Peer review history of the paper Language Effects in Early Development of Number Writing and Reading by Anna Frieda Steiner, Sabrina Finke, Francina J Clayton, Chiara Banfi, Ferenc Kemény, Silke Melanie Göbel, and Karin Landerl published in Special issue Direct and Conceptual Replications in Numerical Cognition in Journal of Numerical Cognition (vol 7, 3), https://doi.org/10.5964/jnc.6929

Table of Contents

Initial Decision Letter

Dear authors,

I now have received the reviews of two expert reviewers on this topic. Both reviewers found it as a merit study and provided some constructive comments, which could help to improve the readability of the paper. The main concern of reviewer 1 is that the paper can benefit from streamlining and shortening. In this regard, consulting with the guideline of the special issue might be helpful (https://osf.io/a9vye/). The main concern of reviewer 2 is to further elaborate on some methodological information and include relevant literature investigating the Arabic language which has the inversion property as well. I agree with the reviewers' comments. Please also report power/sample size calculation in the main text rather than as a footnote as it plays one of the key roles of the current special issue. Accordingly, I decide for a moderate revision and would like you to revise the manuscript using track changes and include a response letter that details how each of the issues raised by the reviewers was addressed. I look forward to your revision.

Best regards,
Mojtaba Soltanlou

Reviewer A:

Overall, this paper systematically explores number transcoding (both reading and writing) in two languages with a specific emphasis on number structure and inversion. The statistical analysis is clear and well written. However, my overarching issue with this paper is that it is very unwieldy- it needs to clearly drive home the theoretical importance of the paper and be more concise in the way it is presented.

Abstract

What does the "their" in this statement refer to: "aimed to replicate their…"

Define "transcoding direction"

Introduction

Overall, the introduction is quite difficult to follow. Perhaps subheadings would help the flow of the material through the introduction section.

In the English this number "five hundred sixty" would actually be stated as "five hundred and sixty"

This sentence "three-digit numbers to dictation." should read "three-digit numbers from dictation."

Page 6: "Average error rates were reported from unbalanced number structures, e.g., twelve XX-numbers in contrast to four X0-numbers." This section is a little confusing and not straightforward to comprehend, it would benefit from a rewrite.

Page 6: This sentence does not appear to be complete: "Barrouillet et al. (2004) suggested early lexicalization of two-digit numbers within the first two years of school."

Page 8: I find it difficult to believe this statement- has a substantial literature review been completed? It is has please state this in the paper: "A further extension to Zuber et al. (2009) was that we did not only use number writing, but asked children to read the very same numbers aloud. Interestingly, this has not been done before."

Page 8: There are no clearly stated hypotheses is this an exploratory study or are there clear predictions?

Page 11: Was any inter-rater reliability conducted for the tasks, especially the number reading task?

Results

Page 15: The standard deviations in the inversion-related errors results are huge and therefore warrant discussion.

Page 16: Please report the results of the inferential test in relation to this sentence: "Zuber et al. (2009) reported significantly more inversion-related errors for number writing of three- (M = 25.20%) than two-digit numbers (M = 24.06 %)."

Typo in Figure 3 bar label for syntactic with inversion error

Page 20: Can the authors clarify that two-digit numbers are taught in Grade 1 in both regions from which children were recruited?

Discussion

Page 31/34: I think the authors make a very important point here: "Note that in number reading, digit order is permanently visible to the child, so working memory demands are clearly lower during number reading than during number writing. Nevertheless, digit order was still a fundamental challenge for German-speaking children also during number reading, at least in the early stages of development." Therefore, further investigations should be suggested to be explored in response to this- assessing working memory demands of transcoding in reading and writing or individual differences in working memory and it's relation to transcoding accuracy.

Page 31: Could an additional explanation for this point "In three-digit numbers, however, XXX-numbers did not reveal language-specific accuracy rates." be that children in both groups simply lacked enough expertise with 3-digit numbers to display language effects?

There are no limitations of the study discussed. Are there any applications of these findings- or is this study really focusing on updating the transcoding model (which is, of course, fine). However, if the focus is really about updating a model of transcoding I would expect more in depth discussion of the updates to the model that are an outworking of these data.

Reviewer B:

The study of inversion effects and transcoding from a cross linguistic perspective is very interesting and contributes to understanding the developmental process of transcoding.
The replication of Zuber et. al. study is fine and even more interesting is the German-English comparison and the 1st 2nd grades comparison besides to the elaboration of error analysis of two and three digit numbers into numbers with\without syntactic load is appreciated.
However, I have some notes that if the authors take into account the paper will have more impact:
1. page 3, I guess the author intended to write "203 might be read as twenty three" or not?!
2. German has the characteristic of number inversion and the author have reviewed the literature but did not review any research conducted in Arabic which have also the inversion feature! integrating studies on numerical cognition in Arabic will enrich the paper and will increase its generalizability.
3. Studies by Ganayim, D. are recommended but the study "Ganayim, D. , Ganayim, S. , Dowker, A. and Olkun, S. (2020) Linguistic Effects on the Processing of Two-Digit Numbers. Open Journal of Modern Linguistics, 10, 49-69. doi: 10.4236/ojml.2020.101004." is a must for the current paper since it deals with linguistic effects of inversion in transcoding in Arabic and Hebrew with two tasks of reading and

writing. in addition the syntactic load of numbers was regarded. I believe the the author will benefit from encompassing this study specifically.

3. I recommend a table with the authors predictions according to the models.

4. more details are needed regarding how the participants were chosen, initial math level (evaluated by teacher or screening test). since the study had a reading task the initial reading level is also important.

5. stimuli, why the two-digit and three digit numbers did not contain the same number of stimuli? besides the xx category is larger (16) than the 1x (4) and x0 (4) this may have a confounding effect since the possibility of an error is larger. please explain that.

6. I recommend to combine figures2a+b into one figure for more visibility.

I hope my notes will advance and improve the paper.

Authors' Response

In the following, we provide a point-by-point response to all comments and concerns.
Editor (Ed): Both reviewers found it as a merit study and but also provided some constructive comments, which could help to improve the readability of the paper. The main concern of reviewer 1 is that the paper can benefit from streamlining and shortening. In this regard, consulting with the guideline of the special issue might be helpful (https://osf.io/a9vye/). The main concern of reviewer 2 is to further elaborate on some methodological information and include relevant literature investigating the Arabic language which has the inversion property as well.

*Response: We thank the reviewers and editor for the overall favourable evaluation of the presented replication of the seminal study by Zuber et al. (2009) on the impact of decade-unit inversion on transcoding of multi-digit numbers in young children. We followed the suggestion by reviewer 1 and included subheadings in the introduction. Following the guidelines of the special issue, we clearly presented and explained our hypothesis. Moreover, to focus on brevity, we omitted parts of the introduction and discussion and concentrated on essential information that explains our motivation for the replication. We extended our method section and included the information requested by reviewer 2. As outlined in our response to reviewer 2 (Q3) we included literature on transcoding in the Arabic language.

*Ed: I agree with the reviewers' comments. Please also report power/sample size calculation in the main text rather than as a footnote as it plays one of the key roles of the current special issue.

*Response: In the revised manuscript on p. 10-11, this has been adapted.

*Ed: Accordingly, I decide for a moderate revision and would like you to revise the manuscript using track changes and include a response letter that details how each of the issues raised by the reviewers was addressed. I look forward to your revision.

*Response: Thank you very much for giving us the opportunity to revise our manuscript and to resubmit the revised manuscript to the Special Issue of the Journal of Numerical Cognition.

To better differentiate following reviewers' points from our responses, each reviewers' point was marked with *Q().

Comments from Reviewer 1
*Q(1): Overall, this paper systematically explores number transcoding (both reading and writing) in two languages with a specific emphasis on number structure and inversion. The statistical analysis is clear and well written.

*Response Q(1): We thank Reviewer 1 for the positive feedback on our statistical analysis.

*Q(2): However, my overarching issue with this paper is that it is very unwieldy- it needs to clearly drive home the theoretical importance of the paper and be more concise in the way it is presented.

*Response Q(2): We have condensed the introduction and the discussion in the revised version to present our hypothesis more clearly and describe the replication motivation. The conclusion was further specified in order to be more specific about our suggested adaptations of current transcoding models. Please see Q(19) for further discussion.

Abstract
*Q(3): What does the "their" in this statement refer to: "aimed to replicate their…"

*Response Q(3): With "their" we referred to Zuber and colleagues, the original study´s authors. We changed the wording in the abstract in order to make that clear.

*Q(4): Define "transcoding direction"

*Response Q(4): Switching between symbolic number representations happens from Arabic numbers to number words (number reading) and vice versa (number writing). Transcoding implies both directions. This was clarified in the revised manuscript.

Introduction
*Q(5): Overall, the introduction is quite difficult to follow. Perhaps subheadings would help the flow of the material through the introduction section.

*Response Q(5): We followed the reviewer´s advice and inserted subheadings in the introduction of the revised manuscript.

*Q(6): In the English this number "five hundred sixty" would actually be stated as "five hundred and sixty"

*Response Q(6): We thank the reviewer for the attentive reading and have changed it accordingly on p. 3.

*Q(7): This sentence "three-digit numbers to dictation." should read "three-digit numbers from dictation."

*Response Q(7): In the revised manuscript on p. 4, this has been corrected.

*Q(8): Page 6: "Average error rates were reported from unbalanced number structures, e.g., twelve XX-numbers in contrast to four X0-numbers." This section is a little confusing and not straightforward to comprehend, it would benefit from a rewrite.

*Response Q(8): We rewrote the section on p. 6 to clarify the characteristics of included items.

"Interestingly, Zuber et al. (2009) did report accuracy rates separately for specific number structures in their Table 1, but all statistical analyses were based on error rates across number structures. This may be relevant as their item set included inversion-demanding numbers such as XX- (e.g., 24, X represents a number between 1 and 9) or XXX-numbers (e.g., 624), and also single digits, double-digit numbers (X0: e.g., 40) and three-digit numbers (X00: e.g., 600, X0X: e.g., 206 or XX0: e.g., 260) that do not require inversion. Obviously, error types (and rates) in mixed item sets depend on number structures included."

*Q(9): Page 6: This sentence does not appear to be complete: "Barrouillet et al. (2004) suggested early lexicalization of two-digit numbers within the first two years of school."

*Response Q(9): We changed to sentence to:
"Barrouillet et al. (2004) suggested that two-digit numbers are lexicalized within the first two school years."

*Q(10): Page 8: I find it difficult to believe this statement- has a substantial literature review been completed? It is has please state this in the paper:
"A further extension to Zuber et al. (2009) was that we did not only use number writing, but asked children to read the very same numbers aloud.
Interestingly, this has not been done before."

*Response Q(10): To identify transcoding studies we conducted a broad literature search using the databases PubMed, ResearchGate, ERIC, Science Direct, and Google Scholar. Besides transcoding we used the search terms number production, number to dictation, numeral writing, and number writing for verbal-visual transcoding. Search terms for visual-verbal transcoding were additionally number reading, reading aloud, and number naming. We found about 25 different studies on children that appeared to be relevant according to the title and abstract. The majority of transcoding studies investigated number writing (Barrouillet, Camos, Perruchet, & Seron, 2004; Byrge, Smith, & Mix, 2014; Camos, 2008; Imbo, Vanden Bulcke, De Brauwer, & Fias, 2014; Johansson, 2005; Lopes-Silva, Moura, Júlio-Costa, Haase, & Wood, 2014; Moura et al., 2015; Pixner et al., 2011; Power & Dal Martello, 1990; Simmons, Willis, & Adams, 2012; van der Ven, Klaiber, & van der Maas, 2017; Van Loosbroek, Dirkx, Hulstijn, & Janssen, 2009; Van Rinsveld & Schiltz, 2016; Zuber, Pixner, Moeller, & Nuerk, 2009). We found one study focusing on number reading (Power & Dal Martello, 1997). Some studies investigated number reading and number writing (Dowker & Roberts, 2015; Habermann, Donlan, Göbel, & Hulme, 2020; Lopes-Silva et al., 2016; Moeller, Zuber, Olsen, Nuerk, & Willmes, 2015; Moura et al., 2013). However, in the study by Moeller et al. (2015) only number writing was analyzed. The studies by Lopes-Silva et al. (2016), Moura et al. (2013), and Habermann et al. (2020) used different stimuli in number writing and number reading. In the study by Dowker and Roberts (2015) we cannot rule out that the same stimuli were tested in number reading as well as writing, as the authors did not provide a detailed description of included items.

We cannot rule out that it has been done for adults. However, studies for adults also lack a detailed description of included items (e.g., Ganayim, Ganayim, Dowker, & Olkun, 2020) We added the information in the text on p. 9 and rewrote the sentence accordingly. "Previous studies comparing number writing and reading typically used different items for the two conditions (Habermann, Donlan, Göbel, & Hulme, 2020; Lopes-Silva et al., 2016; Moura et al., 2013. For some studies, authors did not provide a detailed description of items: Dowker & Roberts, 2015; Moeller, Zuber, Olsen, Nuerk, & Willmes, 2015)."

*Q(11): Page 8: There are no clearly stated hypotheses is this an exploratory study or are there clear predictions?

*Response Q(11): We had clear hypotheses for the replication study. To better introduce our hypotheses, we rewrote the introduction and restructured the part "The current study".

*Q(12): Page 11: Was any inter-rater reliability conducted for the tasks, especially the number reading task?

*Response Q(12): We agree with the reviewer that it is important to report inter-rater reliability and included it in the revised manuscript on p. 13 and 14. This now constitutes a further strength of our replication study as inter-rater reliability was not reported in the original paper by Zuber et al. (2009).

To test the inter-rater reliability for transcoding accuracy, we computed Cohen's κ based on 26 German-speaking first-graders. For both number writing (κ = . 860, p < .001) and reading (κ = . 883, p < .001) inter-rater reliability was almost perfect (Landis & Koch, 1977). This result reflects that different raters were in agreement when distinguishing between correct and incorrect responses.
Error coding, however, was more complex and a certain degree of deviation between raters was observed. In number writing, the inter-rater reliability for single and combined errors ranged from moderate to almost perfect with κ-values between .633 and 1.000 (all ps < .001) for German- speaking sample (N = 170) and with κ-values between .501 and .970 (all ps < .001) for the English-speaking sample (N = 264).
In number reading, we only identified inversion-related errors. Inter-rater reliability was κ = . 948 (p < .001) for German-speaking children (N= 169) and κ = . 822 (p < .001) for English speaking children.

Results
*Q(13): Page 15: The standard deviations in the inversion-related errors results are huge and therefore warrant discussion.

*Response Q(13): The marked difference between standard deviations in our sample and the sample by Zuber et al. (2009) might derive from deviant task setting. Whereas number writing was carried out in single one-to-one sessions by Zuber et al. we conducted the task as a classroom-task broken up into four parts.
In order to provide more information on distribution of errors in our sample, we included the following section in the revised manuscript on p. 16:

"In the current sample variation for inversion-related errors was quite high and clearly higher than reported for the original study by Zuber et al. Most children did not experience problems with inversion as 82 participants made no inversion-related errors in two-digit numbers and 53 children made no inversion-related errors in three-digit numbers. However, some children still struggled with digit order and made inversion-related error rates in more than 50 % of the items (10 children in two-digit numbers and 12 children in three-digit numbers)."

We further added to the discussion on p. 30

"Large standard deviations in inversion-related errors indicate that children do not solve the task in the same way. Differences in (prior) knowledge on number writing and reading lead to a markedly variance in task performance."

*Q(14): Page 16: Please report the results of the inferential test in relation to this sentence: "Zuber et al. (2009) reported significantly more inversion-related errors for number writing of three- (M = 25.20%) than two-digit numbers (M = 24.06 %)."

*Response Q(14): We added results of the computed t test on p. 16:

"Zuber et al. (2009) reported significantly more inversion-related errors for number writing of three- (M = 25.20%, SD = 7.32) than two-digit numbers (M = 24.06 %, SD = 4.19), $t(127) = 3.18$, $p < .01$, $d = 0.23$."

*Q(15): Typo in Figure 3 bar label for syntactic with inversion error

*Response Q(15): We have changed it accordingly.

*Q(16): Page 20: Can the authors clarify that two-digit numbers are taught in Grade 1 in both regions from which children were recruited?

*Response Q(16): Yes, two-digit numbers are taught explicitly in both countries, and this was now clarified in the text (see p.11). As a matter of fact, the high accuracy rates for two-digit numbers across languages indicate that both German- and English-speaking first-graders were familiar with transcoding of two-digit numbers.

"Two-digit numbers are explicitly taught in Grade 1 in both countries. The Austrian national curriculum (Bundesministerium Bildung, Wissenschaft und Forschung, 2012) does not specify separate learning goals for Grades 1 and 2, but it requires that children should be able to count, read and write numbers up to 100 at the end of Grade 2. Most mathematical textbooks used in Year 1 include numbers up to 100. According to the UK national curriculum (Department for Education, 2013) children are expected to be able to count, read and write numbers to 100 by the end of Grade 1. According to both curricula, three-digit numbers are not explicitly taught before Grade 3."

Discussion

*Q(17): Page 31/34: I think the authors make a very important point here: "Note that in number reading, digit order is permanently visible to the child, so working memory demands are clearly lower during number reading than during number writing. Nevertheless, digit order was still a fundamental challenge for German-speaking children also during number reading, at least in the early stages of development." Therefore, further investigations should be suggested to be explored in response to this- assessing working memory demands of transcoding in reading and writing or individual differences in working memory and it's relation to transcoding accuracy.

*Response Q(17): In the revised manuscript we added implication for future studies on p. 33:

"So far, the impact of working memory on transcoding was only investigated for number writing (Imbo et al., 2014; Simmons, Willis, & Adams, 2012; Zuber et al., 2009). Future studies should investigate the association of individual differences in working memory with number reading skills."

*Q(18): Page 31: Could an additional explanation for this point "In three-digit numbers, however, XXX-numbers did not reveal language-specific accuracy rates." be that children in both groups simply lacked enough expertise with 3-digit numbers to display language effects?

*Response Q(18): We agree with the reviewers' suggestion. We mentioned this possibility on page 31:
"Language-independent challenges and a **lack of experience** may have masked the impact of decade-unit inversion. In German, inversion errors often co-occurred with other syntactic errors as shown by the high frequency of combination errors."

Additive composition errors were prominent in German as well as English. Therefore, accuracy was not affected by language and was not specifically lower in German than in English.

*Q(19): There are no limitations of the study discussed. Are there any applications of these findings- or is this study really focusing on updating the transcoding model (which is, of course, fine). However, if the focus is really about updating a model of transcoding I would expect more in depth discussion of the updates to the model that are an outworking of these data.

*Response Q(19): The following adaptations were specified already in the previous version of the manuscript: Models should 1) account for language-specific characteristics in transcoding such as the prominent decade-unit inversion, 2) explain accuracy patterns based on individual number structures. This includes to accommodate (language-independent) higher demands in transcoding XX0- compared to X0X-numbers and 3) pertain to both transcoding directions as similarities between number writing and reading

seemed to overweight differences (this point was emphasised in the new version of the manuscript, please refer to Q17).

In the manuscript (p. 35) we specified necessary adaptations for decade-unit inversion in the ADAPT-model.
"First, adaptations to linguistic peculiarities such as decade-unit inversion are needed. The ADAPT-model (Barrouillet, 2004) suggests for algorithmic transcoding that number words are sequentially parsed and constituent digits are retrieved from long-term memory. In German number writing, an additional procedure would be necessary to reorder digits stored in working memory."

More detailed analyses would be necessary to explain higher accuracy rates for XX0-numbers than for X0X-numbers (language-independent adaptations). So far, only speculations were suggested by current transcoding models. Power and Dal Martello (1990) proposed that superfluous zeros in XX0- but not in X0X-numbers might be due to overwriting problems. Children might only be familiar with overwriting the unit (and not the decade) as this was already practiced in two-digit numbers. In the ADAPT-model problems with superfluous zeros were explained by inadequate rules responsible for programming or managing the number frame. A superfluous slot, however, cannot explain the accuracy difference between X0X- and XX0-numbers. In number reading the syntactic zero in X0X-numbers might help to identify the first number as a hundred because the zero is a place holder for the separate word "hundred". The Arabic number 206 can be sequentially transcoded as 2 = two, 0 = hundred (and), 6 = six.
Clearly, this is a speculative response and we feel that before mentioning this as possibility in a manuscript, more in-depth studies which specifically investigate the impact of lexical and syntactic zeros in three-digit number transcoding would be necessary.

Comments from Reviewer 2
*Q(1): The study of inversion effects and transcoding from a cross linguistic perspective is very interesting and contributes to understanding the developmental process of transcoding.
the replication of Zuber et. al. study is fine and even more interesting is the German-English comparison and the 1st 2nd grades comparison besides to the elaboration of error analysis of two and three digit numbers into numbers with\without syntactic load is appreciated.

*Response Q(1): We thank the reviewer for the positive comments on our manuscript.

However, I have some notes that if the authors take into account the paper will have more impact:
*Q(2): 1. page 3, I guess the the author intended to write "203 might be read as twenty three" or not?!

*Response Q(2): Indeed, both answers have been observed. Our example on p. 3 refers to problems in correctly applying the decimal word "hundred". First, it was "hundred" and "thousand" was mistaken. Second, it was misplaced and inserted after the second instead

of the first digit. "Expressing the multiplicand wrongly" was also reported and discussed by Power and Dal Martello (1997).

The error type mentioned by reviewer 2 has been observed in our study as well. Children split connected digits and in case of 203 they read twenty, three. This error might constitute the counterpart to overwriting problems when the place-value system has not been acquired yet.

In the manuscript's introduction we just wanted to give examples of errors in reading and writing of numbers during early development. However, we can see from the reviewers' query that the incorrectly split number is more intuitive than confusion of the decimal word. Thus, we changed the example accordingly.

*Q(3): 2. German has the characteristic of number inversion and the author have reviewed the literature but did not review any research conducted in Arabic which have also the inversion feature! integrating studies on numerical cognition in Arabic will enrich the paper and will increase its generalizability.

3. Studies by Ganayim, D. are recommended but the study "Ganayim, D. , Ganayim, S. , Dowker, A. and Olkun, S. (2020) Linguistic Effects on the Processing of Two-Digit Numbers. Open Journal of Modern Linguistics, 10, 49-69. doi: 10.4236/ojml.2020.101004." is a must for the current paper since it deals with linguistic effects of inversion in transcoding in Arabic and Hebrew with two tasks of reading and writing. in addition the syntactic load of numbers was regarded. I believe the the author will benefit from encompassing this study specifically.

*Response Q(3): We included the study by Ganayim et al. (2020) in the introduction on p. 4 and also included another study on Arabic adults (Hayek, Karni & Eviatar, 2020). Unfortunately, we could not find any studies on development of transcoding in Arabic-speaking children. The section was adapted as follows:

"Only small effects were reported on transcoding in older children and adults, both in languages with left-to-right writing systems (Authors et al., 2020a; van der Ven, Klaiber, & van der Maas, 2017) and right-to-left writing systems (Ganayim, Ganayim, Dowker, & Olkun, 2020; Hayek, Karni, & Eviatar, 2020). Decade-unit inversion was also reported to influence transcoding in a second (non-inverted) language. Native Arabic speaking adults made (some) inversion errors in their second (non-inverted) language of Hebrew, but hardly any in their first (inverted) language (Ganayim, et al., 2020) "

*Q(4): 3. I recommend a table with the authors predictions according to the models.

*Response Q(4):We added the table (Table 1 on p. 9) to clarify the number of expected transcoding procedures.

Table 1

Predicted Number of Transcoding Procedures by the ADAPT-model (Barrouillet, 2004) and the Model by Power and Dal Martello (1990)

| | Number of transcoding procedures | | | | | |
| | two-digit numbers | | three-digit numbers | | | |
Transcoding model	X0	XX	X00	X0X	XX0	XXX
ADAPT-model (Barrouillet, 2004)	2	2 (3)[a]	3	5	4	4 (5)[a]
Model by Power and Dal Martello (1990)	3	5	3	5	7	9

Note. [a] Shown in parentheses are numbers procedures if XX-numbers are algorithmically transcoded and not lexicalized

Q(5): 4. more details are needed regarding how the participants were chosen, initial math level (evaluated by teacher or screening test) since the study had a reading task the initial reading level is also important.

Response Q(5): All first-graders in the five schools in Austria and eleven schools in the UK were invited to participate. Children in Austria came from a middle-income urban school district. Children in the UK came from four urban, three town and four rural schools, with a mean deprivation index decile score of 8 (indicating the 30% of least deprived neighbourhoods, Department for Communities and Local Government, 2015) and an average of 11% of free school meals. Children in Austria and the UK were matched on experience with double-digit writing because in both countries formal instruction in the writing of double-digit numbers starts in Year 1. We have clarified this in the method section of the revised manuscript. We also added information on children's math and reading levels at the first assessment point (p. 13-14):

"Children completed selected items of the Numerical Operations subtest from the Wechsler Individual Achievement Test 2nd Edition (WIAT-II UK; Wechsler, 2005). Items were adapted for group use and assimilated to language-dependent notation of arithmetic operations. First, children had to master six items that involved identifying and writing Arabic digits to dictation and counting dots. Afterwards, they worked on nine standard arithmetic calculations (addition, subtraction, multiplication and division with one- to three-digit numbers) with increasing difficulty for 15 minutes. German-speaking children ($M = 11.29$, $SD = 1.47$) performed significantly better than English-speaking children ($M = 9.86$, $SD = 2.15$), $t(459) = 8.18$, $p < .001$.
We also individually administered language-specific word reading tests in which children had to read aloud a list of words as quickly as possible for 1 min (in German: SLRT-II, Moll & Landerl, 2010) or 45 s (in English: TOWRE-2, Rashotte, Torgesen, & Wagner, 2012). Both groups showed average percentiles compared to test norms (German: between 81th and 83th percentile, English:77th percentile)"

*Q(6): 5. stimuli, why the two-digit and three digit numbers did not contain the same number of stimuli? besides the xx category is larger (16) than the 1x (4) and x0 (4) this may have a confounding effect since the possibility of an error is larger. please explain that.

*Response Q(6): As we were particularly interested in the impact of inversion we focused on numbers containing decade-unit inversion. We added this point in the manuscript on page 12:

"In line with Zuber et al, we were interested in inversion-related errors, which is why we gave children more numbers that include inversion (1X, XX, XXX) than numbers that do not include inversion in German (X0, X00, X00, X0X, XX0)."

Of course, the reviewer is absolutely right that the number of "invertible" items in an item set influences overall error proportions. This point was already made in the earlier version of our paper (please refer to p. 6):

"Obviously, error types (and rates) depend on included number structures: e.g., decade-unit inversion is specifically required in XX- and XXX-numbers."

However, even if we balance the set according to individual number structures, it is somehow unbalanced according to items with decade-unit inversion (1X, XX, XXX) and without decade-unit inversion (X0, X00, X00, X0X, XX0).
That is why we rendered it important to include a control sample (English) to identify inversion-related errors as language-specific. Moreover, we felt it was essential to analyse error rates by number structure. Therefore, we extended the study by Zuber et al. (2009) and analyzed accuracy for individual number structures.

*Q(7): 6. I recommend to combine figures2a+b into one figure for more visibility.

*Response Q(7): The Figure 2 on p. 18 has been adapted according to the reviewer´s suggestion.

References
Bundesministerium Bildung, Wissenschaft und Forschung (2012). Lehrplan der Volksschule. Available at
https://www.bmbwf.gv.at/Themen/schule/schulpraxis/lp/lp_vs.html
Department for Education (2013). The national curriculum in England: Key stages 1 and 2 framework document. Available at:
https://www.gov.uk/government/publications/national-curriculum-in-england-primarycurriculum
Dowker, A., & Roberts, M. (2015). Does the transparency of the counting system affect children's numerical abilities? Frontiers in Psychology, 6, 945.
https://doi.org/10.3389/fpsyg.2015.00945
Ganayim, D., Ganayim, S., Dowker, A., & Olkun, S. (2020). Linguistic Effects on the Processing of Two-Digit Numbers. Open Journal of Modern Linguistics, 10, 49–69.

https://doi.org/10.4236/ojml.2020.101004

Habermann, S., Donlan, C., Göbel, S. M., & Hulme, C. (2020). The critical role of Arabic numeral knowledge as a longitudinal predictor of arithmetic development. Journal of Experimental Child Psychology, 193, 104794. https://doi.org/10.1016/j.jecp.2019.104794

Hayek, M., Karni, A., & Eviatar, Z. (2020). Transcoding number words by bilingual speakers of Arabic: writing multi-digit numbers in a units-decades inverting language. Writing Systems Research, 11(2), 1–15. https://doi.org/10.1080/17586801.2020.1787298

Johansson, B. S. (2005). Numeral writing skill and elementary arithmetic mental calculations. Scandinavian Journal of Educational Research, 49, 3–25. https://doi.org/10.1080/0031383042000302119

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. Biometrics, 33(1), 159. https://doi.org/10.2307/2529310

Moll K, Landerl K. (2010). Salzburger Lese- und Rechtschreibtest [Reading and Spelling Test] – second edition (SLRT-II). Bern: Hans Huber.

Simmons, F. R., Willis, C., & Adams, A. M. (2012). Different components of working memory have different relationships with different mathematical skills. Journal of Experimental Child Psychology, 111, 139–155. https://doi.org/10.1016/j.jecp.2011.08.011

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). Test of word reading efficiency – second edition (TOWRE-2). Austin, TX: PRO-ED, Inc.

Wechsler, D. (2005). Wechsler Individual Achievement Test – Second UK Edition (WIAT-II UK). London, England: Harcourt Assessment.

Second Decision Letter and Authors' Response

Dear Anna Steiner,

I have now received the comments by the two expert reviewers regarding your submission to Journal of Numerical Cognition, "Language effects in early development of number reading and writing". While both reviewers have suggested publication of your paper, reviewer 1 asked for a few clarifications. Please submit your responses to those comments and according revise your paper if necessary. I will check the responses and possible changes by myself without sending it back to the reviewers.

Best,
Mojtaba Soltanlou

Dear Dr. Soltanlou,

We would like to thank the reviewers and you for the positive feedback and helpful comments for correction or modification. The manuscript has been revised to clarify the comments of Reviewer 1, which are appended to this letter together with our responses. In the revised manuscript, changes are marked by using track changes.
We hope that the revised manuscript is now ready to be accepted for publication in the Journal of Numerical Cognition.

Kind regards,
Anna Steiner and co-authors

Reviewer A:

1. Page 4: I am unsure of what this means: "Inversion seems to have a particularly strong impact on early development." Do you mean that inversion has a particularly strong impact on performance in early development?

2. Thank you for your clarifications in response to this point: Page 8: I find it difficult to believe this statement- has a substantial literature review been completed? It is has please state this in the paper: "A further extension to Zuber et al. (2009) was that we did not only use number writing, but asked children to read the very same numbers aloud. Interestingly, this has not been done before." I now feel that this more fairly reflects the literature base.

3. Having now stated clear research questions address my query around hypotheses, thank you.

4. The discussion of this point is sufficiently addressed: Page 15: The standard deviations in the inversion-related errors results are huge and therefore warrant discussion.

5. Thank you for adding inter-rater reliability.
*Response: We accordingly changed the regarding sentence on page 4:
"Inversion seems to have a particularly strong impact on transcoding performance in early development."
We thank the reviewer for valuing our work and effort and are pleased that we could clarify all comments.

Reviewer B:

the manuscript is better integrated.

Third Decision Letter

Dear Anna Steiner and coauthors,

Thank you for your careful revision. Your article entitled "Language effects in early development of number reading and writing" has now been accepted for publication in the Journal of Numerical Cognition (JNC) – congratulations!

Best,
Mojtaba Soltanlou