 2 does not equal 3 in the everyday sense of “1, 2, 3.” Instead, 1 + 2 equals the identity element zero. Therefore, using a 3/3 cup for zero is not only inconsistent with everyday language, as Reed suggests, but potentially misleading because it obscures the cyclic nature of modular arithmetic. To avoid this issue, we used 1/3, 2/3, and an *empty cup* as the zero element. In our version, children play a game where they “empty a cup once it is full,” which provides a meaningful justification as to why, for example, 2/3 + 2/3 = 1/3 cup “left-over,” while preserving the usefulness of having an empty cup represent zero.

Similar to De Bock et al. (2011), and for the same reason, we developed an additional concrete transfer phase, modeled after the transfer domain in the original study. De Bock et al. repurposed the original study’s “burnt pizza” instantiation. Because that instantiation involved a chef systematically burning predetermined portions of each pizza order, it was arguably the worst in terms of connection-distraction tradeoffs. Instead, we repurposed the “tennis ball containers” concrete instantiation (cover story: “In another country, children play a game by combining containers filled with tennis balls”). In using a concrete transfer test different from that used by De Bock et al., we would be able to evaluate whether the results hold regardless of the test used. We also included a formal transfer phase. Introductory examples are used in math education to accelerate the learning of formal mathematics. To investigate which condition would serve as a better preparation for formal mathematics, we developed a formal transfer test. It involved a group of order 5 (consisting of 0, 1, 2, 3, and 4), and emphasized the underlying mathematical principles (e.g., each element having an inverse). It did not use a cover story.

We expected that the participants in the improved concrete condition would outperform participants in the abstract condition on the initial learning test. We conjectured that the previously reported advantage of abstract examples on abstract transfer would diminish, or disappear altogether, and that the concrete condition would outperform the abstract condition on concrete transfer. We had no particular conjecture regarding performances on formal transfer.

4. Methods

4.1. Participants

The participants were recruited from a large student volunteer pool (>10,000) from ETH and UZH in Zürich. The universities are similarly selective in the sense of being attended by approximately the top 20% of secondary school graduates. Students are able to take courses from both universities. We recruited first- and second-year students not majoring in mathematics or a mathematics-oriented discipline (e.g., computer science, statistics). Altogether, the study participants were (a) entry-level university students who (b) performed well during secondary schooling, (c) elected to attend large research universities, and (d) were *not* majoring in a mathematics-oriented discipline.

As required by the university, participants gave informed consent and were monetarily reimbursed for their time. The study was approved by the local ethics commission.

4.2. Study design

We recruited 114 participants and randomly assigned them to one of three conditions. Two of the conditions were relevant to the hypotheses tested in this study: the original abstract condition (M = 17, F = 21) and the modified concrete condition (M = 14, F = 24). Recall our goal was to test the original abstract condition against an improved concrete condition. We hypothesized that the improved concrete condition would yield better learning than the original abstract condition and that the abstract advantage previously shown for transfer would be reduced or absent. For opportunistic reasons, we also ran a modified abstract condition to examine a separate question of whether the original abstract condition could be improved on initial learning yet lead to worse abstract transfer performance. We did not find evidence that this modified abstract condition was better than the original abstract condition on initial learning, or worse on transfer, and we do not consider it as part of our main analysis here because it distracts from our purpose (but see Data S1 for details and results with all three conditions).

Based on Kaminski et al. (2008a) and De Bock et al. (2011), we aimed for at least 30 participants per condition. Across four sessions over 2 weeks, university students matching participation requirements were invited to participate in the study. We did not turn away additional volunteers, resulting in 38 participants per condition.

Participants completed the phases individually at secluded computer terminals in the following order: training (abstract or concrete), transfer abstract, transfer concrete, and transfer formal. The experiment was conducted by assistants who were blind to the study hypotheses. To account for order effects, the abstract and concrete transfer phases were counterbalanced. Otherwise, the protocol was kept the same as in the original study.

4.3. Materials

Fig. 1 shows the instantiations used in the training and transfer phases (see Fig. 2 for sample items). These instantiations were introduced via cover stories of children playing

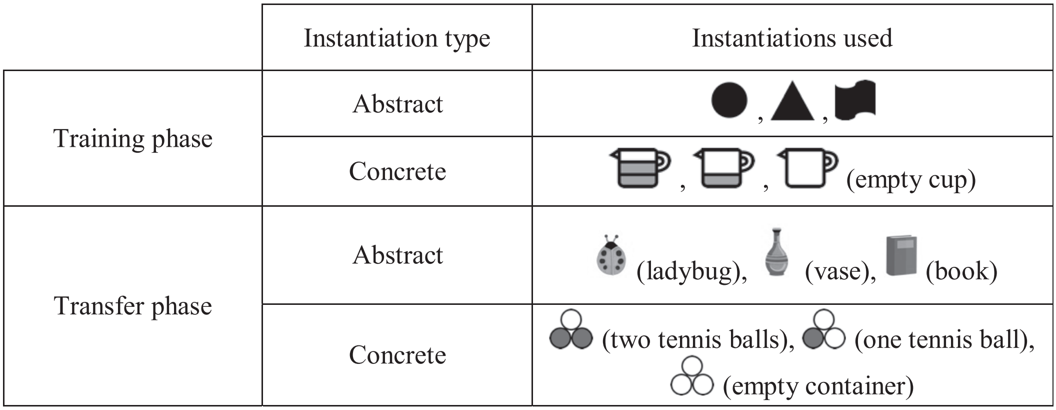


Fig. 1. Instantiations used in the training and transfer phases.

various games. The formal transfer phase used the standard Hindu-Arabic numerals. As in the original, only the training phase had explicit instructions.

Because the students in our study would be asked to solve three transfer tests rather than a single one, a compromise was made to remove four items from the multiple-choice tests (same items from each test), reducing the number of items on the abstract and concrete training and transfer tests from 24 to 20. The formal transfer test had 11 items, consisting of problems that could be encountered in an introductory abstract algebra class. Reliability analysis for the training test, abstract transfer, concrete transfer, and formal transfer tests showed McDonald's ω of 0.893, 0.857, 0.888, and 0.856, respectively.

4.4. Analytical procedures

As mentioned earlier, the abstract and concrete transfer phases were counterbalanced to check for order effects; as none were found, the results were combined. Unlike previous studies, no participants were excluded from the initial analysis.

To compare our findings to Kaminski et al. (2008a) and De Bock et al. (2011), we analyzed the differences between the conditions on initial learning and on transfer tests. For pairwise comparisons, we used the Mann–Whitney *U* test as the data were not normally distributed (see Data S1 for assumption checks); Bayesian Mann–Whitney *U* test was used to quantify nonsignificant results. We complemented this pairwise comparison with a mixed ANOVA. For effect sizes, we report rank-biserial correlations but also Cohen's *d*. To evaluate the effect of initial learning on transfer, nonparametric ANCOVA was used. Analysis was conducted via JASP (2018) and R (2018).

5. Results

Table 3 provides descriptive statistics for all participants. The majority of participants completed the study within an hour, with no one taking more than 75 min.

Table 3
Average scores (and *SD*) as a percentage

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

We report two kinds of analyses. To parallel the replicated studies, we first analyze the differences between the two conditions on initial learning and the three transfer measures. Second, because of clear differences in initial learning, we analyze the impact of initial learning on transfer.

5.1. Differences between the conditions on initial learning and transfer outcomes

Comparing concrete to abstract condition, pairwise tests found the following:

1. On initial learning, a large and significant difference in favor of the concrete training condition, Mann–Whitney $U = 1167.5$, $p < .001$, rank-biserial correlation $r_B = .617$ with 95% confidence interval CI [0.429, 0.754]; Cohen's $d = 1.276$.
2. On abstract transfer, no performance difference, $U = 701.5$, $p = .835$, $r_B = -.028$ with 95% CI [−.282, .229]; $d = -0.176$.
3. On concrete transfer, a significant and small-to-medium difference in favor of the concrete condition, $U = 913.5$, $p = .033$, $r_B = .265$ with 95% CI [0.010, 0.488]; $d = 0.416$.
4. On formal transfer, a small but nonsignificant difference in favor of the concrete condition, $U = 877$, $p = .103$, $r_B = .215$, with 95% CI [−0.043, 0.446]; $d = 0.290$.

To check for the influence of outliers, and to compare results with Kaminski et al. (2008a) and De Bock et al. (2011), we similarly excluded participants who scored two standard deviations away from the mean on any of the tests. Three participants were excluded from each condition (see outlier analysis in Data S1). Table 4 provides descriptive statistics after outlier removal.

After outlier removal:

1. The advantage of the concrete condition on initial learning remained large and significant, $U = 998$, $p < .001$, $r_B = .629$ with 95% CI [0.439, 0.767]; $d = 1.375$.
2. On abstract transfer, the difference remained nonsignificant, $U = 610.5$, $p = .986$, $r_B = -.003$ with 95% CI [−0.269, 0.263]; $d = -0.189$.
3. On concrete transfer, the difference became nonsignificant, $U = 758.5$, $p = .064$, $r_B = .238$ with 95% CI [−0.029, 0.474]; $d = 0.460$.
4. On formal transfer, the difference became larger and significant in favor of the concrete training condition, $U = 777$, $p = .049$, $r_B = .269$ with 95% CI [0.003, 0.499]; $d = 0.436$.

Table 4
Average scores (and *SD*) as a percentage, outliers removed

	Initial Learning	Transfer Abstract	Transfer Concrete	Transfer Formal
Abstract (n = 35)	72 (24.2)	79 (16.2)	92 (9.3)	73 (22.7)
Concrete (n = 35)	96 (5.5)	75 (25.2)	96 (5.7)	83 (21.3)

We supplemented the above pairwise comparisons with a 2 (condition: abstract, concrete) $\times$ 4 (test: initial learning, abstract transfer, concrete transfer, formal transfer) mixed ANOVA, with and without outliers. As ANOVA is a more conservative test, this analysis was conducted to check whether the results hold regardless of the method used. We found a significant interaction between condition and test ($p < .001$, with and without outliers). Simple effects tests indicated a significant difference on initial learning ($p < .001$, with and without outliers). Other differences were not significant (see Data S1 for the full analysis).

Because none of the classical tests indicated a significant difference between conditions on abstract transfer, we ran a Bayesian Mann–Whitney *U* test to quantify evidence for the null. The results indicate that the present data are at least three times more likely under the null model (i.e., no difference) than under the alternative model (see Data S1 for details). This can be considered moderate evidence in favor of there being no meaningful difference between the conditions on abstract transfer.

Finally, because our samples were relatively small (38 participants per condition), we focused our attention on effect sizes, arguably the most important measure when evaluating the effectiveness of educational interventions. Table 5 lists effect sizes, with and without outliers, in favor of the concrete training condition.

Outlier removal led to an overall improvement in favor of the concrete training condition.

5.2. *Differences between the conditions on transfer outcomes, controlling for initial learning*

To compare the two conditions while controlling for the initially superior learning in the concrete condition, we ran a nonparametric ANCOVA (Bowman & Azzalini, 2018; also see Data S1 for additional linear regression analysis) with learning test outcomes as covariates. Results indicate that the participants in the abstract training condition, compared to the concrete condition, had significantly higher abstract transfer ($p = .038$) and concrete transfer ($p = .014$) outcomes. There was no difference on formal transfer ($p = .665$). After outlier removal, however, this advantage of abstract training on transfer became nonsignificant for all transfer tests: abstract ($p = .150$), concrete ($p = .855$), and formal ($p = .598$).

Table 5
Effect sizes (as Cohen's *d*) in favor of the concrete training condition, with and without outliers

	Initial Learning	Transfer Abstract	Transfer Concrete	Transfer Formal
C > A all	1.276***	−0.176	0.396*	0.304
C > A no outliers	1.375***	−0.189	0.460	0.436*

*$p < .05$; ***$p < .001$

6. Discussion

We aimed to critically replicate and extend a well-known study that argued for the advantage of abstract examples in learning mathematics. We made two relatively minor but theoretically motivated modifications to the original study: (a) keeping the "cover stories" relatively similar to each other across the training and transfer tasks, and (b) using an icon of an *empty* cup rather than a *full* cup in the concrete training condition. These modifications were made with the goal of improving the concrete training condition by reducing distractions while connecting to prior knowledge. We also extended the study by including a more concrete transfer task, similar to De Bock et al. (2011), and an original formal transfer task. The results indicate that the new and improved concrete condition supported learning and transfer as anticipated by theory and design.

6.1. The advantage of concrete examples in learning math

6.1.1. Training

While Kaminski et al. (2008a) found no difference in initial learning, and De Bock et al. (2011) found a significant but small difference in favor of the concrete condition, we found a significant, large difference in favor of the concrete condition. Because concrete examples are generally more accessible, this *initial* advantage of concrete examples might be expected, at least when said examples are well-designed.

6.1.2. Transfer to an abstract domain

In contrast to previous work, we did not find compelling evidence for an advantage of abstract examples on abstract transfer. Kaminski et al. (2008a) reported an abstract advantage of $d > 1.6$ with statistical significance. In our study, the differences were nonsignificant. Bayesian analysis, combined with the marginal $d < 0.2$ effect size difference, suggests no meaningful difference between the conditions.

6.1.3. Transfer to a concrete domain

We found partial evidence in favor of the concrete instantiation on the concrete transfer test. However, the advantage was weaker than reported in De Bock et al.

(2011). De Bock et al. repurposed the "burnt pizza" instantiation (see section 3.1), while we modified the "tennis ball container" instantiation. One possible explanation is that, because we used a less distracting and more intuitive transfer test than "systematically burnt pizzas," a better overall performance could be expected regardless of initial training. This may have led to a ceiling effect, masking the difference between the conditions.

6.1.4. *Transfer to formal mathematics*

After outlier removal, we found evidence in favor of the concrete condition. In other words, despite not transferring as well to an abstract domain, concrete examples were more effective in transitioning to more formal mathematics.

6.1.5. *Overall findings*

Our results call into question the reported advantage of abstract examples. The participants in the concrete training condition performed as well or better than the participants in the abstract condition in our study, and generally performed as well or better than the participants in previous studies regardless of the condition (compare Tables 1, 2, and 4). The advantages of concrete examples became more pronounced with outlier removal, as summarized in Table 5.

Overall, the results reveal a modest advantage of concrete examples over abstract examples. As abstract algebra is an advanced (and abstract) area of mathematics, even a modest advantage is notable.

6.2. *The effect of initial learning differences on transfer*

The concrete condition participants performed the best overall. However, when controlling for initial learning (and before removing outliers), we found that the participants from the abstract condition significantly outperformed those from the concrete condition on abstract and concrete transfer, but not on formal transfer. Thus, if participants had learned equally well with abstract examples, this may have led to superior transfer test scores. However, the transfer advantage of the abstract condition was not enough to make up for its inferior initial learning, and the concrete condition participants performed better overall.

6.3. *Explaining the differences in learning and transfer*

What is the best explanation for the differences in initial learning and transfer, compared to the original studies? Were the participants more capable? If this were true, outcomes would have improved uniformly. However, consider the abstract condition performance on the abstract transfer test. There, our results match earlier work: 76% in Kaminski et al. (2008a), 75% in De Bock et al. (2011), and 79% in this study. Another possibility is that the students in the concrete training condition were, by chance, more mathematically capable than the students in the abstract condition.

But randomization and participant selection protocol guarded against this, and statistical tests control for such scenarios. Experimenter bias is also unlikely, since we largely reused materials and the experiment was conducted by assistants blind to the study hypotheses.

We offer that the differences in learning and transfer are best explained by appealing to the modifications made to the training materials. These modifications, while relatively minor, were theoretically motivated and done with the intent of improving the concrete training condition. We wanted to make the concrete example less distracting and more intuitively accessible, with as few modifications as possible. We hypothesized that these modifications would result in relatively better concrete condition outcomes on training and transfer. The results are in line with these expectations. This suggests that the differences in results, regarding the disappearing advantage of abstract examples, can be attributed to the changes made to the learning design. However, without a direct, randomly assigned comparison of the new and old concrete conditions, we cannot be certain that the modified concrete condition is strictly better for learning than the original concrete condition.

6.4. Participant elimination

We conducted our analysis on all participants, and then again after outlier removal. In contrast, the results reported by Kaminski et al. (2008a) and De Bock et al. (2011) entailed outlier removal *and* additionally eliminating participants scoring below chance on the training test, for "failing to learn." We question this method of eliminating participants in an educational study because it privileges students who found the materials useful in the first place. Teachers cannot ignore students who "failed to learn."

There is another reason to question this method of participant removal. In our study, eliminating participants who "failed to learn" did not lead to changing effect trends when compared to the already-implemented (2 SD) outlier removal (see Data S1 for details). However, the "failed to learn" criterion preferentially eliminated low-performing participants from the abstract condition, because that is the more challenging condition. Therefore, excluding students who "failed to learn" could bias the results by preferentially eliminating low-performing participants from the abstract, but not from the concrete condition.

6.5. Replacing the abstract–concrete dichotomy with a more comprehensive framework

Although our study adapted the terminology of abstract and concrete categories for the purposes of an experimental replication, we question the utility of these categories. Indeed, there have always been voices critical of the received abstract–concrete dichotomy (e.g., Wilensky, 1991). Supported by a growing understanding of cognition as embodied and situated (e.g., Barsalou, 2008; Hutchins, 2011; Norman, 1993), these critiques are likely to grow louder. A focus on "the abstract–concrete distinction" is even argued to be "no longer useful and should be replaced with more specific distinctions" (Barsalou, Dutriaux, & Scheepers, 2018, p. 9).

Recent work by Lampinen and McClelland (2018) illustrates this need for more specific distinctions in the context of mathematics education research. "Abstract" and "concrete" are problematic simplifications, the authors argue, because instantiations "that connect to different systems of knowledge may support different aspects of understanding" (p. 680). As evidence, they show that distinct instantiations can result in differences in learning, despite the instantiations being similarly concrete (see also Belenky & Schalk, 2014). A similar argument could be made here. Our replication preserves the "abstract" and "concrete" categories of the original study. If these categories had captured the features of importance, our results should have aligned with the original study's results. The fact that our results were different suggests that mathematics education research should move beyond the simplistic abstract–concrete distinction and investigate how learners actually interact with examples (e.g., Abrahamson & Trninic, 2015; Lobato, 2012; Nemirovsky, 2011).

6.6. Limitations and future directions

This study made modifications to the concrete training condition and compared it with the abstract condition. This allowed us to make claims about the relative advantages of the conditions; however, additional work is required to identify the impact of each modification. Moreover, additional work is needed to directly compare the original concrete condition with the modified one. Both could be accomplished with a 2×2 design randomizing to original or modified cover stories and full cup for zero (original) or empty cup for zero (modified).

Due to how the data were collected, we are unable to match all sample demographics to individual participants. Some of the results reported in Kaminski et al. (2008a) suggest that the efficacy of various kinds of training may be differentially impacted by participant ability. The absence of a means of assessing participant-level academic and mathematical ability is a limitation that should be remedied in future work.

We implemented the formal transfer test to investigate whether concrete or abstract example training transfers better to formal mathematics. While the results indicate an advantage of initial training with concrete examples, every participant trajectory included some exposure to abstract examples (via the abstract transfer test). To tease apart these influences, a follow-up is needed with participants exposed only to concrete or only to abstract examples prior to formal transfer. Moreover, while we used a formal transfer instantiation—with 0, 1, 2, 3, 4—this is not the only formal instantiation. As noted by a reviewer, arithmetic modulo 5 could be formally instantiated in other ways, for instance *e*, *a*, *b*, *c*, *d*. Would different formal instantiation lead to similar outcomes? Did the advantage of concrete examples on formal transfer manifest because of (relevant) surface similarities? Future work could explore this direction.

7. Implications and conclusion

What educational implications can be drawn from the study? We agree with De Bock et al. (2011) that brief interventions should not be used to inform sweeping changes in the classroom. Brief interventions cannot differentiate between a local performance maximum and deeper learning (Jones, 2009b; Kapur, 2016; Lampinen & McClelland, 2018). Indeed, a replication by McNeil and Fyfe (2012) demonstrated that the initial advantage of abstract examples in Kaminski et al. (2008a) decreases over time. The apparent advantage of concrete examples in this study should therefore be interpreted with some caution, and understood in the context of a critical replication. Concrete examples may be more or less useful, depending on their quality and desired educational outcomes.

In weighing the two options for classroom teaching, we would recommend initial training with concrete examples. When considering the impact of initial learning on subsequent transfer, an additional recommendation is possible. When controlling for initial learning, the abstract condition performed better on two of three transfer tasks. Although this advantage disappeared with outlier removal, it could be interpreted as support for instructional approaches such as *concreteness fading* (McNeil & Fyfe, 2012), where the initial learning advantage of concrete examples is leveraged to develop an understanding of increasingly more abstract mathematics.

Despite the conceptual critiques leveled against it, or because of them, Kaminski et al. (2008a) has remained steadily cited over the years, typically as evidence in favor of abstract examples. We offer an alternative explanation for those earlier findings. The "advantage of abstract examples" did not manifest because abstract examples are unambiguously advantageous. It was because, in that particular design, the concrete training condition was suboptimal. What makes the advantages of abstract examples disappear? In this case, the use of better concrete examples.

Acknowledgments

For their valuable feedback, we thank our colleagues at ETH Zürich's Institute of Learning Sciences and Higher Education, as well as Hillary Swanson and Ricardo Nemirovsky. We are also grateful to the anonymous reviewers and to Nicole McNeil, whose constructive critique and recommendations significantly improved the report.

Open Research badges

This article has earned Open Data badges. Data are available at available at https://osf.io/f56jm/.

Notes

1. An abstract group of order 3 is a mathematical structure: set of three elements and a binary operation satisfying certain rules: closure, associativity, identity, and inverse.
2. Modular arithmetic of groups of order 3 is a special arithmetic. One example is the set {0, 1, 2} under addition where 1 + 0 = 1, 2 + 0 = 2, 1 + 1 = 2, yet 1 + 2 = 0 and 2 + 2 = 1.
3. Indeed, Kaminski (2006) makes a similar observation: "Irrelevant concreteness is extraneous information that is unrelated to the to-be-learned concept and [may] hinder both learning and transfer" (p. 104).
4. Notably, Kaminski et al. (2008a) and De Bock et al. (2011) disagreed on whether the play example in the original study is more concrete or abstract. Play is sometimes viewed as a transitional stage between everyday sensemaking and more abstract forms of reasoning (Vygotsky, 1967).

References

Abrahamson, D., & Trninic, D. (2015). Bringing forth mathematical concepts: Signifying sensorimotor enactment in fields of promoted action. *ZDM Mathematics Education*, *47*(2), 295–306.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617–645.

Barsalou, L. W., Dutriaux, L., & Scheepers, C. (2018). Moving beyond the distinction between concrete and abstract concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1752). https://doi.org/10.1098/rstb.2017.0144.

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*(1), 27–50.

Bowman, A. W., & Azzalini, A. (2018). R package 'sm': nonparametric smoothing methods [computer software]. Available at: https://cran.r-project.org/web/packages/sm. Accessed November 8, 2019

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

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*, *26*(1), 9–25.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38.

Gravemeijer, K. (1994). *Developing realistic mathematics education*. Utrecht, The Netherlands: Freudenthal Institute.

Greeno, J. G., Moore, J. L., & Smith, D. R. (1993). Transfer of situated learning. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 99–167). Norwood, NJ: Ablex Publishing.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, *15*, 332–340.

Hutchins, E. (2011). Enaction, imagination, and insight. In J. Stewart, O. Gapenne, & E. A. Di Paolo (Eds.), *Enaction: Toward a new paradigm for cognitive science* (pp. 425–450). Cambridge, MA: MIT Press.

Inoue, N. (2005). The realistic reasons behind unrealistic solutions: The role of interpretive activity in word problem solving. *Learning and Instruction*, *15*(1), 69–83.

JASP. (2018). JASP [computer software]. Available at: https://jasp-stats.org/. Accessed January 22, 2019

Jones, M. G. (2009a). Transfer, abstraction, and context. *Journal for Research in Mathematics Education*, *40*, 80–89.

Jones, M. G. (2009b). 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. (2013). Extraneous perceptual information interferes with children's acquisition of mathematical knowledge. *Journal of Educational Psychology*, *105*(2), 351–363.

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. Available at: http://www.sciencemag.org/cgi/content/full/320/5875/454. Accessed January 20, 2019.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. (2009). Transfer of mathematical knowledge: The portability of generic instantiations. *Child Development Perspectives*, *3*, 151–155.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2013). The cost of concreteness: The effect of nonessential information on analogical transfer. *Journal of Experimental Psychology: Applied*, *19*(1), 14–29.

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist*, *51*(2), 289–299.

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.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, *17*(2), 248–294.

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.

Lave, J. (1992). Word problems: A microcosm of theories of learning. In P. Light & G. Butterworth (Eds.), *Context and cognition: Ways of learning and knowing* (pp. 74–92). New York: Harvester Wheatsheaf.

Lobato, J. (2012). The actor-oriented transfer perspective and its contributions to educational research and practice. *Educational Psychologist*, *47*(3), 232–247.

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

Nemirovsky, R. (2011). Episodic feelings and transfer of learning. *Journal of the Learning Sciences*, *20*(2), 308–337.

Norman, D. A. (1993). Cognition in the head and in the world: An introduction to the special issue on situated action. *Cognitive Science*, *17*, 1–6.

Palm, T. (2008). Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics*, *67*(1), 37–58.

R. (2018). R: A language for statistical computing [computer software]. Available at: https://www.R-project.org/. Accessed January 20, 2019

Reed, S. K. (2008). Concrete examples must jibe with experience. *Science*, *322*(5908), 1632c–1633c.

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

Siler, S. A., & Willows, K. J. (2014). Individual differences in the effect of relevant concreteness on learning and transfer of a mathematical concept. *Learning and Instruction*, *33*, 170–181.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, *8*, 247–261.

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.

Vygotsky, L. S. (1967). Play and its role in the mental development of the child. *Soviet Psychology*, *5*, 6–18.

Wilensky, U. (1991). Abstract meditations on the concrete and concrete implications for mathematics education. In I. Harel & S. Papert (Eds.), *Constructionism* (pp. 193–203). Westport, CT: Ablex.

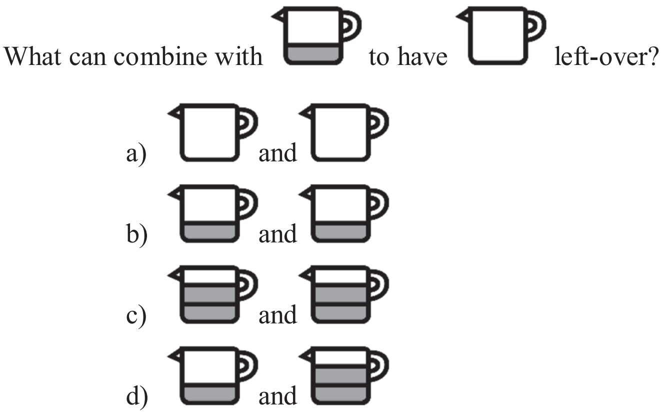
Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Data S1. Supplementary Materials for The Disappearing “Advantage of Abstract Examples in Learning Math”.

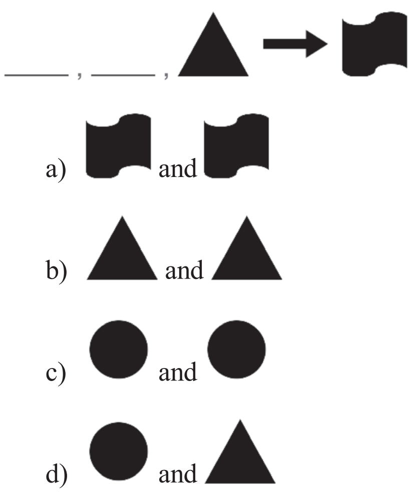
Appendix: Sample Test Items

Training phase: Concrete



Training phase: Abstract

What goes into the blanks to make a correct statement?



Transfer: Concrete

a)  and 

b)  and 

c)  and 

d)  and 

Transfer: Abstract

Children want the winner to point to  . They first point to  and then  .

What objects should they point to next?

a)  and 

b)  and 

c)  and 

d)  and 

Transfer: Formal

In the following expression, X could be any number in the system.

Solve: X + X + X + X + X

a) 1
b) 2
c) 3
d) 4
e) 0
f) Impossible to determine