We thank the reviewers for their insightful comments on our work, and especially for the helpful references, which we have read with interest, and which we believe have strengthened our understanding and presentation of our work. We have responded below to the individual comments.

Reviewers' comments:  
  
Reviewer #1:  
This study has carefully explored the effects of two types of presentations (concrete and abstract) on learning the targeted concepts (cyclic group) and related concepts (identity, inverse, generator). It also explored the effects of a hybrid presentation (containing both concrete and abstract) on learning. The overall experimental study design is rigorous. The materials seem to be well-written for subjects without prior knowledge of group theory, and the questions are well-targeted to answer the research questions. In addition, the authors have carefully analyzed and interpreted their data with thoughtful future research considerations. Therefore, I think this study may light on an important subject. Despite the strength of this paper, I do have a couple of concerns along with suggestions:  
  
1. The authors use both the terms "presentation" and "representation" in the manuscript, but redominantly use the first one. However, I was confused by the word choice as I read through the literature because many prior studies used the word "representations," and the use of the two terms in the paper did not seem clear. Initially, in the Introduction section, I thought the authors used "pedagogical presentation" to refer broadly to classroom instruction. Later, I thought it referred to "classroom discourse" specifically (e.g. "students may have some preconceptions about the objects included in the presentation"). It wasn't until the later reading of the literature, I believe, that the authors actually talked about "representation." Why did the authors choose to use "presentation" rather than "representation?" Is there a specific technical difference between the two as they are used in this paper, and how does the use of "presentation" in this paper relate to its use in other studies? If the authors clarified the language use at the beginning of the paper, it would make it easier to read.

We have added a paragraph at the beginning of the paper clarifying the distinction we are trying to draw with the two terms: We use the term ``presentation'' here to refer to the details of the pedagogical materials, while we reserve ``representation'' to refer to participants mental representation of the concepts, which, while grounded in the presentation used, may differ from it in important ways.

2. Literature Review  
(1) Literature is convincingly and critically reviewed. However, the motivation of your study is still not leveraged to the level required. At the end of the literature, the authors connected their studies like Gick and Holyoak (1983), Schwartz and Goldstone (2015), and Fyfe, McNeil, Son, & Goldstone (2014). Since prior studies have already reached the conclusion that both types of representations have advantages and disadvantages and a connection between them (even with specific sequence like "concreteness fading") can better support student learning, what motivates the current study for further exploration? In fact, the IES recommendations (Pashler et al, 2007) based on reviews of numerous cognitive psychology studies and classroom experiments have already recommended making connections between concrete and abstract representations as an instructional principle. And, the conclusion of this study is basically similar to these existing literature assertions. Therefore, how may the authors argue for their specific contributions? Can the authors pinpoint some specific gaps in the existing literature in which to situate their study? This would better-highlight the novelty of the current study. Right now, even though I appreciate the rigor of this study design, I am not very clear about the novelty of the findings. The authors may explicitly discuss this so as to help readers better appreciate their study.

We did not adequately clarify our intentions in this study. We attempted to make three contributions that elaborate upon the present state of the field:

* showing that presentations can affect learning of later concepts,
* broadening the discussion beyond simply concrete vs. abstract and irrelevant features (as in Belenky & Schalk, 2014) to different ways presentations may link to systems of understanding that students are familiar with and showing that these differences may impact learning.
* showing that combining presentations is beneficial for these phenomena, as has been shown in simpler cases, like concreteness fading (Fyfe, 2014).

In order to make this more clear, we have substantially rewritten the introduction of the paper to better highlight these issues and make the connection and contribution of our work to the previous literature more clear.

The second issue in particular was inadequately highlighted previously, so we have added new sections to the introduction and discussion where we discuss it. We hope to broaden the discussion in the field further beyond the simple “concrete vs. abstract’’ dimension, by illustrating that different presentations can relate a concept in different ways (in our case, visuospatially vs. connecting more directly to prior arithmetic concepts). These differences can accentuate or obscure different aspects of the concept in question. We suggest that rather than concrete vs. abstract, it is these connections that are the relevant feature for assessing presentations.

We hope that these changes make the motivation for and contribution of our study more clear.

(2) Additionally, the introduction could benefit from more grounding in cognitive science and mathematics education literature. The authors reviewed the interrelatedness of math concepts and mention that learned concepts build on prior knowledge. This paper would be stronger, as a whole, if they accordingly included a stronger theoretical framework for this notion. For instance, the mathematics education field generally viewed "understanding" as connection making among concepts (Hiebert et al). Stigler and others argued for the conceptual structure of mathematics.  
- Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). Teaching the conceptual structure of mathematics. Educational Psychologist, 47, 189-203.  
- Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K. C., Wearne, D., Murray, H., et al. (1997). Making sense: Teaching and learning mathematics with understanding. Portsmouth: Heinemann.

These were helpful resources, we have added some reference to this literature both in the introduction and in the Relationships Among Mathematical Concepts section, and increased our discussion of the importance of relationships between concepts to understanding. We think this nicely integrates with the results of Hazzan (1999) and DeBock et al. (2011) to provide clearer motivation for our study.

3. Research questions. On p.8, the authors proposed a list of research questions as measure for the aspect of understanding. The authors admit that it was not a comprehensive list. However, why are these components, in particular, key to assessing understanding? What is the conceptual framework for assessing understanding that justifies focusing on these aspects? The following two pieces may be helpful in this regard:  
  
- Greeno, J., & Riley, M. (1987). Processes and development of understanding. In R. E. Weinert & R. H. Kluwe (Eds.), Metacognition, motivation, and understanding (pp. 289-313). Hillsdale: Lawrence Erlbaum Associates.  
- Bisanz, J., & LeFevre, J. (1992). Understanding elementary mathematics. In J. I. D. Campbell (Ed.), The nature and origins of mathematical skills (pp. 113-136). Amsterdam: North Holland, Elsevier Science.

We have added reference to this literature and some discussion to clarify our perspective on why these aspects are important, both in the section where we introduce these questions and in the discussion. Specifically, we highlight the complexity of the word “understanding” and the fact that examining “understanding” in different ways may lead to different conclusions, as highlighted by both Greeno & Riley, and Bisanz & LeFevre. We use this to highlight the limitations of previous work which has evaluated only a limited type of understanding.

4. Methods (Experiments)  
  
(1) Design. Did the hybrid group receive more total instruction than the modular or geometrical groups? It seems that they necessarily would have to. If so, was there any correction for this or other attempt to establish that the observed effects were due to the differing nature of instruction rather than the increased quantity?

The hybrid presentation subjects received two more sentences of instruction on the operation than the other experimental groups, but performed the same amount of practice problems, and received the same amount of instruction on all later parts of the experiment. We have added a paragraph to the “Group Presentations” section discussing this in more detail.

(2) Hypothesis  
On p.15, the authors have provided a list of hypotheses. They have clearly stated that hypotheses for Experiments #2 and #3 were based on the results from Experiment #1. Maybe for experiment #1, they can add some literature support for their hypothesis.

This idea was inspired by the results from De Bock et al. which showed that subjects were making different inferences about higher level rules or relationships, which we hypothesized could better prepare subjects for learning later concepts. We have added material clarifying this to the hypotheses section, as well as slightly expanding our discussion of it in the introduction to make the connection more clear.

(3) Statistics tests - Provide citations for some statistics tests or terms and provide some justification as to why they are appropriate for your study (e.g., "naïve Bayes classifiers")

We have added references for a number of the methods we used:

* We have added a reference to an article which gives an overview of naïve Bayes and describes when it might be useful.
* We have added a reference to an article about hierarchical modeling and when it is useful in the hierarchical modeling section.
* -Added references about bootstrapping and bootstrapped logistic regression, and brief justification of our use of this approach.
* Added a reference on the Bayesian Information Criterion for model comparison.
* Added a reference to a paper on the benefits of within-paper meta-analysis.

5. Results. The format of report is somewhat monotonous. It might be helpful if the authors were to add some key interpretations along with the main findings in each section so as to highlight the main points (of course, they can save major discussions for later sections).

We did include some brief highlights after each section of results, we have expanded these somewhat and moved them to the beginning of each result subsection in order to frame the following results, which we hope improves readability.

6. Discussion  
(1) The authors provide convincing literature review to situate their study. For instance, I agree with the author's interpretation of Kaminski, Sloutsky, and Heckler's (2009) findings. Therefore, I expect they could link back to this literature more closely. In the current version, the majority of discussion is limited to the results; only a few places mention prior studies. Is it possible for the authors to more substantially link to the reviewed literature in order to explain or discuss their results? For instance, in comparison with Goldstone and others' research assertions, the authors could clarify what new ideas and issues your findings have brought to the field in terms of connections between concrete and abstract.

We have made a number of edits throughout the discussion in order to make the connection between our work and the previous literature more clear:

* Connected more directly to the results of De Bock et al. & Kaminski et al. in the discussion of our results.
* In the new Concrete vs. Abstract section we connect back to Belenky & Schalk (2014) and the concreteness fading literature (Goldstone & Son 2005, Fyfe 2014) and highlight the importance of our contribution.
* Added reference to various relevant literature in formalization section (Richland (2012), Hazzan (1999), Greeno & Riley (1987) and Burger & Shaughnessy (1986)).
* Added reference to complexity of understanding literature in Polygon vs. Modular Presentations section.
* Added reference to previous literature on multiple representations in the Hybrid Group section.

(2) The authors should be careful with their discussions - it seems that some are not based on their findings. For instance, on p.28, the authors claim, "It is important to clarify that to be beneficial, a hybrid presentation must be constructed from distinct presentations with distinct advantages." It seems that their findings do not necessarily lead to this conclusion.

We have clarified that this claim is a hypothetical suggestion based on our interpretation of our results and Kaminski’s, rather than a direct result of our experiment.

(3) In the discussion of the subjects' difficulty with formalization and generalization, the authors do not mention the possibility that subjects struggled with this because the act of formalization/generalization itself is a learned skill. Thus, there appear to be large potential differences in what might be observed here between teaching an unfamiliar concept to math students, who might have facility with formalization as a thinking skill (such as late-undergraduate math students), and to subjects without extensive mathematical training. In fact, the authors mentioned that participants who have sufficient knowledge background have been removed from analysis. Therefore, a lack of necessary prior knowledge/math skills may be a factor that cause difficulty in formalization/generalization. This may be connected back to the literature about the interrelated math concepts.

We think this is an important point, which we tried to convey in our discussion section on formalization and generalization. We have expanded our comments on this (including references back to Richland (2012), Hazzan (1999), Greeno & Riley (1987) and Burger & Shaughnessy (1986)) in hopes of making this more clear.

7. Other  
- p.7 The authors mentioned student may "[substitute] a variable in as the solution of a problem the students can already solve" (e.g. "5 + 6 =? " to "5 + 6 = x , solve for x "). I think a better example that shows the connection between arithmetic and algebra would be, 5 + ? = 11 and 5 + x = 11, as a part-unknown problem more clearly requires algebraic (rather than arithmetic) reasoning.  
- p.12 introduces a new term - process-level understanding - all of a sudden. Please define and explain.  
- Even though the authors strategically arranged another 15 figures in Appendix B, there are still 16 figures, which is too many for a regular manuscript. The authors could remove some of them. For instance, Figure 3 is similar to Figure 2. The authors may just describe Figure 3 in natural language and delete this figure.  
- Is Appendix B necessary? I would not include it in the manuscript.

-the suggested example is more clear, we have updated it.

-This is a vestige of an earlier draft of the paper, we have replaced it with a description that more accurately reflects the current argument.

-We have removed the nonagon diagram, but left the n-gon diagram since it is substantially different. We have removed the forest plots from the results section, and will provide them in supplementary material instead. We have removed two of the which representation figures showing less interesting results, and removed the hierarchical model figure.

-We have removed it.

Reviewer #2: Two Presentations of a Mathematical System: Complementary Advantages that Can Be Combined  
  
This paper investigates the effects of two types of presentations (i.e., modular and polygon) and their combination on learning about various aspects of cyclic groups (e.g., operator, inverse, generator). The authors find that there are advantages and disadvantages for each kind of presentation for learning these concepts. They also found that the hybrid presentation appears to help some participants improve their performance, behaving as if they used each representation in an optimal way. Although this paper has some potentially interesting results, I had difficulty locating the contribution of the work in relation to the broader literature. Below I describe my concerns with some suggestions for improving the paper.  
  
Relations to the Broader Literature  
The general message that there is no single best way to present instructional material and that different presentations can have different affordances is an important one (see Koedinger, Corbett, & Perfetti, 2012; Nokes & Ohlsson, 2005 for similar points). However, the overall goal of the paper and the contribution to the broader literature was not clear. The authors begin the paper by a detailed description of the Kaminsky et al. (2008) and follow-up papers. This situates the current work as addressing the issues pursued in and questions derived from that work. Only somewhat later in the introduction do the authors describe some of the more general aspects of the literature of abstract versus concreteness of presentations (e.g., on p. 8 referencing the Gick & Holyoak and Goldstone work). Because there were multiple issues at play it was unclear exactly which the authors aimed to directly address and test in the current work.  
  
Here is a list of the issues discussed at some point in the paper:  
What is learned from different types of presentations? (general)  
What is learned from abstract versus concrete presentations? (more specific)  
What is learned from hybrid, or multiple presentations?  
What is transferred to new problem solving or learning situations?  
How does initial learning help or hinder later learning?  
  
Each of these questions could connect to a much broader literature than were cited in the current work. For example, if the primary focus was further testing of the advantages and disadvantages of abstract versus concrete presentations, there is a very large body of prior work that was not considered (see Belenky & Schalk, 2014 for a review). Further, if this was the primary focus, the mapping and justification of the conditions to abstract and concrete materials was not clear. Presumably, modular presentation was the more abstract condition and polygon presentation was the more concrete condition. However, the authors do not go into details about the potential differences in what constitutes abstract or concrete in the current materials and how those distinctions relate to the prior literature. Perhaps if there were more connections to these categories they would have further informed possible hypotheses about why one presentation may afford certain types of learning over another.  
  
The authors included a hybrid condition that included both representations. However, they do not review much literature on the impact of multiple representations on learning outcomes (e.g., see Ainsworth, 2006; Rau, 2016). There is a growing literature on how multiple representations can impact learning that was not connected to in the current paper. Again, it was not clear how this hybrid condition relates to prior work.  
  
The authors also describe wanting to know what transfers (p. 7). This is also a large literature that is not well connected to (e.g., Day & Goldstone, 2012; Detterman & Sternberg, 1993; Gick & Holyoak, 1987). The analysis also does not deeply explore the question of what is transferred between the group 6 and 9 problems. Answering this question would help provide an understanding of what was learned from the initial instruction on group 6. For example, many results show an improvement in group performance from 6 and 9 which suggest positive transfer and further learning. However, a few tasks show a decrement such as demonstrated in Figure 25. These differences and their implications are not described in depth.  
  
Finally, the authors describe the importance of initial learning on later learning and that "nothing is learned in isolation", which indeed is an important point and concept. There has been much work on a related idea of "Preparation for Future Learning" by John Bransford and Daniel Schwartz (see Schwartz, Bransford, & Sears, 2005 for a review). However, much of the past work examining the question of how initially different experiences (i.e., different presentations) affect later learning from a resource usually give both groups the same later learning resources. If I understand the current design correctly, the different presentation symbols were maintained through the later learning phases of identity, inverses, and generators. This context is different than that past work asking similar questions. The current work appears to ask how do two different types of representational forms (modular vs. polygon) affect learning multiple related concepts.  
  
Given that these issues were reviewed at different depths and the connections to the particular conditions of the current experiments were tenuous in some cases, it was difficult to assess the contribution to the literature.  
  
Suggestion: clearly highlight which primary issue(s) the current work is addressing in the introduction (adding more related literature to that issue(s)), clearly describe how the current work will address that issue, test, and contribute to it, and then in the discussion relate back to that particular issue.

As above: We did not adequately clarify our intentions in this study. We attempted to make three contributions that elaborate upon the present state of the field:

* showing that presentations can affect learning of later concepts,
* broadening the discussion beyond simply concrete vs. abstract and irrelevant features (as in Belenky & Schalk, 2014) to different ways presentations may link to systems of understanding that students are familiar with, and showing that these differences may impact learning.
* showing that combining presentations is beneficial for these phenomena, as has been shown in simpler cases, like concreteness fading (Fyfe, 2014).

In order to make this more clear, we have substantially rewritten the introduction of the paper to better highlight these issues and make the connection and contribution of our work to the previous literature more clear.

The second issue in particular was inadequately highlighted previously, so we have added new sections to the introduction and discussion where we discuss it. We hope to broaden the discussion in the field further beyond the simple “concrete vs. abstract’’ dimension, by illustrating that different presentations can relate a concept in different ways (in our case, visuospatially vs. connecting more directly to prior arithmetic concepts). These differences can accentuate or obscure different aspects of the concept in question. We suggest that rather than concrete vs. abstract, it is these connections that are the relevant feature for assessing presentations.

We have also addressed some of the specific points you raised:

* Added more literature on Abstract vs. Concrete, including Belenky & Schalk’s review and some of the concreteness fading literature (Goldstone & Son, 2005; Fyfe et al., 2014).
* Added reference to Preparation for Future Learning (Bransford & Schwartz, 1999) and the idea that different ways of learning may support different types of understanding (Nokes, 2005) to provide context for and more clearly motivate the interest in learning later concepts.
* Added references to Ainsworth (2006) and Rau (2016) to better motivate the idea that multiple presentations may be beneficial or detrimental depending on their features..
* Added a paragraph to the new Limitations & Future Directions section acknowledging the fact that different presentation symbols were used for the two groups throughout, and suggesting that future work should investigate whether this is necessary to observe effects on subsequent concepts.

We hope that these changes make the motivation for and contribution of our study more clear.

Minor Comments  
There were too many graphs. They could be cut down to a critical number of core findings of at most 4 or 5 figures.

As above: We have removed the nonagon diagram, but left the n-gon diagram since it is substantially different. We have removed the forest plots from the results section, and will provide them in supplementary material instead. We have removed two of the which representation figures showing less interesting results, and removed the hierarchical model figure.

References  
  
Ainsworth, S. (2006). DeFT: a conceptual framework for considering learning with multiple representations. Learning & Instruction, 16, 183-198.  
  
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  
  
Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. Educational Psychologist, 47, 153-176.  
  
Detterman, D. K., & Sternberg, R. J. (Eds.). (1993). Transfer on trial: intelligence,cognition, and instruction. Norwood, NJ: Ablex.  
  
Gick, M. L. & Holyoak, K. J. (1987). The cognitive basis of knowledge transfer. In S. M. Cormier & J. D. Hagman (Eds.), Transfer of learning: Contemporary research and applications (pp. 9-46). New York, NY: Academic Press.  
  
Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction (KLI) framework: Bridging the science-practice chasm to enhance robust student learning. Cognitive Science, 36, 757-798.  
  
Nokes, T. J., & Ohlsson, S. (2005). Comparing multiple paths to mastery: What is learned? Cognitive Science, 29, 769-796.  
  
Rau, M. A. (2016). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. Educational Psychology Review, 1-45.  
  
Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. Mestre (Ed.), Transfer of learning from a modern multidisciplinary perspective (pp. 1-51). Greenwich, CT: Information Age.  
  
  
  
Reviewer #3: I think that the topic is quite timely and relevant. The complexity of the topic and design of the study requires a clearly-written MS and I think that the authors have achieved this. Brief summaries and examples help with the density of some of the sections. I offer a few general comments below.  
  
Introduction/Literature Review  
  
The specific literature supporting the current study has been sufficiently reviewed. Providing a bit more detail regarding the broader context and importance of the study seems warranted. For example, when the set of specific guiding questions are proposed on p. 8 the reader should be made aware of the importance of these questions for mathematical cognition and pedagogy; beyond their particular relevance to furthering Kaminski's work and related studies.

We have made note of the broader relevance of our specific questions about presentations, while attempting to still maintain the position that these questions are just a starting point rather than a definitive list. Also, in revising the introduction and discussion, we have tried to include more connections to broader issues.

Discussion  
  
Include a more thorough discussion of the limitations of the current set of studies. Again, this section can also provide space for discussion related to a few of the broader implications this work has for learning and instruction related to the findings.

We have added a Limitations & Future Directions section to the end of the discussion to address these points.