



Universiteit
Leiden
The Netherlands

Protective teaching mechanisms in case of mild perinatal adversity

Merkelbach, I.

Citation

Merkelbach, I. (2018, November 7). *Protective teaching mechanisms in case of mild perinatal adversity*. Print Service Ede, Ede. Retrieved from <https://hdl.handle.net/1887/66720>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/66720>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/66720> holds various files of this Leiden University dissertation.

Author: Merkelbach, I.

Title: Protective teaching mechanisms in case of mild perinatal adversities

Issue Date: 2018-11-07

Chapter 7

General Discussion

Mild perinatal adversities such as being born late preterm or small for gestational age have long been considered vulnerability factors for academic success. However, a small scale study (Van der Kooy-Hofland, Van der Kooy, Bus, Van IJzendoorn, & Bonsel, 2012) established that kindergarten children with such mild perinatal adversities showed increased susceptibility to their literacy environment, both for better and for worse. In this dissertation we aim at both strengthening this body of knowledge through means of replication, and at expanding it by addressing new hypotheses. As a result this dissertation addresses the following issues:

1. Can findings of an earlier study on the effects of a digital program called *Living Letters* on the reading performance of young children with mild perinatal adversities be replicated, by using a larger sample size and a planned missing data approach?
2. If replication of initial program findings (partially) fails, are potential reasons for non-replication found in:
 - a) Heterogeneous responses to *Living Letters* by children born late preterm and children small for gestational age?
 - b) Sensitivity and quality of post testing?
 - c) Fidelity of implementation, possibly explained by teacher opinions on digital material?
3. Does *Clever Together*, a program comparable to *Living Letters* in substantive features, design, and duration and dosage, but targeting early numeracy instead of early literacy skills, yield similar results in children born late preterm, children born small for gestational age, and children without mild perinatal adversities?
4. Can we identify biological mechanisms possibly explaining differential effects?

To strengthen earlier conclusions, we aimed to replicate results of a small scale study (Van der Kooy-Hofland et al., 2012) in which kindergartners, who either were or were not subject to mild perinatal adversities (i.e. being born late preterm or small for gestational age), were exposed to a digital early literacy intervention (i.e. *Living Letters*). Initially, replication was sought through analyzing complete data of a new, larger study with a similar design as was used in the Van der Kooy-Hofland et al. (2012) study. When a partial lack of replicability was observed, we examined the utility of a planned missing data approach (Graham, Taylor, Olchowski, & Cumsille, 2006) to account for potentially diminished validity of outcome measures. As this approach did not fully explain the partial lack of replication, we shifted focus to the potential influence of the teacher on intervention effects, and to this end explored teacher opinions on the possibilities of digital interventions for supporting kindergarten children with learning vulnerabilities.

Van der Kooy-Hofland et al. (2012) found increased susceptibility in kindergartners

with mild perinatal adversities to the effects of a digital early literacy intervention, *Living Letters*, offering continuous feedback, guidance and support. Adversity-exposed children fell behind in the control condition, but outperformed their peers when assigned to *Living Letters*. Effects were found both in the short term (directly after the intervention) and in the long term (one year later). In the current study however this increased susceptibility was replicated only for children born late preterm, but not for those born small for gestational age. In the Van der Kooy-Hofland (2012) study the distinction between these two types of mild perinatal adversities was - and could not - be made, due to a small sample size and consequently a lack of power (MacCallum, Browne, & Sugawara, 1996). In addition to only partial replication, effects found in the current study were clearly more modest than earlier findings: in the Van der Kooy-Hofland (2012) study Cohen's d s exceeded 1.00 both in the long and the short term, while in the current study the values for Cohen's d in the late preterm group ranged between .30 and .40.

The differences between these two studies might originate in decreased quality of the outcome measure in the current study. In the Van der Kooy-Hofland (2012) study, researchers administered both pre- and posttests, while in the current study teachers administered these tests. Differences in level of researcher control might partly explain why effect sizes in the Van der Kooy-Hofland (2012) study and the current study differ so strongly: teacher-administered testing might have introduced a higher amount of error into the scores. To evaluate this assertion a planned missing data approach was used to account for possibly diminished validity of outcome measures without losing power (Graham, Taylor, Olchowski, & Cumsille, 2006): high quality reference (gold standard) measures were administered to a randomly selected subsample of children. Gold standard measures were expected to be of superior sensitivity compared to the general posttest, because these measures both comprised more items and were administered by trained research assistants instead of teachers. By ensuring that the subsample was randomly selected, data were missing completely at random (MCAR) (Garnier-Villarreal, Rhemtulla, & Little, 2014). By using the general posttest as an auxiliary variable to the gold standard, a shared variance factor of these measurements could be identified, which could be considered a valid estimate of performance. We found that results could be replicated using such an approach, however only when only gold standard measures closely approaching the skills targeted by the intervention were included in the model. Addition of gold standard measures that focused on related, but not similar, skills as those targeted by the intervention (e.g. vocabulary as opposed to alphabetic knowledge and phonemic awareness) led to poorer model fit. However, for results that *could* be replicated using such planned missing data models, effects were in the same ballpark as those found without the use of planned missing data and did thus not approach the effects found in the Van der Kooy-Hofland et al. (2012) study. Because suboptimal quality

of outcome measures in the current study does not seem to offer a valid explanation for the non-replicability of effect sizes of the Van der Kooy-Hofland et al. (2012) study, we might conclude that the original outcome measures used in the current study – as administered by the teachers - showed satisfactory sensitivity.

By using a planned missing data approach, we could thus conclude that lack of sensitivity of the outcome measure, possibly due to intrinsic qualities of the test or to administration of the test by the teacher instead of the researcher, did not explain why findings of Van der Kooy-Hofland et al.'s (2012) study could only be partly replicated. Teachers could however still be of influence, because they were not only in charge of pre- and post-testing but also managed the implementation of the intervention. To test if teacher characteristics were likely to influence the quality of implementation, teachers were asked to complete a questionnaire on their use of and opinions on digital (educational) material. Digital use and skill of the teacher, (perceived) results of digital material by the teacher, and teachers' expectations and beliefs regarding digital material, were all, directly or indirectly, associated with involvement in and effective implementation of such material. These associations show that teacher characteristics can predict quality of intervention implementation and thus suggest that teachers can influence study results in more ways than just through lower post-test fidelity. Presumably, either teachers have not been as steadfast in implementing the intervention at regular intervals as researchers would have been, or teachers have put children to work under suboptimal circumstances (e.g. noisy environment), which might have led to diminished results in the current study. Teachers who have negative beliefs and expectations about the potential effectivity of interventions might be more inclined to make such undermining choices. Additionally, teachers might have transferred negative expectations or opinions about digital interventions to their pupils, in turn possibly resulting in less investment by or motivation of pupils when working with *Living Letters*. Future studies should focus on the potential influence teachers have on the results of digital interventions. If the influence of the teacher indeed turns out to be substantial, motivating and training teachers for the use of digital material in the curriculum is of great importance.

Our study did not only aim at replication of previous results and explanation of deviations between these previous and current findings, but also at extending knowledge. Aiming to expand on the current body of knowledge, a series of new hypotheses was generated and tested. These addressed new, but related, areas of interest: we explored results of an intervention with an approach similar to *Living Letters*, but targeting a different academic area (i.e. early numeracy) in order to establish whether differential effects were early literacy specific, or if similar effects would also hold for other early academic skills. Furthermore we examined biological mechanisms possibly underlying differential effects found.

We looked into the effects of the digital program *CleverTogether*, highly similar in design to *Living Letters* but targeting another area of academic development in kindergartners, namely early numeracy. We found effects similar to those of *Living Letters*, suggesting that differential effects of supportive digital programs for late preterm kindergartners are not specific for early literacy but can apply to a broader range of academic skills. Additionally, we found support for the notion that children born late preterm show higher levels of stress reactivity as compared to their full term peers. This increased stress reactivity offers a plausible explanation for the good fit between highly supportive and guiding intervention programs, like *Living Letters* and *Clever Together*, and children born late preterm. These findings offer new insights and leads for new studies into the effectiveness of digital programs for children born late preterm and for other groups of stress reactive children. However, replication and further specification of these differential effects and their underlying mechanisms is necessary in order to optimally exploit the practical use of supporting digital programs in the daily classroom environment.

In conclusion, this dissertation shows that children born late preterm are susceptible to digital programs with features comparable to those of *Living Letters* and *Clever Together*, and that such programs can substantively contribute to the learning performance of these children. Additionally, we might assume that high levels of stress reactivity play a central role in explaining this increased susceptibility. However, especially in the light of the replication crisis (Maxwell, Lau, & Howard, 2015), we must be careful to draw firm and final conclusions. In order to fully understand why and under which circumstances digital programs can contribute to the learning gains of kindergartners born late preterm and/or kindergartners with high levels of stress reactivity, a lot of hypotheses remain to be formulated and repeatedly tested. We would therefore like to conclude with some recommendations and considerations for future researchers studying this topic, or using comparable designs to study the (differential) effects of interventions:

When studying intervention effects, differential effects for subgroups should be taken into account. Considering only main effects might disregard important information and might thus deny vulnerable subgroups valuable learning opportunities: both *Living Letters* and *Clever Together* did for example not show main effects, but proofed supportive for a subgroup of children. When considering such possibly susceptible subgroups, researchers should be aware that subgroups that seem highly similar at first sight (and might even fall under the same denominator: e.g. late preterm and small for gestational age are both mild perinatal adversities) might still react differently to their (learning) environment. Researchers should therefore formulate hypotheses about which (biological) mechanisms are bound to underlie susceptibility, before determining which subgroups of children might show increased susceptibility and should be studied and which interventions should be used. While still in the phase of generating such hypotheses, using a range of

small scale (e.g. pilot) studies to explore the effects of making small differences in study design, intervention design, and target group might be more suitable than using one large study (Van Teijlingen & Hundley, 2001) and testing only one specific hypothesis. When considering such small-scale studies, interpreting significance levels should not be the prime focus, because such studies are bound to have limited power. Instead, the direction of effects over a series of related studies should be examined to determine whether patterns can be observed, which could serve as the basis for formulating a well-founded 'final' hypothesis. At this stage large scale replication studies like the current project are of optimal value, because they can be used to test the ecological validity of the final intervention. Additionally, although researchers should be aware that using planned missing data is only useful when the gold standard measure closely approaches those skills specifically targeted by the intervention, the current dissertation shows that using planned missing data can be a valuable approach to diminish the influence of error in the outcome measure and could improve model fit. Making use of planned missing data designs thereby offers an easily accessible way to improve the quality of future experimental studies.

References

- Garnier-Villarreal, M., Rhemtulla, M., & Little, T. (2014). Two-method planned missing designs for longitudinal research. *International Journal of Behavioral Development*, 411-422.
- Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, 323-343.
- MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 130-149.
- Maxwell, S., Lau, M., & Howard, G. (2015). Is psychology suffering from a replication crisis? What does 'failure to replicate' really mean? . *American Psychologist*, 487-498.
- Merkelbach, I., Plak, R., & Rippe, R. (2018). Reproducibility of young learners' susceptibility to the learning context. *Learning and Individual Differences*, 167-175.
- Van der Kooy-Hofland, V. A., van der Kooy, J., Bus, A. G., van IJzendoorn, M. H., & Bonsel, G. J. (2012). Differential susceptibility to early literacy intervention in children with mild perinatal adversities: Short- and long-term effects of a randomized controlled trial. *Journal of Educational Psychology*, 337-349.
- Van Teijlingen, E., & Hundley, V. (2001). The importance of pilot studies. *Sociology at Surrey*.