PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Getting it Right from the Start (GIRFTS): protocol for a stepped-wedge cluster randomised controlled trial of a school-based framework to improve children's oral language and reading outcomes
AUTHORS	Quach, Jon; Siew, Melissa; Sinclair, Cecilia; Snow, Pamela; Eadie, Patricia; Poed, Shiralee; Shingles, Beth; Gold, Lisa; Orsini, Francesca; Connell, Judy; Edwards, Stuart; Goldfeld, Sharon

VERSION 1 - REVIEW

REVIEWER NAME	***Raghu Lingam***
REVIEWER AFFILIATION	University of New South Wales, School of Women's & Children's Department
REVIEWER CONFLICT OF INTEREST	
DATE REVIEW RETURNED	18-Feb-2024

GENERAL COMMENTS	Thank you so much for the opportunity to review this paper. The authors highlight a very important area for further research i.e. an intervention study to improved child reading and understanding of language for children in primary school aged children. The basis of the study is really interesting and important. However, I have a few major methodological queries that I hope the authors can address to allow publication of this important work. Major Points 1. The introduction needs to emphasis the primary and secondary outcomes of this work i.e. primary outcome reading and secondary outcome receptive language (i.e. understanding). This is clear in line 12-13 but needs to be highlighted elsewhere. The term the author uses is "poor oral language skills" this is expressive and receptive. There is no mention that the intervention will impact expressive language and this needs to be made overt in the introduction and the references used. 2. The order of the outcomes in the title, summary abstract etc needs to be changed. The intervention does NOT try and improve children's language skills it tries to improve 1. Reading and 2. Language comprehension. This should be made overt throughout. 3. Trial design – this is a major point. I don't think this can be defined as a step wedge trial. My understanding of the trial is that there are in summary two steps – intervention and delayed intervention. As such I do not think this can be considered a step wedge trial. See Hemming et al The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting (bmj.com) (fig 1). In many ways the design I think from my

	understanding of the paper is a (parallel) cluster randomised controlled trial with a delayed implementation of the intervention phase (for equity). As such it is unclear what the design is from the manuscript. I would therefore change this nomenclature throughout and clarify intervention, control and data collection periods in figure 1, table 1 and table 3. A step wedge trial needs far more steps. 4. Table 1 I think it would be useful to highlight in table 1 or table 3 when data will be collected i.e. baseline T1 and T2. These timepoints need to be specified in the outcomes i.e. at what timepoint is the primary outcome? 5. Table 3 needs timepoints 6. Implementation evaluation - this needs far more detail, a theoretical framework, metrics and theory. 7. Sample size justification of ICC and other parameters used in the sample size calculation needs to be justified. Did you get some data from your pilot? 8. Ethics – school based education ethics needs to be applied for or a statement needs to be made why it is not applied for. 9. I would add the logic model to the main text and highlight what data is being collected. In the logic model I would remove acronyms. 10. I don't think we need information and consent forms in the publication – would remove. 11. Would add a discussion/ saying how this work will add to the literature and reviewing the existing literature

VERSION 1 – AUTHOR RESPONSE

Reviewer Comment
1. The introduction needs to emphasis the primary and secondary outcomes of this work i.e. primary outcome reading and secondary outcome receptive language (i.e. understanding). This is clear in line 12-13 but needs to be highlighted elsewhere. The term the author uses is "poor oral language skills" this is expressive and receptive. There is no mention that the intervention will impact expressive language and this needs to be made overt in the introduction and the references used.
Our response: We have updated the introduction to be more explicit about the intervention and outcomes we are testing as part of this project.
2. The order of the outcomes in the title, summary abstract etc needs to be changed. The intervention does NOT try and improve children's language skills it tries to improve 1. Reading and 2. Language comprehension. This should be made overt throughout.
Our response: Thank you for your comments. We have modified our manuscript in the title and introduction to specify that this project aims to improve *oral* language and reading.

3. Trial design – this is a major point. I don't think this can be defined as a step wedge trial. My understanding of the trial is that there are in summary two steps – intervention and delayed intervention. As such I do not think this can be considered a step wedge trial. See Hemming et al The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting (bmj.com) (fig 1). In many ways the design I think from my understanding of the paper is a (parallel) cluster randomised controlled trial with a delayed implementation of the intervention phase (for equity). As such it is unclear what the design is from the manuscript. I would therefore change this nomenclature throughout and clarify intervention, control and data collection periods in figure 1, table 1 and table 3. A step wedge trial needs far more steps.
Our response: Thank you for your detailed feedback on our trial design. We appreciate your thoughtful consideration of the methodology. Allow me to clarify some aspects regarding the design of our study. Our study indeed follows a stepped wedge design, which involves an initial period where no clusters are exposed to the intervention, followed by regular intervals where groups of clusters transition from control to intervention. This process continues until all clusters have been exposed to the intervention. The inclusion of an initial period without intervention, subsequent steps of intervention introduction, and eventual exposure of all clusters to the intervention align with the characteristics of a stepped wedge design. Specifically, all schools in the study are in the control condition, i.e. business as usual, during school year 2021 (period 0). Then, one cohort (comprising 9 schools) switches to the intervention at the beginning of 2022 (period 1) while the other cohort continues with business as usual for the whole school year. At the beginning of school year 2023, the other cohort also switches to the intervention, while the first cohort continues delivering the intervention with it for another year (period 2, 2023). As suggested by Hemming et al The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting (bmj.com), we can in fact summarise our stepped wedge trial as follows: - the number of clusters = 18 - number and length of steps = 2 steps, length = 1 school year - number of clusters randomised at each step = 9 clusters/schools While we acknowledge that our study may not have as many steps as some traditional stepped wedge trials, it still adheres to the fundamental principle of staggered introduction of the intervention across clusters over time. Each transition from control to intervention represents a step in the implementation process, even if the number of steps is fewer than in other designs. Furthermore, our data collection strategy ensures that each cluster contributes observations under both control and intervention periods (students outcomes are collected at the end of each school year), allowing for a robust evaluation of the intervention's effectiveness over time. We will take your suggestion into account and provide additional clarity on the intervention, control, and data collection periods in our figures and tables to enhance the understanding of our study design. We have updated Figure 1 to this effect. Overall, while a parallel cluster randomized controlled trial with a delayed implementation and a stepped wedge design both aim to evaluate the effectiveness of interventions implemented over time, they differ in their approach to timing, control group composition, data collection, and statistical analysis. The choice between the two designs depends mainly on the research question, logistical considerations, and ethical considerations. Particularly, our trial is not a parallel cluster randomized controlled trial with a delayed implementation because our aim was to determine the impact of an **RTI intervention, compared**

with 'business as usual', in the first two years of school on students' oral language and reading outcomes. Had it been a parallel cluster randomized controlled trial with a delayed implementation, our aim would have been to determine the impact of an **immediate RTI framework, compared with a delayed RTI intervention**, in the first two years of school on students' oral language and reading outcomes.

While both a parallel cluster randomized controlled trial with a delayed implementation of the intervention phase and a stepped wedge design involve the implementation of interventions over time, there are key differences between the two:

1. Timing of Intervention Implementation:

- In a parallel cluster randomized controlled trial with delayed implementation, clusters are randomized to receive the intervention either immediately or after a delay. The delay allows for comparisons between groups that receive the intervention at different times (immediately vs delay).

- In a stepped wedge design, all clusters eventually receive the intervention, but the timing of implementation is staggered over multiple time periods. Clusters start as controls and transition to the intervention group at predefined time points, forming "steps" as the study progresses.

2. Control Group Composition:

- In a parallel cluster randomized controlled trial with delayed implementation, there are typically separate control and intervention groups. Clusters randomized to receive the intervention later serve as the control group.

- In a stepped wedge design, all clusters eventually receive the intervention, so there is no distinct control group throughout the study. Instead, clusters serve as their own controls before transitioning to the intervention group.

3. Data Collection Timing:

In both designs, data collection occurs before and after the implementation of the intervention. However, the timing of data collection relative to the intervention differs:

- In a parallel cluster randomized controlled trial with delayed implementation, data collection occurs at baseline (before the immediate and the delayed intervention is implemented) and then after the intervention phase for both groups.

- In a stepped wedge design, data collection occurs at multiple time points for all clusters, with some clusters serving as controls during early time points and transitioning to the intervention group at later time points. Data are collected before and after each cluster transitions to the intervention phase.

Thank you for your valuable input, which will help strengthen the clarity and accuracy of our manuscript.

4. Table 1 I think it would be useful to highlight in table 1 or table 3 when data will be collected i.e. baseline T1 and T2. These timepoints need to be specified in the outcomes i.e. at what timepoint is the primary outcome?
Our response: Thank you for your suggestion. Data collection points have been added in Table 1. A statement clarifying the time of primary outcome collection has also been added in section “Measures and data collection” > “Student outcome measures”.
5. Table 3 needs timepoints
Our response: Thank you for your feedback. Timepoints have now been added in the last column.
6. Implementation evaluation - this needs far more detail, a theoretical framework, metrics and theory.
Our response: Thank you for your feedback. Additional detail has been included in the manuscript. However, we are limited by the fact we have exceeded the word limit. An implementation framework (CFIR) and data collection tools and timepoints have been added to the description.
7. Sample size justification of ICC and other parameters used in the sample size calculation needs to be justified. Did you get some data from your pilot?
Our response: Thank you for your inquiry regarding the justification of the sample size calculation parameters, including the intraclass correlation coefficient (ICC) and other relevant factors. While we acknowledge the importance of justifying the selection of these parameters, it's important to note that coefficients like intraclass correlation coefficient (ICC) and cluster autocorrelation coefficient (CAC) are often not readily available in trial publications. In our case, despite efforts to gather prior information, we found a lack of data on ICC and CAC specific to our intervention and study population. Given this limitation on CAC, we followed the recommendation in the literature to assume a range of values for CAC, typically between 0.6 to 0.8, as this reflects common practice in similar studies. In many real-world settings, ICC values tend to be relatively low, especially in trials conducted in educational or healthcare settings. Research across various domains has shown that ICC values often fall within the range of 0 to 0.05, making 0.02 a reasonable estimate for many studies. Recognizing the impact that varying assumptions can have on sample size calculations, we conducted sensitivity analyses to assess the robustness of our results across a range of CAC and ICC values. After careful examination of these sensitivity analyses, we decided to adopt a CAC value of 0.8 and a ICC of 0.02 for our sample size calculation. This decision was informed by our aim to ensure adequate statistical power while accounting for potential clustering effects within the data. Unfortunately, due to the absence of pilot data specific to our intervention's context, we were unable to derive CAC estimates from a pilot study. Nevertheless, we believe that our approach of conducting sensitivity analyses provides a reasonable basis for the sample size calculation, considering the available information and the complexities of our study design.
8. Ethics – school based education ethics needs to be applied for or a statement needs to be made why it is not applied for.

Our response: Thank you for your feedback. School-based education ethics does not need to be applied for in Victoria. To conduct research in schools in Victoria HREC approval is required as well as approval from RISEC (relevant for Department of Education schools) and Melbourne Archdioceses Catholic Schools (MACS). For us to obtain both RISEC and MACS approvals we required approval from a HREC, which we receive from the Royal Children's Hospital HREC. We have made this clearer in our manuscript and provided the relevant application/study numbers.
9. I would add the logic model to the main text and highlight what data is being collected. In the logic model I would remove acronyms.
Our response: Thank you for your suggestion. We have moved the logic model into the main text (See Figure 2). Acronyms have been removed or explained. We had attempted to indicate in the logic model what data was being collected and when, however this made the figure more difficult to understand. Instead we have included a new table, Table 4, which highlights the study data that is being collected, what it is and when it is collected.
10. I don't think we need information and consent forms in the publication – would remove.
Our response: Thank you for your suggestion. These have been removed.
11. Would add a discussion/ saying how this work will add to the literature and reviewing the existing literature
Our response: Thanks for your comments. We have double checked and the protocol guidelines for BMJ Paediatrics Open and the journal does not include a discussion section for protocols. However, we acknowledge your point about the importance of adding commentary around how our study would add to literature and some additional literature and therefore have included these aspects into the introduction.

VERSION 2 – REVIEW

REVIEWER NAME	***Peter Flom***
REVIEWER AFFILIATION	Peter Flom Consulting
REVIEWER CONFLICT OF INTEREST	
DATE REVIEW RETURNED	08-Aug-2024

GENERAL COMMENTS	I usually confine my remarks to statistical and methodological aspects of articles, however, here, I recognized some other problems as well. It is good to see protocol papers and I commend BMJ Pediatrics for accepting them. Unfortunately, there are quite a few problems here, but, since it is a protocol paper, they should be fixable. The dates given in the paper are now in the past. That needs to be changed if it is indeed a protocol. But the years start with 2021, which concerns me.

	The introduction needs some editing. For one thing, correlation does not imply causation. So, the fact that lack of education is related to poor health outcomes does not imply that it causes them. In observational studies, it's impossible to firmly establish causation, but we can go some ways towards it by controlling for other variables. What variables were controlled for in the studies cited? Other issues here are more related to sentence structure and so on, making the text of the intro hard to follow. In the intro, the educational system in Australia should be very briefly described. Is "foundation" equivalent to "kindergarten"? What are the "first two years of school"? I'm not sure Table 1 is needed, it could certainly be reduced to two lines from 18. Also, the footnote does not match the table regarding the symbols. The last line on p. 5 is very troubling. It implies that only disadvantaged students are at risk of "learning difficulties" (whatever they are, they don't appear to be defined). This is not so. The risk is higher for them, but plenty of middle and above class kids have issues. On p. 6, we see that "data will be collected from the start of grade 2" (line 14) but then that the kids will be in "foundation and grade 1" (line 22). The first line in the randomization section is unclear. What are the sectors? Why variable block sizes? Near the bottom of p. 6 - different schools doing different things will add a lot of noise, making it harder to detect effects. What will be done with schools that decline? How about families that decline? Missing data is always a problem in studies like this. There is also going to be loss to follow up. How will this be dealt with? It seems to me that the missing data here may well be nonignroable nonresponse, which is hard to deal with. In table 3, the years changed (but are still in the past). Also, it would be good to include reliability and validity data. On p. 11, what about the number of kids? I'm guessing that different schools have different numbers of kids. This will surely affect power as it will affect the precision of the estimates in each school. Also, which student level variables will be added and how will the model be built? Sorry to be so negative. I do think this can all be fixed. Peter Flom

VERSION 2 – AUTHOR RESPONSE

Dr. Flom has provided thoughtful and detailed feedback to provide greater clarity and accuracy to our paper, and have enabled us to further refine the paper after the reviews provided by Dr Lingam in

May 2024. We are grateful for the opportunity to address the feedback in the attached file 'GIRFTS protocol_Response to reviewers_Flom'. In summary we have:

• Explained the history of the timing of the submission of this protocol

• Improved the language used in the introduction

• Provided definitions for terms that might be used differently in Australia compared to other parts of the world

• Updated the trial design table

• Included references for assessment reliability and validity

• Commented on how missing data will be handled and how we've considered the variation in the number of participants per school in the power calculation.

Thank you again for the opportunity to address these comments on our revised protocol paper.