

Food for Education Project phase 2

Implemented by Planet Aid International and ADPP Mozambique

Funded by the US Department of Agriculture

Midterm Evaluation Report

Author: Simone Doctors

with Forcier Mozambique and NFER

30 October 2019

Agreement Number: FFE-656-2015/009-00

Dr Simone Doctors

CHARTERED MEMBER OF THE CHARTERED INSTITUTE OF PERSONNEL AND DEVELOPMENT (CIPD)

11 Oakwell Drive, Leeds
LS8 4AE, United Kingdom
+44(0)113 2403308
+44(0)7523645522
simone@simonedoctors.com
www.simonedoctors.com



Table of contents

Acknowledgments	5
List of abbreviations and acronyms	6
Executive summary.....	7
Introduction.....	14
Background.....	15
Overview of the FFE2 Midterm Evaluation.....	17
Alignment with previous and future evaluations	17
Methodology	18
Information gathered for the midterm evaluation.....	18
Sampling for pupil-level data collection	20
Challenges, delays and lessons learned.....	26
Midterm findings	29
Improved literacy of school-age children (SO1)	29
Improved Quality of Literacy Instruction (MGD 1.1)	52
Access to School Supplies and Materials (MGD 1.1.2)	52
Literacy Instructional Materials (1.1.3).....	56
Skills and Knowledge of Teachers (1.1.4)	56
In-service training in literacy of primary school teachers.....	57
Initial teacher education in teacher training colleges (EPFs).....	79
Profiles of EPF trainers.....	81
Specialized subjects: trainer supply and demand.....	83
Skills and Knowledge of School Administrators (1.1.5)	87
Attentiveness (1.2)	88
Short-term hunger (1.2.1)	91
Access to food (school feeding) (1.2.1.1)	93

Student attendance (1.3)	99
School infrastructure (1.3.3)	103
Enrolment (1.3.4).....	104
Capacity of government institutions (1.4.1)	104
Government support (1.4.3).....	105
Engagement of Local Organizations and Community output (1.4.4)	105
Use of Health, Nutrition and Dietary Practices (SO2)	106
Knowledge of Health and Hygiene Practice (2.1)	106
Knowledge of Safe Food Preparation and Storage Practices (2.2)	109
Knowledge of Nutrition (2.3)	109
Access to Clean Water and Sanitation Services (2.4).....	121
Access to Clean Water	121
Access to Sanitation Services.....	123
Access to preventative health interventions (2.5).....	125
Access to requisite food preparation and storage tools and equipment (2.6).....	126
Capacity of government institutions (2.7.1)	126
Engagement of Local Organizations and Community output (2.7.4)	127
Transversal themes.....	128
Human Resources, capacity, collaboration and ownership.....	128
Transport	131
Administrative and financial systems and procedures	131
Monitoring and Evaluation (M&E).....	132
Sustainability and relevance to the local and national school feeding policy and program environment.	134
Conclusions.....	136
External evaluator's concluding remarks	138
Annexes	139
Annex 1: Results Framework of the Food for Education Project in Mozambique.....	139
Annex 2: McGovern-Dole Results Frameworks	140

Annex 3: Technical appendix 1	144

Acknowledgments

The evaluators gratefully acknowledge the many individuals and institutions who have contributed to this midterm evaluation report. Thanks are due to all the Planet Aid, ADPP, Humana People to People and Food for Education project leadership and staff who have contributed their time, effort and knowledge, supporting fieldwork, providing information and answering numerous queries. In particular, the new project M&E team who “hit the ground running” during the midterm evaluation made immense efforts to support the process, accompany field visits and obtain information on request. The helpful information and support provided by Cade Fields-Gardner, who supported the M&E team while key positions were vacant, is also gratefully acknowledged. The project leadership team were consistently supportive, helpful and skilled in finding solutions to problems and making things happen. Component leaders and staff were welcoming and flexible, providing valuable information and detailed explanations of their work practices. The evaluators sincerely thank all the individuals who welcomed evaluation visits, taking part in interviews and focus groups and sharing their lived experience of the project.

The contribution of all the representatives of the Government of Mozambique, at national, provincial and district levels, who supported the external data gathering process, made education and health sector data available and provided other invaluable inputs, is acknowledged with gratitude.

Grateful thanks are also due to the representatives of the World Initiative for Soy in Human Health and Cambridge Education for their support in developing evaluation tools, training data gatherers, providing and cross-checking large amounts of information, often at short notice, and reviewing the draft report.

Immense thanks are also due to Valerie Legg, for data verification.

List of abbreviations and acronyms

ADPP	<i>Ajuda de Desenvolvimento de Povo para Povo</i> Development Aid from People to People
ASA	American Soybean Association
BMI	Body Mass Index
CE	Cambridge Education
CSB+	Corn Soy Blend-Plus
DNUSE	<i>Direção Nacional de Nutrição e Saúde Escolar</i> National Directorate for Nutrition and School Health, MINEDH
DPEDH	<i>Direção Provincial de Educação e Desenvolvimento Humano</i> Provincial Directorate of Education and Human Development
EGRA	Early Grade Reading Assessment
EPC	<i>Escola Primária Completa</i> Complete primary school, which includes Grades 1-7
EPF	<i>Escolas de Professores do Futuro</i> (teacher training colleges run by ADPP)
FFE	Food for Education (project title)
FFK	Food for Knowledge (alternative project title)
GoM	Government of Mozambique
GPS	Global Positioning System
HGSFG	Home Grown School Feeding Garden
HIV/AIDS	Human immunodeficiency virus infection/acquired immunodeficiency syndrome
IMF	International Monetary Fund
INDE	<i>Instituto Nacional para Desenvolvimento da Educação</i> National Institute for Education Development
ISSET/OWU	<i>Instituto Superior de Educação e Tecnologia</i> One World University (ISSET/OWU)
M&E	Monitoring and Evaluation
MGD	McGovern-Dole International Food for Education and Child Nutrition Program
MINEDH	Ministry of Education and Human Development
NFER	National Foundation for Education Research
ODK	Open Data Kit (data collection software for resource-constrained environments)
PAI	Planet Aid, Inc.
PRONAE	<i>Programa Nacional de Alimentação Escolar</i> National School Feeding Program
QA	Quality assurance
SDEJT	<i>Serviço Distrital de Educação, Juventude e Tecnologia</i> District Office for Education, Youth and Technology
SDSMAS	<i>Serviço Distrital de Saúde Mulher e Acção Social</i> District Service for Women Health and Social Action
SEN	Special Educational Needs
SFC	School Feeding Committee
TB	Tuberculosis
ToC	Theory of Change
USAID	United States Agency for International Development
USDA-FAS	United States Department of Agriculture - Foreign Agricultural Service
WFP	World Food Program
WHO	World Health Organization
WISHH	World Initiative for Soy in Human Health
WPM	Words per minute

Executive summary

Phase Two of the 2015-2020 Food for Education project (FFE2) follows the 2012-2016 project of the same name (FFE1), supported and funded by the Foreign Agricultural Service (FAS) of the United States Department of Agriculture (USDA). A comprehensive school-feeding project, complemented by a major literacy component, it also includes child health, nutrition education, teacher training, water and sanitation components. The project is implemented in Mozambique by Planet Aid (PAI) and *Ajuda de Desenvolvimento de Povo para o Povo* (ADPP), Planet Aid's local partner, with partners World Initiative for Soy in Human Health program (WISHH) of the American Soybean Association (ASA), Cambridge Education (CE) and Mozambique's Ministry of Education and Human Development (MINEDH).

This midterm evaluation offers a "photograph" of the project in March-April 2019 in order to measure the project's "state of health". Key objectives of this evaluation are to permit comparison against the baseline evaluation to assess and document changes since the beginning of phase two; report how recommendations and lessons from phase one are being implemented; and make recommendations in relation to planning and implementation for the remaining 16 months of the project. While education access has generally improved, Mozambique faces high dropout rates, especially for girls, in the context of the combined challenges of political, economic and environmental crises (two major cyclones occurred during the preparation of this report). In addition to providing accountability to donors and other stakeholders and allowing learning and improvement, this report presents further information requested by project staff and managers.

There are three major components of the project, implemented by three different agencies: i) 88,100 students in 271 target schools in 4 districts of Maputo Province, receive a daily meal, consisting of nutritious porridge made from corn soya blend-plus (CSB+). School gardens and eight large-scale Home Grown School Feeding Gardens (HGSFGs) seek to encourage diversification and sustainability of school feeding. These "core" areas of the project are implemented by the FFE project team; ii) the project literacy team, with technical assistance from CE, is supporting literacy teaching in schools with materials developed to reinforce reading and writing skills for 1st, 2nd and 3rd grade children, teacher training and coaching; iii) a team managed by the WISHH is implementing a comprehensive nutrition education campaign in the beneficiary districts of Maputo province.

The midterm evaluation of FFE2 was conducted by a consortium composed of: a) Dr Simone Doctors (evaluation lead, qualitative data collection and analysis, main report author), b) Forcier Consulting (quantitative data collection) and c) the National Foundation for Educational Research (NFER; quantitative data analysis). The three members worked collaboratively under the coordination of Simone Doctors. To allow comparisons over time and build on lessons learned during previous evaluations, the FFE2 baseline survey conducted in 2017 was based on a newly developed evaluation plan; this commenced a new cohort study but continued to collect data on many of the same indicators as used in FFE1. Where possible, the FFE2 plan follows up on many of the research

questions used during FFE1, assessing these at three time points: baseline, midterm and end point. The pupil-level data at FFE2 baseline was collected from a new cohort of pupils, who can be expected to remain in primary school for the duration of the project lifespan, using a three-ways sampling frame developed by NFER; teacher-level data was collected from a new sample of teachers. The present midterm evaluation follows the progress of these cohorts of intervention and counterfactual pupils, and of the teachers and schools studied at baseline.

The midterm evaluation used a mixed method approach to gather and analyze a combination of quantitative and qualitative information consisting of: i) existing information provided by ADPP / the Provincial Directorate of Education and Human Development (DPEDH), and reviewed, collated and analyzed by Simone Doctors, ii) quantitative data collected by Forcier Mozambique and analyzed by NFER, and iii) qualitative data collected and analyzed by Simone Doctors. This information included:

- Project documents, surveys of (PAI/ADPP-run teacher training colleges (EPFs), pre- and post- Portuguese language tests of trainee teachers attending EPFs, DPEDH enrolment and dropout data, de-worming data;
- Literacy testing of 3,956 learners (from project schools and control/counterfactual schools), interviews and anthropometric measurements (weight and height) of the same 3,956 learners, a school survey of 170 project schools, a teacher survey of 190 teachers from project schools, classroom observation of teaching in 246 project and comparison schools, and statistical modelling (by NFER) to account for measured background differences between target learners and the comparison group;
- Focus groups with key stakeholder groups, semi-structured interviews with key stakeholders, visits and observation of a sample of schools, visits of the project headquarters and project warehouse, visits and observation of three EPFs, observation of food preparation and distribution processes, visits and observation of school gardens and HGSFGs, observation of teaching practice by EPF students, an internal self-review by the project implementation team, a stakeholder review workshop.

The midterm report is structured according to the logic of the project Theory of Change (ToC), with particular attention given to the two new components which began with the second phase: i) the program to improve literacy instruction in primary schools, particularly literacy in local languages, and ii) the HGSFGs, which strive to make school feeding more sustainable.

The project aims to improve literacy of school-age children in Mozambique (Strategic Objective 1) and to increase the use of health and dietary practices (Strategic Objective 2). In line with the project's ToC, interventions are designed to improve the quality of literacy instruction through training teachers in project schools in literacy, a significant addition in phase two of the project, giving learners improved access to school supplies and materials and other interventions to improve attendance. Interventions designed to improve health and dietary practice include nutrition education, school gardens and deworming. Achievement of both SO1 and

SO2 depend upon increased engagement of local organizations and community groups and increased capacity of government institutions. Improving school infrastructure and increasing enrolment and attendance, as a result of the school feeding program and other interventions to increase enrolment, is expected to further contribute to improving child literacy.

Assessing the literacy levels of students who have benefited from the literacy program supported by CE is key to evaluating the success of this intervention and of the wider project. At midterm, the target for students able to read and understand the meaning of grade level text by the end of two grades of primary schooling was 50% achieved by project students (67% by students benefiting from the literacy program in Portuguese). 92,237 individuals are benefiting directly from USDA-funded interventions, resulting in a midterm target 108% achieved. The number of individuals benefiting indirectly from USDA-funded interventions was 336,000, which achieves the project's target.

The cohort of pupils tested at baseline, when they were in grade 1, performed an Early Grade Reading Assessment (EGRA) comparable to that used at baseline. The results of each EGRA subtask were analyzed across the following comparisons: a) FFE + literacy (all) versus comparison (all); b) FFE + literacy (Portuguese) versus comparison (Portuguese); c) FFE + literacy (local languages) versus comparison (local languages); d) FFE + literacy (Portuguese) versus FFE only (Portuguese); e) FFE + literacy (Portuguese) versus FFE + literacy (local languages).

EGRA results demonstrate a low level of reading overall, with only 8% of the students benefiting from the literacy intervention meeting the national benchmark. However, taken overall, the results suggest that the literacy intervention is beginning to have an impact. Although the FFE + literacy group did not perform significantly better than the comparison group overall, in 8 of the 18 cases reported, of the students tested in local languages, the FFE + literacy group performed significantly better than the comparison group; there were no cases where the comparison (local languages) group performed significantly better than the FFE + literacy (local languages) group.

In 12 of the 18 cases reported, the FFE + literacy group performed significantly better than the FFE only group amongst students tested in Portuguese, providing promising early evidence that the literacy intervention is having a positive impact on the literacy acquisition of students within FFE project schools.

This finding provides promising early evidence of the impact of the literacy intervention, although more time may be needed for the effects of the intervention to translate into more systematic results. The final evaluation of FFE2 in 2020 will allow the initial achievements reported here to be tested after an additional year of the literacy intervention and cast more light on how the initial progress made translates into later literacy. For the students taught in local languages, the final evaluation will capture how literacy instruction begun in local languages then transitions into literacy in Portuguese, the national language of instruction.

The project ToC encapsulates the assumption that the quality of literacy instruction will be improved by a) better access to school supplies and materials and b) increased skills and knowledge of teachers. At midterm, 71,932 textbooks and other teaching and learning materials had been provided as a result of USDA assistance, resulting in 182% achievement of the final target. 264 schools had received school supplies and materials as a result of USDA assistance, resulting in 103% achievement of the final target. The project has contributed to teachers and pupils having improved access to teaching and learning materials.

Awards have been provided to both pupil and teachers since the beginning of the project to encourage both to make efforts in teaching and learning. The final target for awards to teachers has been surpassed, and the midterm target for awards to pupils has been 90% achieved.

A range of literacy books and materials in local languages Xichangana and Xirhonga and in Portuguese has been produced by the project literacy team, and targets for distributing these have been surpassed. The midterm evaluation revealed widespread praise for these innovative materials, for the quality of their content and their culturally appropriate messages.

Two major dimensions of the project interventions target improving the skills and knowledge of teachers: 586 teachers in project schools have received training in teaching reading and writing using phonics - a significant innovation for Mozambique, where phonics have not been commonly used. Lessons were observed in 240 project schools and comparison schools in order to attempt to evaluate whether teachers in project schools were using “new and quality teaching techniques or tools” as a result of the literacy training. FFE + literacy teachers were more likely to exhibit these characteristics than FFE only teachers or comparison group teachers; FFE only and comparison group teachers were almost equally likely to exhibit these characteristics; a higher percentage of teachers using local languages demonstrated the characteristics than did teachers teaching in Portuguese.

Prior evaluations of the FFE project have underlined the important numerical and quality contribution the eleven ADPP-run teacher training colleges (EPFs) are making to Mozambique’s teacher pool. At the time of the midterm evaluation, 6,059 trained teachers had graduated during the current phase of the project, well exceeding the target. The EPF teacher training is providing significant added value in Portuguese language competence through remedial support to student teachers who arrive at the EPF with poor levels of Portuguese. The EPFs provide a culture in which a “different sort of teacher” may be formed, with a focus on the more holistic aspects of education, which provides trainees with skills, attributes and attitudes appropriate to the realities of teaching in low-resource rural schools. The target for the number of school administrators and officials in target schools who demonstrate use of new techniques or tools as a result of USDA assistance, has been 110% achieved; the target for the number of school administrators and officials trained or certified as a result of USDA assistance has been 161% achieved.

Short-term hunger leads to lack of attentiveness in the classroom, leading to problems in learning. At midterm, 11% of teachers report increased attentiveness of students in the classroom and 6% of students report being more attentive in the classroom. 42% of pupils in intervention schools (FFE + literacy and FFE only combined) said they sometimes felt hungry in the classroom, compared with 58% of students in comparison schools. These levels of hunger are lower than the baseline levels for both groups. 23% of pupils in intervention schools (FFE + literacy and FFE only combined) said they felt hungry during the interview, compared with 33% of students in comparison schools.

The distribution of school meals is on target. At midterm, 28,162,260 daily school meals (breakfast, snack, lunch) were provided to school-age children as a result of USDA assistance, which is 104% of the midterm target. 84,026 school-age children were receiving daily school meals, 114% of the midterm target. When students missed school meals, the principle reason for non-feeding was pupil absence. Of reasons attributable to the project itself, the main reasons for schools not feeding are lack of CSB+ and lack of cooks.

During school visits for the midterm evaluation, as during other evaluations of the project, teachers, school directors, parents, government officials and other stakeholders consistently claimed that school feeding and the other project interventions had led to increases in student attendance.

One project activity intended to promote increased school attendance is support to extra-curricular learning clubs. Of the schools surveyed at midterm, 89% reported having one or more learning clubs. 113% of the final target has been achieved for the number of after-school learning clubs active in project, and 109% of the final target for the number of school children participating. In accordance with a recommendation made in the baseline report, this evaluation found that in project schools, pedagogical directors and teachers responsible for clubs have been trained in the facilitation of extracurricular clubs, play-based and other activities to promote learning and in the use of provided materials.

Targets for maintaining the school infrastructure provided by the project have been surpassed, with 1,002 maintenance interventions performed in kitchens, storerooms, and firewood-saving stoves at midterm, or 127% of target.

The project ToC considers enrolment as a prerequisite to attendance and therefore to improved learning. Targets for enrolment of students have been surpassed, and informants overwhelmingly claimed that enrolment had increased, and dropouts decreased as a result of the projects' activities, particularly school feeding. It was not possible to verify this claim using information provided by the DPEDH, due to inconsistencies in the data.

The project's second Strategic Objective (SO2) concerns the use of health, nutrition and dietary practices. Targets for training in good hygiene practice, including training volunteer cooks in health and hygiene, have been greatly surpassed. In regard to the percentage of students that demonstrate acceptable knowledge of health

and hygiene practices, the final target has been 170% achieved; the midterm target for food preparers at target schools trained in hand washing, safe food preparation and storage practices has been 118% achieved. FFE2 targets for training individuals in child health and nutrition have been greatly surpassed.

School gardens have been encouraged and supported since the inception of the project, and the final target for the number of school children benefiting from school gardens has been achieved. 8 new HGSFGs have been established; these are an innovation of FFE2 to address the question of sustainability of school feeding after the end of the project activities. Access to safe water continues to be a condition for school feeding, local production of food, health and wellbeing; despite the water team's best efforts, this also continues to be a challenge for some project schools. The final target for the number of schools using an improved water source has been achieved; however, in many schools the absence of safe water some or all of the time is a major problem.

The evaluation provides further analysis of transversal themes which cut across the different components and activities and concern the overall operations, strategy and vision of the project. These are Human Resources, capacity, collaboration and ownership; Transport; administrative and financial systems and procedures; monitoring and evaluation; and sustainability and relevance to the local and national school feeding policy and program environment.

Most activities planned have been implemented as intended, often in the face of considerable obstacles, and most midterm targets have been met or surpassed, with a number of final targets met at midterm; in the small percentage of cases where targets have not been met, these are likely to be met by the end point. The exception is the target concerning the percentage of school children meeting the literacy benchmark; it may be that the target set was just too ambitious to be achieved within such a short time, given the low baseline levels.

Unlike many projects operating in the extremely challenging context of the Mozambican education system, FFE2 is largely managing to do what it set out to do within the planned timeframe. It is recommended that project staff and leaders (and evaluators) remain aware of the reasons why given activities are conducted, what positive change they are intended to bring and how they are experienced by all those involved in their implementation, including but not limited to the beneficiaries. Project interventions are beginning to make a difference: teachers are beginning to use improved methods in their teaching; students are beginning to read better; community volunteers in the HGSFGs are beginning to apply new methods to produce large amounts of food; thanks to this produce, schools are beginning to diversify the menus consumed by students; school water committees are beginning to take responsibility for maintenance of their water system. While project interventions are beginning to make a difference, more time is needed in order for this initial change to be consolidated and integrated into durable behavioral changes.

The project applies many of the principles of good international development practice and is providing services and skills which are highly relevant to both the beneficiary population and the country as a whole. While it is

beginning to produce impact in the form of positive change, questions remain about the sustainability of these achievements in the current Mozambican context. Whether there is an extension to the current project, or the remaining period is spent implementing an exit plan to prioritize activities which can be continued after the close of the project, maximizing future sustainability is now imperative.

Every effort should be made to prolong the FFE2 project to capitalize on the human and material investment made so far.

Introduction

The 2015-2020 Food for Education project (phase 2 of the 2012-2016 project of the same name, or FFE2)¹² is supported and funded by the Foreign Agricultural Service (FAS) of the United States Department of Agriculture (USDA). FFE2 is a comprehensive school-feeding project with child health, nutrition education, teacher training, water and sanitation components, complemented by a major literacy component. The project is implemented in Mozambique by Planet Aid (PAI) and *Ajuda de Desenvolvimento de Povo para o Povo* (ADPP), Planet Aid's local partner in Mozambique, in partnership with the World Initiative for Soy in Human Health program (WISHH) of the American Soybean Association (ASA), Cambridge Education (CE) and the Ministry of Education and Human Development (MINEDH). The project is intended to serve as one of the models for the PRONAE, the national school feeding program being developed and implemented by the Government of Mozambique (GOM) with collaboration and assistance from the World Food Program (WFP) and other partners.

The midterm evaluation seeks to measure the “state of health” of the project at midterm, offering a “photograph” of phase 2 of the project in March-April 2019. This allows:

- comparison against the baseline evaluation to assess and document changes since the beginning of phase 2;
- monitoring of how recommendations and lessons from phase one are being implemented;
- recommendations to be made in relation to planning and implementation for the remaining 16 months of the project.

Like the baseline report for phase 2, this midterm evaluation report:

- accounts to the project donors and other stakeholders for how resources have been used;
- collects and analyzes information identified as important by project staff and managers to allow them to make informed decisions about their operations;
- identifies lessons and makes recommendations for the remainder of the project and beyond, including in relation to sustainability.

Despite progress in improving access to education, education quality remains a cause for concern in Mozambique. Dropout rates are high, with over half the girls who enroll in primary school dropping out by the fifth grade, only 11 percent continuing to secondary level, and 1 percent to college (USAID, 2019). Of those who

² often referred to internally as Food for Knowledge (FFK2) or *Comida para o Saber*, in Portuguese.

finish primary school, “nearly two-thirds leave the system without basic reading, writing, and math skills” (USAID, 2019)³.

The overall context in Mozambique has deteriorated since the baseline report. Residents speak of a three-fold political, economic and environmental crisis. The ongoing saga of the illegal debt has led to the suspension of much international development aid. Armed conflict in certain areas of the country is leading to violence and population displacement. Extreme weather events disrupt daily life for many citizens: during the preparation of this report, between March and April 2019, cyclones Idai and Kenneth both led to widespread flooding, loss of life and infrastructure and population displacement.

Background

The FFE 2 project aims to promote the health, wellbeing and education of school aged children, through holistic interventions combining school feeding with health, water, sanitation, literacy and nutrition components. Based on the Theory of Change captured in the McGovern Dole Results Framework (see Annexes1 and 2), the project’s Strategic Objectives are to improve literacy of school-age children in Mozambique (SO1) and to increase the use of health and dietary practices (SO2). In order to achieve these Strategic Objectives, interventions are structured according to a hierarchy of assumptions about the conditions leading to change.

Within SO1:

- the quality of literacy instruction is improved by providing intensive, targeted literacy training to teachers in project schools;
- the skills and knowledge of primary school teachers are increased through on the spot coaching of in-service teachers within primary schools⁴;
- providing teachers and learners with improved access to school supplies and materials contributes to improved literacy instruction;
- school feeding reduces short-term hunger in the classroom, leading to improved student attentiveness and therefore learning;
- student attendance is improved through a combination of improved school infrastructure (including kitchens, water and sanitation);

³<https://www.usaid.gov/mozambique/education>

⁴This dimension was a significant addition to Phase 2 of the project. Evaluations of Phase 1 revealed that, although attendance and retention rates and reported pupil concentration in the classroom had improved since the onset of school feeding, in the absence of improved teaching, educational outcomes were unchanged.

- increased enrolment results from students having increased access to food increased as a result of school feeding.

Within SO2:

Use of health and dietary practices increases through:

- improved knowledge of health and hygiene practices, such as hand washing;
- increased knowledge of safe food preparation and storage practices, for example through training volunteer cooks;
- increased knowledge of nutrition, through nutrition education and access to school gardens;
- increased access to clean water and sanitation (latrines);
- increased access to preventative health interventions, such as de-worming and TB screening;

Critical to the achievement of both SO1 and SO2 are increased engagement of local organizations and community groups and increased capacity of government institutions.

In order to bring about the desired changes, the project consists of three major components, implemented by different agencies:

1. School feeding, water supply development, school gardens, and related activities

The project strives to provide 88,100 students in 271⁵ target schools in 4 districts of Maputo Province, Mozambique (Moamba, Magude, Manhiça and Matutuine) with a daily meal, consisting of nutritious porridge made from corn and soya. School gardens and two large-scale Home Grown School Feeding Gardens per district aim to encourage diversification of students' diet contribute to sustainability of school feeding after the end of the project. Schools are provided with a safe and adequate school water supply, latrines and hand-washing facilities. Students receive regular de-worming. Actions to provide an improved learning environment, enrolment and attendance for all pupils include after-school clubs and kits of educational and recreational materials. These "core" areas of the project are implemented by the FFE project team.

2. Early grade reading and writing intervention

FFE2 includes an innovative intervention to support literacy teaching, in both Portuguese and local languages (Xirhonga or Xichangana)⁶. Schools benefiting from the literacy intervention receive specially developed

⁵ Significantly increased from the original proposal and contract with USDA to provide meals to 74,000 students in 264 schools.

⁶ Children in the beneficiary districts speak one of several local Mozambican languages when they enter school, but they have traditionally been taught to read in Portuguese, the official national language. Approaching early literacy in a language which children do not speak or understand (and which many teachers master imperfectly) creates a major obstacle in their acquisition of reading and writing skills as well as in their wider education.

materials to reinforce reading and writing skills for 1st 2nd and 3rd grade children. Teachers receive training in how to use the new methods and reading coaches provide ongoing support to the trained teachers, through classroom observation and lesson coaching. The literacy intervention is implemented by a dedicated team (referred to hereafter as “the project literacy team”) supported by CE.

3. Nutrition education program

Since the onset of FFE1, the project has implemented a comprehensive nutrition campaign in the 4 beneficiary districts of Maputo province and the 11 teacher training colleges run by ADPP (the *Escolas de Professores do Futuro*: EPFs). Teachers currently teaching in primary schools are trained as Nutrition Educators, known as “coaches”, and train other primary school teachers in the schools in which they work. Nutrition education is an integral part of the teacher training provided by the EPFs. Nutrition and health educational materials and information documentation have been produced. The Nutrition Education program is implemented by a team based in the project office but managed by the World Initiative for Soy in Human Health (WISHH).

Overview of the FFE2 Midterm Evaluation

Alignment with previous and future evaluations

Regular evaluations of the FFE project have been conducted, in accordance with the project Evaluation Plan⁷:

- a baseline study of FFE1 in May 2013 by project staff with support from an external evaluator;
- a midterm evaluation of FFE1 in 2014 by an external evaluation team;
- a final evaluation of FFE1 in late 2016 (report dated July 2017) by an external evaluation team;
- a baseline study of FFE2 in 2017 (report dated July 2018) by project staff with support from an external evaluator (to avoid redundancy and in order to make the best use of resources, the baseline of FFE2 built on the final evaluation of FFE1).

The current evaluation is of the FFE2 project at midpoint, without direct reference to FFE1.

The FFE2 evaluation plan builds on that of FFE1. For the sake of coherence, to allow comparisons over time and build on lessons learned during previous evaluations, the FFE2 baseline survey was based on a newly developed evaluation plan, commencing a new cohort study but continuing to collect data on many of the same indicators as used in FFE1. Where possible, the FFE2 plan follows up on many of the research questions used during FFE1, assessing these at three time points: baseline, midterm and end point. Alignment between evaluations of FFE1

Bilingual education seeks to introducing reading and writing in mother tongue, whilst simultaneously learning the language of instruction, into which teaching and learning transitions in grade 3.

⁷The reports of these evaluations are available on request.

and FFE2 has contributed to the amassing of a considerable body of evaluation evidence which is relevant to the current project.

The pupil-level data at FFE2 baseline was collected from a new cohort of pupils, who can be expected to remain in primary school for the duration of the project lifespan, using a three-ways sampling frame developed by NFER; teacher-level data was collected from a new sample of teachers.

The present midterm evaluation of FFE2 measures progress against the baseline in achieving objectives; it formulates recommendations to improve effectiveness and efficiency for the remainder of the project and to ensure sustainability thereafter. The indicators and methods used are based on those of the FFE2 baseline, to follow the progress of the cohorts of intervention and counterfactual pupils and of the teachers and schools studied at baseline.

A final external evaluation of the second phase will be conducted at the end point of the project in 2020, in order to assess the project's relevance, effectiveness, efficiency, impact and sustainability. Using the same indicators, measures and methods, the final evaluation of FFE2 will assess whether the project objectives have been achieved, what has changed in the lives of the project beneficiaries and whether recommendations made following the midterm evaluation have been integrated and implemented.

Methodology

The midterm evaluation used a mixed method approach to gather and analyze a combination of quantitative and qualitative information. The methods used built on those used at baseline (and, to some extent, during the evaluations of FFE1). These will also form the basis of the approach to be used for the final evaluation of FFE2, ensure comparability and consistency of the information gathered at the three time points and evaluate change over the duration of the project.

Information gathered for the midterm evaluation

Three types of information were gathered and analyzed:

Existing information provided by ADPP / DPEDH, and reviewed, collated and analyzed by Simone Doctors

- **Project documents:** project records, in the form of individual record level data in unanalyzed /non-tabulated form (for example, participant lists from training courses); project documentation; training materials produced; reports (for example the semi-annual Logmon reports submitted to USDA) and other monitoring information (for example the monthly indicator tracking files and commodity tracking files), were provided by ADPP and WISHH to the external evaluator for analysis upon request. The external evaluator reviewed these sources extensively, extracting elements to analyze and engaging in an iterative conversation with the project leadership and M&E teams to discuss these and reconcile records as

necessary. During visits to the project headquarters, she performed spot checks on records and evidence, asking M&E staff to produce different types of information and to demonstrate the systems for gathering and archiving these.

- **Surveys of EPFs** (PAI/ADPP-run teacher training colleges) regarding infrastructure and equipment, profiles of teaching staff and staffing levels relative to teaching needs. The instruments used at baseline were updated as necessary by the external evaluator and sent to all EPF directors to complete, through the ADPP EPF focal point.
- **Results of pre- and post- Portuguese language tests** of trainee teachers attending EPFs showing pre-test and post-test scores for all trainee teachers in 6 of the EPFs, allowing their performance before training began to be compared with that after one year's training over the course of 2015; this information was gathered by 6 EPFs at the beginning and end of the 2018 academic year and made available to the external evaluator by the ADPP EPF focal point.
- **Enrolment data** provided by the DPEDH Maputo for all districts of Maputo province for 2012, 2013, 2014, 2015, 2016, 2017 and 2018;
- **Dropout data** provided by the DPEDH Maputo for all districts of Maputo province for 2012, 2013, 2014, 2015, 2016, 2017 and 2018;
- **Data on de-worming** provided by the district health services in the intervention districts, either directly to the external evaluator or through the intermediary of the project district coordinators. This was collated and analyzed by the external evaluator

Quantitative data collected by Forcier Mozambique and analyzed by NFER

- **Literacy testing** of 3,956 learners (from project schools and control/counterfactual school) using the Early Grade Reading Assessment (EGRA) used at baseline with minor modifications;
- **Interviews** with the same 3,956 learners, using the instrument used at baseline with minor modifications;
- **Anthropometric measurements (weight and height)** of the same 3,956 learners to establish BMI z scores;
- **School survey** of 174 project schools, using a reduced version of the instrument used at baseline;
- **Teacher survey** of 190 teachers from project schools, using a slightly modified version of the instrument used at baseline;
- **Classroom observation** of teaching in 246 project and comparison schools, using a standardized observation sheet;
- **Statistical modelling (NFER)** to account for measured background differences between target learners and the comparison group, allowing potential changes associated with the program to be measured.

Qualitative data collected and analyzed by the external evaluators ⁸

- **Focus groups** with key stakeholder groups: learners, teachers, parents, volunteer food preparers, school-feeding committee members, student teachers, teacher trainers, project “professionals” (the project field officers);

⁸ Ten of the focus groups (two focus groups each with female pupils, male pupils, parents, volunteer cooks, school feeding committee members) were conducted by Forcier since they required the use of local languages;

- **Semi-structured interviews** with key stakeholders: school heads, community representatives, government education officials (national, provincial and districts level), district health officers, EPF heads, project component managers or team representatives, representatives of USDA, representatives of PAI, representatives of WISHH, representatives of CE, representatives of other organizations providing school feeding, other partners);
- **Visits and observation** of a sample of schools in all four project districts, using a standardized observation framework and photographic recording of infrastructure;
- **Visit** of the project headquarters in Manhica, including observation and discussion of working practices and M&E systems;
- **Visit** of the main project warehouse in Manhica;
- **Visits and observation** of three EPFs (Tete, Gaza and Maputo)⁹;
- **Observation** of food preparation and distribution processes, using a standardized observation framework and photographic recording;
- **Visits and observation** of school gardens and Home Grown School Feeding Gardens in all four project districts;
- **Observation** of teaching practice by EPF students and of the support and feedback provided by EPF tutors and school mentors;
- **Results of an internal self-review** by the project implementation team, facilitated by the external evaluator on 22 February 2019 and attended by approximately 40 project leaders and staff, including the entire leadership team;
- **Results of a stakeholder review workshop** to validate initial findings and suggest corrections or changes in strategy for phase 2 of the project. This was held on 10 April 2019 and attended by approximately 50 project leaders and staff, including the entire leadership team and ADPP's Executive Director. The initial findings as presented were recognized, owned and well received and generated invaluable comments and responses from project leaders and staff.

Sampling for pupil-level data collection

The evaluation plan for FFE2 is conceived as a quasi-experimental cohort study at three time points. The collection of quantitative data at midterm was based on the stratified sampling frame of schools drawn up by experts in planning statistical analysis for education studies from NFER and already used at baseline. Since almost all schools in the four project districts (Manhica, Magude, Moamba and Matutue) are benefitting from the project interventions, the control group schools were selected from the four non-project districts (Namaacha, Matola, Marracuene and Boane).

An explanation of the selection of the sample is summarized below; more details can be found in the FFE2 baseline report.

all other qualitative information was collected by Simone Doctors using Portuguese, except in a small number of cases where respondents preferred to speak English. All translations of quotes from interviews are by Simone Doctors.

⁹ The sample of EPFs visited aimed to achieve a geographical spread and balance between those EPFs considered by WISHH to be implementing the Nutrition education program well. The choice was constrained by logistical considerations and time constraints and by the disruption caused to two EPFs by cyclones in March and April 2019.

- Schools in the project districts not receiving the intervention were excluded
- Schools in the comparison districts known to be receiving a different feeding program were excluded
- Xichangana and Xirhonga were combined into one “local language” group.
- The sampling frame contains five categories of schools:
 - FFE+literacy (local language)
 - FFE+literacy (Portuguese)
 - FFE+literacy (local language and Portuguese)
 - FFE only
 - Comparison
- It was not possible to achieve perfect matching between the categories, since the sample was drawn based on the schools actually benefiting from the project, rather than based on their characteristics (as would be the case in a randomized control trial). The project districts were selected to benefit from the project on the basis of their relative poverty and food insecurity; the comparison districts are likely to have lower levels of socioeconomic deprivation than the intervention districts.

Whereas at baseline, pupils in grade 1,2 and 3 were tested, the midterm only followed up on the cohort of 16 pupils per school, who had been in grade 1 at baseline and, for the most part, were in grade 3 at midterm (where schools benefit from literacy interventions in both Portuguese and the local language, two cohorts of 16 pupils were tested: one from the class learning in Portuguese, the other from the class learning in the local language).

At baseline, 4,150 grade 1 students were tested: 1,887 in the FFE + literacy group, 1,104 in the FFE only group and 1,159 in the comparison group¹⁰. 115 students tested were removed from the sample at midterm, following verification of the intervention groups, which revealed that these students no longer qualified to be part of the sample, for example because they had been part of the FFE only group but their school had begun receiving the literacy intervention between baseline and midterm. At midterm, the FFE only (bilingual) group was removed from the study, since there are no longer project schools teaching the bilingual curriculum which are not receiving the literacy intervention. At midterm, therefore, there was one FEE only group, tested in Portuguese. Of the remaining students, 2,819 were successfully recontacted at midterm: 1,180 FFE + literacy; 668 FFE only and 971 comparison. Data gatherers were unable to recontact 902 at midterm, for a variety of reasons, including leaving school, changing schools or residence, absence from school on the day of data gathering, and death: these were deemed “lost” to the project. This high level of attrition had been anticipated and the sample drawn accordingly. Where possible, the lost students were replaced with substitutes (628 FFE + literacy, 284 FFE only and 110 comparison), using a rigorous sampling protocol based on the sampling method used at baseline. This information is summarized in table 1 below.

¹⁰ Throughout the report, the terms “FFE + literacy”, “FFE only” and “Comparison” are used, without quotation marks or italics, to differentiate teachers, students and schools belonging to the three groups.

Students who participated at baseline and at midterm				
Row Labels	Lost	Recontacted	Replacement	Grand Total
Baseline	902	2819		3721
Comparison (Bilingual)		89		89
Comparison (Portuguese)	165	882		1047
FFE + lit (Bilingual)	8	530		538
FFE + lit (Portuguese)	390	650		1040
FFE only	339	668		1007
Midline		2819	1022	3841
Comparison (Bilingual)		89	20	109
Comparison (Portuguese)		882	90	972
FFE + lit (Bilingual)		530	231	761
FFE + lit (Portuguese)		650	397	1047
FFE only		668	284	952

Table 1: Students who participated in the baseline and midterm evaluations

Schools within the sample were either classified as “bilingual”, meaning they teach in both a local language (Xichangana or Xirhonga) and Portuguese, or “Portuguese”, meaning they teach only in Portuguese. Students from the bilingual schools are tested in local languages and students from Portuguese schools are tested in Portuguese (cohorts of 16 per school, in each case). However, in some bilingual schools, both local language and Portuguese students are tested (so 2 cohorts of 16 students are tested, one in Portuguese, the other in local languages). For the purposes of the evaluation plan, bilingual students from Moamba, Magude and Manhiça are tested in Xichangana and those from Matutuine in Xirhonga¹¹. For details of the number of schools within each group within the sample, see table 2 below. The groups are not balanced, due to the relative scarcity of schools delivering the bilingual program, compared to those teaching in Portuguese (see figure 1 below).

¹¹ This decision is based on the languages taught by the project in these districts. In reality, the distribution of the two languages does not follow the district boundaries but there is some overlap. Furthermore, Xichangana and Xirhonga are mutually intelligible and speakers of Xichangana frequently use words and expressions from Xirhonga and *vice versa*.

Groups	Number of schools
Comparison (Bilingual)	5
Comparison (Portuguese)	67
FFE + lit (Bilingual)	42
FFE + lit (Portuguese)	51
FFE only	71
Grand Total	236

Table 2: Numbers of schools in the evaluation sample at midterm

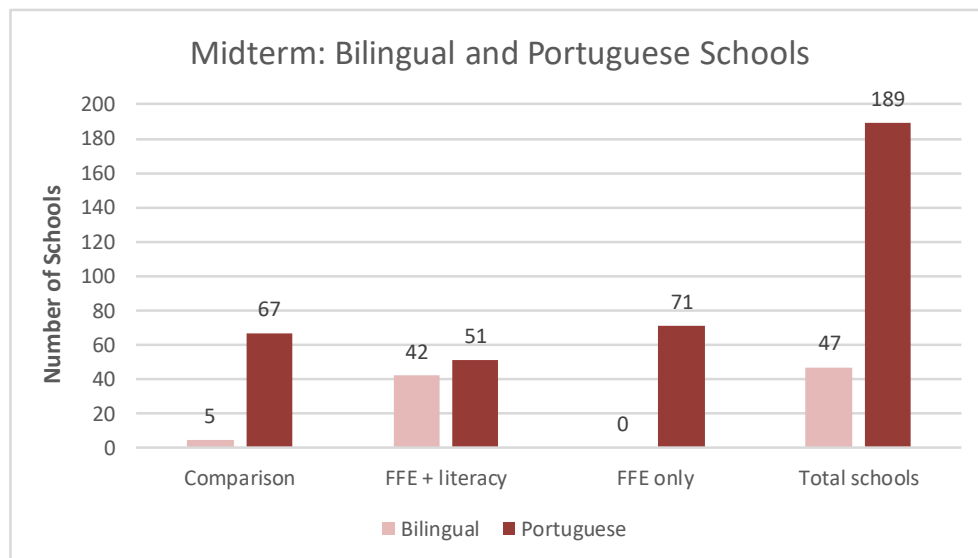


Figure 1: distribution of bilingual and Portuguese schools within the sample

Table 3 (below) show the distribution of students who participated in the midterm evaluation by language of instruction. Students were tested in their language of instruction. However, where students in the Portuguese group did not speak Portuguese sufficiently well to understand the instructions of the EGRA subtasks, for example, these were given in the local language, with the items themselves (names of object, words, letter sounds, etc.) presented in Portuguese.

Midterm: breakdown of students by language of instruction			
Row Labels	Bilingual	Portuguese	Grand Total
Comparison (Bilingual)	5		5
Comparison (Portuguese)		67	67
FFE + lit (Bilingual)	42		42
FFE + lit (Portuguese)		51	51
FFE only		71	71
Grand Total	48	189	236

Table 3: numbers of students who participated in the midterm evaluation by language of instruction

This composition of the three groups is balanced in terms of gender. 50% of the bilingual students are female and 50% male. 49% of the Portuguese students are female and 51% male. Because students were selected at random from within the classes identified by the sampling plan, the gender distribution within the sample should reflect that of the school population. For details of the breakdown of each group by gender, see table 4 below.

Midterm: breakdown of students by gender						
Intervention groups	boy	%	girl	%	Grand Total	%
Comparison (Bilingual)	57	3%	52	3%	109	3%
Comparison (Portuguese)	469	24%	503	27%	972	25%
FFE + lit (Bilingual)	379	19%	382	20%	761	20%
FFE + lit (Portuguese)	531	27%	516	28%	1047	27%
FFE only	531	27%	421	22%	952	25%
Grand Total	1967		1874		3841	

Table 4: breakdown of student who participated in the midterm evaluation by gender

Management and quality assurance of the Midterm Evaluation

The midterm evaluation of FFE2 was conducted by a consortium of three members who responded to Planet Aid's Request for Proposal of 27 November 2018 in December and were awarded the contract. The consortium is composed of: a) Dr Simone Doctors (evaluation lead, qualitative data collection and analysis, main report author), b) Forcier Consulting (quantitative data collection) and c) the National Foundation for Educational Research (quantitative data analysis). The three members worked collaboratively under the coordination of Simone Doctors. The midterm evaluation was designed and planned by Simone Doctors with extensive input from and collaboration with the other members. Throughout the planning and data collection and analysis, regular three-way Skype meetings were held to ensure these processes ran as smoothly as possible and to optimize communication between the members.

The evaluation instruments were developed by Simone Doctors, based on those used at baseline, with input from members of the FFE team where necessary, so that the information collected would meet the needs and vision of the project team¹². These were scripted and coded in the Open Data Kit (ODK) application by the Forcier team, so data collection could take place using smartphones. Simone Doctors worked closely with the Forcier team during this process, testing the tools on the phones and requesting amendments to the script so that they worked as intended.

12 supervisors and 48 enumerators were recruited by Forcier and trained in Maputo by Simone Doctors and the Forcier team. A team of 20 potential supervisors received 5 days intensive training, following which 12 were

¹² All evaluation instruments used are available on request.

selected; the following week, these 12 supervisors participated in and contributed to the enumerator training, which also lasted 5 days, including one day piloting of the instruments. Again, the 48 enumerators were selected from a group of 60 based on their performance during training and piloting. Groups consisting of 4 enumerators and one supervisor were formed; care was taken to ensure groups were balanced in terms of gender, experience and languages competence (Portuguese, Xichangana and Xirhonga).

The evaluation instruments as coded in ODK were piloted in two schools using ODK-encoded smart phones¹³. Simone Doctors and three members of the Forcier team observed the piloting and made recommendations; recommendations from the enumerators and supervisors themselves were also considered. Forcier uploaded and collated the results from the smartphones and Simone Doctors and the Forcier team analyzed these. Following the pilot, some improvements were made to the instruments and an additional day's enumerator training was held to address issues identified by the enumerators themselves and their supervisors.

NFER were responsible for the sampling framework (see Sampling for Pupil-level data collection above). Forcier planned fieldwork and logistics for the quantitative data collection using this, with considerable assistance from the FFE project's M&E assistants. Forcier collected the quantitative data between 11 March and 16 April 2019, beginning with the comparison group schools, where there were fewer observations to conduct. Data collected using the ODK interface was uploaded to Forcier's secure server by supervisors at the end of each day's fieldwork. After cleaning, it was uploaded to NFER's secure server. Forcier and NFER liaised extensively with the objective of ensuring the data was presented in suitable form for NFER to work with.

Qualitative data were collected by Simone Doctors during two field visits to Mozambique (10 February – 3 March and 22 April – 11 May 2019). She was assisted in making logistical arrangements by the FFE2 management and M&E teams. She was accompanied in visits to the project districts to visit schools, HGSFGs and government officials by a member of the M&E team but all interviews were conducted without the presence of project representatives. When visiting EPFs, she travelled alone, for logistical reasons; however, she was accompanied on school visits to observe teaching practice by members of the EPF management and staff. Again, all interviews were conducted without EPF/ADPP staff being present. Semi-structured interviews and focus groups were recorded using a digital voice recorder and later transcribed by Forcier. These were then analyzed by Simone Doctors, using a version of framework analysis.

Quality assurance protocols were in place to ensure the consistency of the data collection process and validity of the data. Random spot checks were conducted by Forcier to ensure consistency between enumerators and the validity of the data gathered. Forcier will provide more detailed information about this process on request.

¹³ Two schools, one peri-urban, one rural, were selected. Both schools are receiving the FFE literacy program in local languages but do not form part of the evaluation sample.

Spot checks of approximately 10% of the transcriptions of qualitative data were performed, comparing this against the original recordings: these revealed the transcription quality to be acceptable.

Rigorous data security protocols were in place. These provided for all data to be uploaded to NFER's secure server, so no data containing personal information was transmitted using other means. In particular, email was not used to transmit data containing personal information. In the field, all information was stored on encrypted computers, then uploaded to NFER's secure server. Enumerators were not permitted to use the smart phones used for quantitative data collection for any other purpose (all enumerators had additional phones and were provided with airtime to allow them to make logistical arrangements and to ensure their own security).

The internal self-review by project management and staff and the stakeholder review workshop were facilitated by Simone Doctors with assistance from a local researcher from Forcier.

The draft report was prepared by Simone Doctors then sent to Planet Aid, ADPP and the different implementing partners for factual verification. The evaluators wished to ensure that the report fairly and accurately reflected the lived experience of the implementing teams and incorporated any information which provided context for the report's findings. Simone Doctors and NFER responded to all comments and requests for clarification received as a result of this process and received a small number of follow-up comments. During this dialogue, the evaluators retained a strictly independent stance, dealing with each comment on a case-by-case basis and deciding not to incorporate comments judged not to be relevant or supported by other sources of evidence. Following this dialogue, the final draft was prepared by Simone Doctors.

Challenges, delays and lessons learned

A number of challenges during field work and data analysis led to delays and lessons learned for the future:

- The team of enumerators and supervisors found the logistics of accessing some of the more remote schools quite challenging, as suitable vehicles were not always available. Extreme weather impacted the fieldwork, with rain leaving some roads impassable: one school could not be visited, despite the enumerator team making three attempts. This was exacerbated by the fact that the enumerators and drivers were not familiar with the location of some of the schools, particularly within the comparison schools not under FFE2, although the FFE2 team had provided guidance and contact information for the school directors. The importance of having reliable GPS coordinates for each school and more information about the conditions of the roads, particularly leading to the comparison group schools, is a lesson learned when planning for the final evaluation.
- Last minute edits to the ODK scripts to deal with problems identified during the piloting and initial field work were not fully tested due to pressure of time, leading to a small number of questions in the pupil

questionnaire being skipped and some data relating to school clubs not being collected.¹⁴ Although regrettable, this loss of data is not actually critical, since information regarding school clubs was collected both from teachers and in the school survey. The evaluation is therefore able to report on the question of school clubs from two perspectives rather than three as intended.

- The data collected by Forcier and uploaded to NFER initially included a number of duplicate entries, likely caused by confusion amongst enumerators about the substitution process. In order to resolve this issue, the Forcier team manually verified the physical data tracking sheets to identify the observations, in distance liaison with Simone Doctors. In a small number of cases, where it was not possible to ascertain with certainty which entry was correct, both were deleted from the data set. This time-consuming process led to delays in the data analysis workplan.
- Previous evaluations have documented the challenges caused by turnover and capacity challenges within the project M&E team. At the beginning of the evaluation process, there was only a skeleton M&E staff in post, and the external evaluators were supported with basic arrangements by the project leadership team and two M&E assistants. During the evaluation fieldwork, a new M&E manager and a new M&E specialist took up their functions. Both appear to be capable and competent; with support from the leadership team, they made considerable efforts to assist the external evaluators, accompany field visits and obtain information on request. Although M&E of the project has continued throughout the changes and gaps in personnel, these have inevitably led to some challenges in data collation and monitoring. It is promising that a fully-staffed M&E team is now in place.
- Delays in NFER receiving useable data to analyze led to delays in the workplan overall and required previously unscheduled time to correct.
- Some adjustments to the evaluation plan were necessary due to the two cyclones which devastated two regions of Mozambique. Two ADPP-run EPFs were affected, leading to one being evacuated and both being preoccupied with other vital matters during the period of evaluation. Some EPF information was therefore incomplete. Despite these events affecting many ADPP and FFE staff and their families, through involvement in the emergency response or because family members live in the affected areas, the ADPP leadership insisted that the evaluation should continue as planned and continued to provide all support necessary.
- The discovery by the project team that the lists of schools supplied to the evaluators included some inaccuracies meant that, late in the report preparation process, it was necessary to verify the status of all schools in the sample, reassign certain schools to different intervention groups and rerun the statistical analysis of the pupil-level information. The other data sets were not affected. In the future, it should be noted that, at least for the purposes of the evaluation plan, it is preferable to avoid changes to the status of

¹⁴Since these questions were only asked at the intervention schools and fieldwork began in the comparison schools, this problem was not picked up by Forcier's usual QA processes.

schools, such as extending the literacy program to new schools. Where judged to be of such value to the schools as to outweigh other considerations, such changes should be rigorously noted for evaluation purposes, so the evaluation sample can be adjusted accordingly.

Midterm findings

Previous evaluations have been structured around the three evaluation questions:

- Internal validity (*is the project doing /did the project do what it said it would do?*)
- Initial impact (*is it making /has it made a difference?*)
- Strategic relevance regarding effectiveness, efficiency, impact and sustainability (*are/ were these the right things to do? what can be learned?*)

The current report does not follow this approach but is structured according to the logic of the project Theory of Change (see Annex 1), with findings presented in function of the project indicators, followed by a number of other transversal themes. Each section addresses the three evaluation questions to some extent. Given the intrinsic links between the areas covered by the indicators, the second and third evaluation questions regarding initial impact and strategic relevance are covered in more or less detail as appropriate, avoiding unnecessary repetition. Given that the project has already been evaluated several times, although the current report covers all project activities, it gives particular attention to the two new components which began with the second phase: i) the program to improve literacy instruction in primary schools, particularly literacy in local languages, and ii) the Home Grown School Feeding Gardens, which aim to make school feeding more sustainable, by promoting local production and diversification of food sources. Other project activities are often considered in less detail, unless there have been particular changes since baseline or the midterm data reveals something particular of note. The conclusion returns to the three evaluation questions listed above and addresses each of them in turn in the light of the midterm findings.

Where data collected at midterm from students, teachers or schools are reported, it should be noted that these refer to a sample of those populations. Where appropriate, results are extrapolated to the whole population, as is the case where targets are based on indicators formulated in terms “percentage of students who...” or “number of schools which...”. In these cases, informed estimations have been made, based on the behavior within a representative sample of the population. This should be borne in mind when, for example, results of the schools survey (conducted in 170 schools at midterm) expressed as percentages of those schools are compared with those of the baseline school survey (conducted in all 271 project schools at baseline) and expressed as percentages of all project schools.

Improved literacy of school-age children (SO1)

The first of the two large Strategic Objectives of the project within the project ToC, (SO1) concerns Improved literacy of school-age children. Each project activity is presented following the structure and logic of the ToC: each section begins with boxes summarizing the relevant project results in relation to the specific indicators and

targets associated with each activity. Midterm targets have been calculated as of March or April 2019 on the basis of targets for 2019¹⁵.

Indicator 26 (Outcome): Percent of students who, by the end of two grades of primary schooling, demonstrate that they can read and understand the meaning of grade level text. Final target: 45% Midterm target: 12%

Midterm results (April 2019): 6% (Female: 7%; Male 5%)

Midterm target 50% achieved (Female: 67%; Male 42% achieved)

Indicator 27 (Output): Number of individuals benefiting directly from USDA-funded interventions. Final target: 85,560 (Female: 45,450; Male: 40,110; Continuing: 78,748; New: 6,812)

Midterm target: 77,088 (Female: 38,701; Male: 38,387; Continuing: 76,307; New: 781)¹⁶

Midterm results (March 2019): 92,237; Female: 48,119; Male: 44,118; Continuing: 92,081; New: 156

Midterm target 108% achieved (Female: 106%; Male: 110% achieved; Continuing: 121% achieved; New: 20% achieved)

Indicator 28 (Output): Number of individuals benefiting indirectly from USDA-funded interventions. Final target: 336,000

Midterm target: 336,000

Midterm results (March 2019): 336,104

Final target 100% achieved

Assessing the literacy levels of students who have benefited from the literacy program delivered by ADPP with technical assistance from CE is key to evaluating the success of this intervention and of the wider project. The report assesses the literacy program itself within the section entitled Improved Quality of Literacy Instruction (MGD 1.1) below. The cohort of pupils tested at baseline, when they were in grade 1, performed an Early Grade Reading Assessment (EGRA) comparable to that used at baseline. The EGRA used at baseline had been developed by a team of specialists from CE, language specialists from INDE, the FFE literacy team and the FFE M&E team, working closely with an international consultant. Three equivalent versions were developed in Portuguese, Xichangana and Xirhonga. The version used at midterm was improved by the external evaluator working closely with the FFE literacy team. The task was then coded using ODK and administered by enumerators using smartphones.

The EGRA consists of 11 progressively ordered subtasks to assess pre-reading then reading skills, designed to reflect the progressive, cumulative stages of literacy acquisition. Pupils continue until they have reached the

¹⁵ Midterm targets have been calculated on a pro rata basis, or as appropriate for the nature of the activity, taking the targets for 2019 as set out in the document entitled Attachment E in Amendment # 1: Performance Indicators Modification 1.

¹⁶New midterm target = 25% of 2019 new target. Continuing midterm target = all the rest in midterm target

threshold at which they can no longer perform the tasks; at this point, discontinuation filters stop them attempting tasks beyond their level of competence, to minimize student distress. However, all pupils attempt the writing task.

Each of the three language versions of EGRA consisted of the following subtasks (all of which are language specific, reflecting the phonological and orthographic rules of the target language):

- ST1. Oral vocabulary (name the object depicted in stimuli pictures – 10 items)
- ST2. Comprehension of an oral text (respond orally to comprehension questions after hearing a short text – 4 items)
- ST3. Phonological awareness (indicate the picture of an object whose name begins with the same initial sound as that of a stimulus object – 10 items)
- ST4. Concepts of print (perform tasks to demonstrate familiarity with how printed language functions – 10 items)
- ST5. Letter sounds (produce the letter sounds when shown printed lower and upper case letters – 100 items)
- ST6. Syllable recognition (read aloud syllables consisting of consonant and vowel combinations permitted in the target language – 50 items)
- ST7. Reading simple words (read aloud a list of words progressively longer and less common - 30 items)
- ST8. Reading fluency (read a short passage – number of words read correctly within 60 seconds – 70 items; and Reading comprehension (respond to questions based on the passage just read – 4 items)
- ST9. Writing (correctly write first name and family name – 2 items); and Dictation - words (correctly write 5 words dictated – 5 items); Dictation – phrase (correctly write 5-word phrase - 5 items)

The 11 subtasks were analyzed as 9 subscales (reading fluency and reading comprehension are considered as one subscale; writing tasks were considered as one subscale). The results of each subtask were analyzed across the following comparisons: a) FFE + literacy (all) versus comparison (all); b) FFE + literacy (Portuguese) versus comparison (Portuguese); c) FFE + literacy (local languages) versus comparison (local languages); d) FFE + literacy (Portuguese) versus FFE only (Portuguese); e) FFE + literacy (Portuguese) versus FFE + literacy (local languages)¹⁷.

The results for each subtask are presented below (for more details, please see Technical Appendix 1, prepared by the statistical team from NFER).

¹⁷ Following changes to the project since baseline, there is no longer a group of FFE only schools in project districts learning in local languages; all such schools are now benefiting from the literacy intervention. The sample and the evaluation design have been changed to reflect the fact that there is no longer an “FFE only (local languages)” intervention group.

Oral vocabulary

The oral vocabulary subtask consists of ten rows of three pictures of everyday objects; pupils are asked to indicate the object named by the enumerator in their language of instruction (Portuguese, Xichangana or Xirhonga, as appropriate).

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** overall, FFE + literacy students achieved significantly greater increases between baseline and midterm than did comparison students.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** overall, FFE + literacy students studying and tested in Portuguese achieved significantly greater increases between baseline and midterm than did comparison students studying and tested in Portuguese.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** overall, FFE + literacy students studying and tested in local languages (Xichangana or Xirhonga) achieved slightly smaller mean increases between baseline and midterm than students in the comparison group studying and tested in the same languages, although this difference was not significant.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** overall, FFE + literacy students studying and tested in Portuguese achieved significantly lesser increases between baseline and midterm than FFE only students studying and tested in Portuguese.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in Portuguese achieved significantly greater increases between baseline and midterm than did FFE + literacy students studying and tested in local languages.

This is a straightforward task which would not be expected to cause difficulty for the majority of grade 3 pupils. In effect, project students performed well overall on this pre-reading task, with all groups achieving mean scores of over 9.5/10. Since the analysis looked at the difference between baseline and midterm, rather than the midterm results, it is the comparative performance since baseline which was measured, rather than the midterm result in isolation. For example, with regard to d) contrast 4 (FFE + literacy versus FFE only tested in Portuguese), although the FFE + literacy group outperformed the FFE only group at midterm, because they had outperformed them still further at baseline the FFE only group achieved significantly better results.

The complete results are presented in Technical Appendix 1, pp 5 - 15.

Oral comprehension

Pupils listened to a short story (48 words in length) read by the enumerator, then answered four simple questions relating to what they had heard.

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** overall, comparison students achieved significantly greater increases between baseline and midterm than did FFE + literacy students.

- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** overall, comparison students studying and tested in Portuguese achieved significantly greater increases between baseline and midterm than did FFE + literacy students studying and tested in Portuguese.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** overall, FFE + literacy students studying and tested in local languages (Xichangana or Xirhonga) achieved smaller mean increases between baseline and midterm than students in the comparison group studying and tested in the same languages, although this difference was not significant.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** overall, FFE + literacy students studying and tested in Portuguese achieved significantly greater increases between baseline and midterm than FFE only students studying and tested in Portuguese.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in local languages achieved smaller increases between baseline and midterm than did FFE + literacy students studying and tested in Portuguese, although this difference was not significant.

Although results overall have increased since baseline, this remains a task which causes some difficulties to many pupils, both those tested in Portuguese and in local languages, with mean scores for all groups of below 3 out of 4, and mean scores for both the FFE only group and the comparison (local languages) of less than 2 out of 4, suggesting that some grade 3 pupils are struggling to process oral language effectively (in all groups, the minimum score was 0 and the maximum 4).

The complete results are presented in Technical Appendix 1, pp 16 - 25.

Phonological awareness

Pupils were presented with a series of ten rows of three photos of everyday objects. For each row, they responded to a stimulus word pronounced by the enumerator by indicating the object whose name begins with the same initial consonant sound.

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** overall, comparison students progressed more between baseline and midterm than FFE + literacy students, although the difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** overall, FFE + literacy students outperformed comparison students, although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** overall, FFE + literacy students studying and tested in local languages (Xichangana or Xirhonga) achieved higher scores at midterm than students in the comparison group studying and tested in the same languages. Nevertheless, the comparison group made relatively more progress between baseline and midterm, although the difference was not statistically significant.

- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** overall, FFE + literacy students studying and tested in Portuguese achieved significantly greater increases between baseline and midterm than FFE only students studying and tested in Portuguese.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in Portuguese achieved slightly greater increases between baseline and midterm than did FFE + literacy students studying and tested in local languages, although this difference was not significant.

All groups progressed in this activity between baseline and midterm, with most groups achieving mean midterm scores of between 6.3 and 7.8 out of 10 (in all groups, the minimum score was 0 and the maximum 10).

The complete results are presented in Technical Appendix 1, pp 26 - 35.

Concepts of print

Within this sub task, pupils are asked to perform ten activities as evidence of their familiarity with printed materials.

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students achieved greater increases between baseline and midterm than comparison students overall, although the difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, comparison students achieved greater increases between baseline and midterm than FFE + literacy students, although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students achieved greater increases between baseline and midterm than comparison students, although the difference was not significant.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students achieved greater increases between baseline and midterm than FFE only students, although the difference was not significant.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in local languages achieved greater increases between baseline and midterm than did FFE + literacy students studying and tested in Portuguese, although this difference was not significant.

All groups progressed in this activity between baseline and midterm, with all groups achieving mean midterm scores of between 6.2 and 7.6 out of 10 (in all groups, the minimum score was 0 and the maximum 10).

The complete results are presented in Technical Appendix 1, pp 6 - 45.

Letter sound recognition

Pupils were asked to produce sounds of 100 upper- or lower-case letters, presented in ten rows of ten, producing as many as possible within 60 seconds.

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students progressed more than comparison students overall although the difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, FFE + literacy students progressed more than comparison students overall although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students progressed significantly more than comparison students overall.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students progressed significantly more than FFE only students overall.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in local languages achieved smaller increases between baseline and midterm than did FFE + literacy students studying and tested in Portuguese, although this difference was not significant.

Scores for this subtask were low overall (means of between 8.6 and 8.8 out of 100 in the project groups, with means of as low as 3.0 out of 100 in the comparison groups; the overall range of scores was between 0 and 100). The low scores overall for this subtask reflect the fact that a large number of pupils failed to produce a single one of the first ten letter sounds correctly and therefore discontinued this activity (only 41% of the students tested at midterm completed this activity). As noted at baseline, in Mozambique, due to the ways reading is traditionally taught, pupils shown a written letter are not generally able to produce the corresponding letter sound but tend to respond by naming that letter. Since teaching reading does not traditionally involve phonics or letter sound recognition, experience shows that teachers, trainee teachers and teacher trainers may also have difficulty producing letter sounds. Given this knowledge, considerable focus was given to preparing enumerators to administer this sub-task during enumerator training; it is nevertheless possible that either pupils' performance of the task or enumerator grading of this may have been impacted by the enumerators' relative lack of familiarity with the task.

The complete results are presented in Technical Appendix 1, pp 46 - 55.

Reading syllables

Pupils were asked to read 50 syllables constituting plausible vowel-consonant combinations of the languages of testing, reading as many as possible within 60 seconds.

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students progressed less than comparison students overall although the difference was not significant.

- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, FFE + literacy students progressed more than comparison students overall although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students progressed more than comparison students overall although the difference was not significant.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students progressed significantly more than FFE only students overall.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in Portuguese progressed significantly more than did FFE + literacy students studying and tested in local languages.

All groups progressed overall in this activity between baseline and midterm and performed better overall than on the previous activity (reading letter sounds). However, low scores overall for this subtask (means of between 4.0 and 9.7 out of 50 in the project groups, with means of as low as 2.5 out of 50 in the comparison groups; the overall range of scores was between 0 and 50) reflect the fact that only approximately 40% of pupils completed this activity, the others having discontinued after 60 seconds or when they failed to produce any one of the first ten syllables correctly. Of the 41% of all students who attempted this activity across all groups, 45% scored between 0 and 10 out of 50; 19% between 11 and 20; 13% between 21 and 30; 12% between 31 and 40 and 11% between 41 and 50 out of 50.

The complete results are presented in Technical Appendix 1, pp 56 - 65.

Reading words

Pupils were asked to read a series of 30 progressively more difficult words aloud. This task was not timed.

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students progressed more than comparison students overall although the difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, FFE + literacy students progressed more than comparison students overall although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students progressed significantly more than comparison students overall.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students progressed significantly more than FFE only students overall.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in local languages progressed more than did FFE + literacy students studying and tested in Portuguese, although the difference was not significant.

Students who attempted this task performed relatively better than on the previous one (with means of between 7.1 and 8.2 out of 30 in the project groups, and means of between 2.4 and 6.4 out of 30 in the comparison groups; the overall range of scores was between 0 and 30). The results are extremely polarized: of the 41% of all students who attempted this activity across all groups, 43% scored between 0 and 10; 17% between 11 and 20; and 40% between 21 and 30. At either extreme, 16% scored 0, whereas 13% scored 30 out of 30. The complete results are presented in Technical Appendix 1, pp 66 – 75.

Reading fluency (words per minute)

Pupils were asked to read a short text (70 words in Portuguese or 53 words in Xichangana or Xirhonga: the length of the local language texts is equivalent to the Portuguese text but has fewer words due to the morphological structure of those languages). The task was timed and the number of words read correctly within one minute recorded to give a fluency rating; pupils then had a further two minutes to finish reading the text if needed. The results below relate to reading fluency (words per minute).

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students progressed less than comparison students overall, although the difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, FFE + literacy students progressed less than comparison students overall although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students progressed more than comparison students overall although the difference was not significant.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students progressed significantly more than FFE only students overall.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in local languages progressed less than did FFE + literacy students studying and tested in Portuguese, although the difference was not significant.

Of the 41% of students who attempted this activity, mean scores for the number of words read correctly in 60 seconds were between 6.7 and 8.3 in the project groups, and between 2.4 and 8.7 in the comparison groups; the overall range of scores was between 0 and 70 for students tested in Portuguese and between 0 and 53 for students tested in local languages. Of the students who attempted this activity, 66% read between 0 and 20 words in 60 seconds; 22% read between 21 and 40 words; 13% read between 41 and 70 words. The relationship between fluency and accuracy is discussed on page 39 below.

The complete results are presented in Technical Appendix 1, pp 76 - 85.

Comparative reading fluency

In order to compare the performance of students tested in Xichangana and Xirhonga with those tested in Portuguese, overall fluency scores were divided by 70 for students tested in Portuguese and 53 for those tested in Xichangana and Xirhonga.

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students progressed more than comparison students overall although the difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, FFE + literacy students progressed less than comparison students overall although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students progressed more than comparison students overall although the difference was not significant.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students progressed significantly more than FFE only students overall.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in local languages progressed less than did FFE + literacy students studying and tested in Portuguese, although the difference was not significant.

The complete results are presented in Technical Appendix 1, pp 86 - 95.

Overall reading ability

In order to assess students' overall ability to read the text, regardless of timing, as opposed to fluency (words per minute), students who had not finished reading the text within 60 seconds continued reading for up to a further 2 minutes or until they completed the text or discontinued the activity. This condition did not apply at baseline, when all students discontinued reading the text after 60 seconds. It was included at midterm at the request of the literacy team in order to assess pupils' reading of a text without pressure of time.

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students performed less well than comparison students overall although the difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, FFE + literacy students performed less well than comparison students overall although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students performed significantly better than comparison students overall.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students performed better than FFE only students overall although the difference was not significant.

- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in local languages performed less well than did FFE + literacy students studying and tested in Portuguese, although the difference was not significant.

Due to the lack of baseline values, these differences cannot conclusively be attributed to the intervention.

The complete results are presented in Technical Appendix 1, pp 96 - 105.

It is interesting to compare the number of words students are able to read correctly when they are allowed three minutes to read the test (overall reading ability or accuracy), with the number of words they are able to read correctly in a minute (fluency). Overall, there is an inverse relationship between fluency and accuracy: accuracy increased as fluency decreased (see figure 2 below). Of the 41% of students who attempted the task, only 5% were able to correctly read between 61 – 70 words in a minute, whereas 24% were able to correctly read between 61 to 70 words in three minutes.

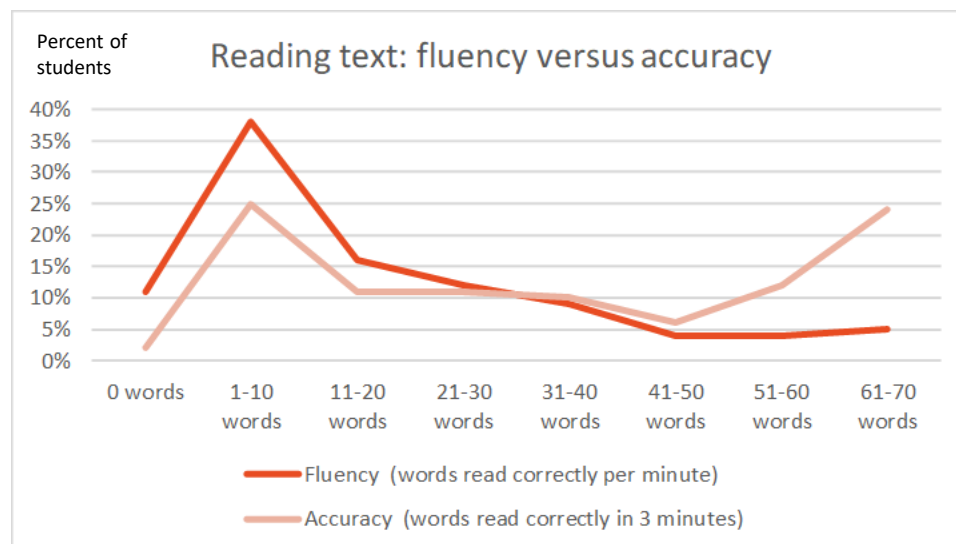


Figure 2: Comparative fluency and accuracy scores reading a text (all students who attempted task)

Comparative overall reading ability

In order to compare the overall reading ability (total number of words correctly read) of students tested in Xichangana and Xirhonga with those tested in Portuguese, overall reading ability scores were divided by 70 for students tested in Portuguese and 53 for those tested in Xichangana and Xirhonga.

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students performed slightly better than comparison students overall, although the difference was not significant.

- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** FFE + literacy students progressed slightly less well than comparison students between baseline and midterm, although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students performed significantly better than comparison students overall.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students performed better than FFE only students overall although the difference was not statistically significant.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in local languages performed better than did FFE + literacy students studying and tested in Portuguese, although the difference was not significant.

The complete results are presented in Technical Appendix 1, pp 105 - 115.

Comprehension of reading

To assess whether pupils understood what they had read, they were asked between one and four comprehension questions corresponding to the point in the text they had reached when reading.

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students made more progress between baseline and midterm than comparison students overall, although the difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, FFE + literacy students performed less well than comparison students overall although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students performed significantly better than comparison students overall.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students performed significantly better than FFE only students overall.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in local languages performed less well than did FFE + literacy students studying and tested in Portuguese, although the difference was not significant.

The comprehension results overall were low: overall the mean scores of the intervention groups were between 0.49 and 0.57 out of a possible total of 4; overall the mean scores of the comparison groups were between 0.11 and 0.60 out of a possible total of 4; scores ranged between 0 and 4 overall. Comprehension results were polarized: of the 41% of students who attempted this activity, 37% scored 0 out of 4, whereas 25% scored 4 out of 4. 16% students scored 1 out of 4; 13% scored 2 out of 4; 8% scored 3 out of 4. It should be remembered that students only attempted the comprehension questions for those portions of the text they had finished reading.

The complete results are presented in Technical Appendix 1, pp 115 - 125.

All pupils, regardless of whether they had completed the reading tasks or discontinued these, were asked to perform a series of short writing tasks: writing their first name, writing their family name, writing five discrete words as a dictation exercise and writing a short phrase of five words. Discontinuation protocols operated within the writing subtasks.

Writing first name correctly

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students made less progress between baseline and midterm than comparison students overall, although the difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, FFE + literacy students performed less well than comparison students overall although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students performed better than comparison students overall although the difference was not significant.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students performed significantly better than FFE only students overall.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in local languages made more progress between baseline and midterm than FFE + literacy students studying and tested in Portuguese, although the difference was not significant.

All participating students were asked to write their first and family names. Given that all students are in the third grade, it is striking that only 73% of the students from intervention groups and between 39% and 74% from comparison groups were able to perform this task correctly.

The complete results are presented in Technical Appendix 1, pp 155 - 161.

Writing family name correctly

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students performed less well than comparison students overall although the difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, FFE + literacy students performed less well than comparison students overall although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students performed better than comparison students overall although the difference was not significant.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students performed better than FFE only students overall although the difference was not significant.

- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in local languages performed less well than FFE + literacy students studying and tested in Portuguese, although the difference was not significant.

Overall, pupils were less likely to be able to write their family names correctly than their first names: of the intervention groups, between 42% and 45% of students were able to write their family name correctly; within the comparison groups this ranged from between 23% to 47%. Again, this is a striking finding given the age of the students.

The complete results are presented in Technical Appendix 1, pp 161 - 167.

Writing whole name correctly

A further analysis of the same items (first name and family name) sought to assess whether students were able to write their complete name correctly.

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students performed less well than comparison students overall although the difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, FFE + literacy students performed less well than comparison students overall although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students performed significantly better than comparison students overall.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students performed better than FFE only students overall although the difference was not significant.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in local languages performed less well than FFE + literacy students studying and tested in Portuguese, although the difference was not significant.

The complete results are presented in Technical Appendix 1, pp 167 - 173.

Writing dictation: individual words

At baseline, pupils had been asked to write ten progressively more difficult individual words as a dictation. At midterm, at the request of the literacy team, this task was divided into two discrete elements in order to investigate whether pupils would perform better if the words were presented in the context of a simple phrase: a dictation of five individual words followed by a dictation of a short phrase containing five words. Due to this change in the format of the task, it was not possible to compare the midterm results directly with those from baseline.

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students performed less well than comparison students overall although the difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, FFE + literacy students performed less well than comparison students overall although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students performed significantly better than comparison students overall.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students performed significantly better than FFE only students overall.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in local languages performed less well than FFE + literacy students studying and tested in Portuguese, although the difference was not significant.

Of the students who performed this task, the mean results ranged between 0.93 to 1.07 out of 5 for the intervention groups and between 0.21 and 1.12 out of 5 for the comparison groups; the range was from 0 to 5 out of 5. The complete results are presented in Technical Appendix 1, pp 136 - 145.

Writing dictation: phrase

In writing the short phrase consisting of five words:

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students performed better than comparison students overall although the difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, FFE + literacy students performed better than comparison students overall although the difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students performed significantly better than comparison students overall.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students performed significantly better than FFE only students overall.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in local languages performed less well than FFE + literacy students studying and tested in Portuguese, although the difference was not significant.

Of the students who performed this task, the mean results ranged between 1.06 to 1.09 out of 5 for the intervention groups and between 0.20 and 0.92 out of 5 for the comparison groups; the range was from 0 to 5 out of 5. The students receiving the literacy intervention performed the writing task better when the five words formed a coherent phrase than when they were five discrete words; this was not the case for the students not receiving the literacy intervention (FFE only and comparison students). The complete results are presented in Technical Appendix 1, pp 145 - 155.

Total writing

A total writing score was calculated based on the sum of scores for all the writing tasks (writing first name, writing family name, writing dictation words, writing dictation phrase).

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, FFE + literacy students performed less well than comparison students overall, although the difference was not statistically significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, FFE + literacy students performed significantly less well than comparison students overall.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, FFE + literacy students performed better than comparison students overall, although the difference was not significant.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, FFE + literacy students performed significantly better than FFE only students overall.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** overall, FFE + literacy students studying and tested in Portuguese performed better than FFE + literacy students studying and tested in local languages, although the difference was not significant.

The mean total writing results ranged between 2.10 to 2.22 out of 12 for the intervention groups and between 0.92 and 3.05 out of 12 for the comparison groups; the range was from 0 to 10 out of 12 for all groups except for the comparison (Portuguese) group, which had a range of 0 - 12. Students in the comparison (Portuguese) performed significantly better in writing overall than the other groups.

The complete results are presented in Technical Appendix 1, pp 126 - 135.

EGRA results: Discussion

At baseline, the EGRA results for the cohort tested, then in grade 1, had revealed very low literacy outcomes overall, with almost no grade 1 learners able to proceed beyond the first four sub tasks (this was not surprising, given that, at the time of testing, they had only been attending school for three to four months). At midterm, the situation is much more nuanced: although many learners discontinued the reading part of the EGRA after the first five tasks, since they were not able to read any syllables, 41% continued and completed the subtasks on reading syllables, reading words, reading a text and reading comprehension. The results of these subtasks should therefore be considered in this light; for example, although the overall mean score for reading words in the FFE + literacy group was 7.13/30, the overall mean score of those who actually attempted the subtask was closer to 20/30.

Overall, then, the pupils tested have progressed since baseline, although almost 60% are still not able to read letter sounds or syllables. Two questions must be answered:

1. How have the pupils tested progressed compared to expectations (national benchmarks)?
2. What has been the impact of the literacy intervention? Have the students in the FFE + literacy group progressed more than the other students tested?

Progression in relation to national benchmarks

The indicator “Percent of students who, by the end of two grades of primary schooling, demonstrate that they can read and understand the meaning of grade level text” assumes a nationally accepted benchmark, or general agreement as to what children of a particular grade should be able to do. In the absence of official national benchmarks, but in line with recommendations developed by the USAID *Vamos Ler* (Let’s Read) project¹⁸, the project literacy team has been working with the following definition: 30 words per minute (WPM) for Portuguese; 25 WPM for local languages (Xichangana/Xirhonga) and 3/4 (75%) comprehension questions correct in each case. This is the measure of “demonstrate that they can read and understand the meaning of grade level text” which is therefore adopted for the purposes of this evaluation.

At midterm, pupils within the FFE + literacy group were compared with those in the FFE only and comparison groups, using the same 5 comparisons as for the individual EGRA subtasks, in order to investigate the role of the literacy intervention on their ability to meet the benchmark.

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, 6.7 % of FFE + literacy students met the benchmark compared with 8.3% of comparison students; FFE + literacy students performed less well than comparison students overall, although this difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, 7.8 % of FFE + literacy students met the benchmark compared with 9.1% of comparison students; FFE + literacy students performed less well than comparison students overall, although this difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, 5.2% of FFE + literacy students met the benchmark, compared with 1.9% of comparison students; FFE + literacy students performed better than the comparison students, although the difference was not significant.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, 7.8% of FFE + literacy students met the benchmark, compared with 4.5% of FFE only students; FFE + literacy students performed better than FFE only students, although this difference was not statistically significant.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** at midterm, 7.8% of FFE + literacy students studying and tested in Portuguese met the benchmark, compared with 5.2% of FFE + literacy students studying and tested in local languages; FFE + literacy students studying and tested in

¹⁸ The five-year USAID-funded Mozambique *Vamos Ler!* Program, which began in 2016, develops bilingual education pedagogical tools and activities to improve national early grade literacy.

Portuguese, performed better than FFE + literacy students studying and tested in local languages, although the difference was not significant .

The complete results are presented in Technical Appendix 1, pp 173 - 178.

Progression in relation to national benchmarks: boys

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, 5.9 % of FFE + literacy students met the benchmark compared with 7.2% of comparison students; FFE + literacy students performed less well than comparison students overall, although this difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, 7.6 % of FFE + literacy students met the benchmark compared with 7.9% of comparison students; FFE + literacy students performed less well than comparison students overall, although this difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, 3.4% of FFE + literacy students met the benchmark, compared with 1.8% of comparison students; FFE + literacy students performed better than the comparison students, although the difference was not significant.
- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, 7.6% of FFE + literacy students met the benchmark, compared with 2.5% of FFE only students; FFE + literacy students performed significantly better than FFE only students.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** at midterm, 7.6% of FFE + literacy students studying and tested in Portuguese met the benchmark, compared with 3.4% of FFE + literacy students studying and tested in local languages; FFE + literacy students studying and tested in Portuguese, performed better than FFE + literacy students studying and tested in local languages, although the difference was not significant.

The complete results are presented in Technical Appendix 1, pp 178 - 183.

Progression in relation to national benchmarks: girls

- a) **Contrast 1: FFE + literacy (all) vs Comparison (all):** at midterm, 7.6 % of FFE + literacy students met the benchmark compared with 9.4% of comparison students; FFE + literacy students performed less well than comparison students overall, although this difference was not significant.
- b) **Contrast 2: FFE + literacy (Portuguese) vs comparison (Portuguese):** at midterm, 8.1 % of FFE + literacy students met the benchmark compared with 10.1% of comparison students; FFE + literacy students performed less well than comparison students overall, although this difference was not significant.
- c) **Contrast 3: FFE + literacy (local languages) vs comparison (local languages):** at midterm, 6.9% of FFE + literacy students met the benchmark, compared with 1.9% of comparison students; FFE + literacy students performed better than the comparison students, although the difference was not significant.

- d) **Contrast 4: FFE + literacy (Portuguese) vs FFE only (Portuguese):** at midterm, 8.1% of FFE + literacy students met the benchmark, compared with 7.1% of FFE only students; FFE + literacy students performed better than FFE only students, although this difference was not significant.
- e) **Contrast 5: FFE + literacy (local languages) vs FFE + literacy (Portuguese):** at midterm, 8.1% of FFE + literacy students studying and tested in Portuguese met the benchmark, compared with 6.9% of FFE + literacy students studying and tested in local languages; FFE + literacy students studying and tested in Portuguese, performed better than FFE + literacy students studying and tested in local languages, although the difference was not significant.

The complete results are presented in Technical Appendix 1, pp 183 - 188.

8% of students (8% of boys; 8% of girls) in the FFE + literacy (Portuguese) group met the benchmark, compared with 5% (2% of boys; 7% of girls) in the FFE only group and 5% (3% of boys; 7% of girls) in the FFE + literacy (bilingual) group (see table 5 below). Amongst project students, the percentage of those reaching the national benchmark is thus 6% (5% boys; 7% girls). The midterm target of 12% for this indicator has thus been 50% achieved (42% for boys; 58% for girls). This result should be considered both in the light of the overall results presented above, and of the differences discussed below.

Intervention groups	Did not achieve benchmark	Achieved benchmark	Percent achieved benchmark	Total
Comparison (Bilingual)	106	2	2%	108
Comparison (Portuguese)	884	88	9%	972
FFE + lit (Bilingual)	716	39	5%	755
FFE + lit (Portuguese)	956	81	8%	1037
FFE only	908	43	5%	951
Total project students	2,580	163	6%	2,743
Grand Total	3,570	253	7%	3,823

Table 5: Students achieving national benchmarks

What has been the impact of the literacy intervention? Have the students in the FFE + literacy group progressed more than the other students tested?

The analysis of the EGRA results shows a nuanced picture. On the one hand, the overall results demonstrate a very low level of reading overall, with only 6% of project students (7% of those benefiting from the literacy intervention) meeting the national benchmark, compared with 8% of comparison students overall. The results do not demonstrate a resounding impact of the literacy intervention to date, in which FFE + literacy children consistently outperform the comparison group children. However, taken overall, the results suggest that the literacy intervention is beginning to have an impact: Table 6 below summarizes the significant differences in results of EGRA subtasks between the intervention groups (all the results reported are included, meaning that

Subtask	Significant difference observed?					
	FFE+literacy (all) vs comparison (all)	FFE+literacy (Portuguese) vs comparison (Portuguese)	FFE+literacy (local langs) vs comparison (local langs)	FFE+literacy (Portuguese) vs FFE only (Portuguese)	FFE+literacy (local langs) vs FFE+literacy (Portuguese)	
Oral vocabulary	+ significant	+ significant	--ns	-- significant	-- significant	
Oral comprehension	-- significant	-- significant	-- ns	+ significant	-- ns	
Phonological awareness	-- ns	+ ns	--ns	+ significant	-- ns	
Concepts of print	+ns	--ns	+ ns	+ ns	+ns	
Letter sound recognition	+ ns	+ ns	+ significant	+ significant	-- ns	
Reading syllables	-- ns	+ ns	+ ns	+ significant	-- significant	
Reading words	+ ns	+ ns	+ significant	+ significant	+ns	KEY + significant treatment group performed significantly better -- significant treatment group performed significantly worse + ns treatment group performed better but difference not statistically significant --ns treatment group performed worse but difference not statistically significant treatment groups are in red in column heads
Reading fluency (words per minute)	--ns	--ns	+ ns	+ significant	-- ns	
Comparative reading fluency	+ ns	-- ns	+ ns	+ significant	-- ns	
Overall reading ability	--ns	--ns	+ significant	+ ns	-- ns	
Comparative overall reading	+ ns	= ns	+ significant	+ ns	+ns	
Comprehension of reading	+ns	--ns	+ significant	+ significant	-- ns	
Writing first name correctly	--ns	--ns	+ ns	+ significant	+ns	
Writing family name correctly	--ns	--ns	+ ns	+ ns	-- ns	
Writing whole name correctly	--ns	--ns	+ significant	+ ns	-- ns	
Writing dictation: individual words	--ns	--ns	+ significant	+ significant	-- ns	
Writing dictation: phrase	+ ns	+ ns	+ significant	+ significant	-- ns	
Total writing	--ns	-- significant	+ ns	+ significant	-- ns	

Table 6: Statistically significant differences in results of EGRA subtasks between intervention groups

some are reported twice. For example, “writing first name correctly”, “writing family name correctly” and “writing whole name correctly” are presented on three separate lines, so some of the information presented

overlaps). Across all comparisons, in 16 of the 18 lines presented in Table 6, the FFE + literacy group performed significantly better than either the comparison group or the FFE only group, in one or more instances ¹⁹. The FFE + literacy group did not perform significantly better than the comparison group overall. However, in 8 of the 18 cases reported, of the students tested in local languages, the FFE + literacy group performed significantly better than the comparison group; there were no cases where the comparison (local languages) group performed significantly better than the FFE + literacy (local languages) group.

There was only one instance where the FFE + literacy group performed significantly better than the comparison group when students were tested in Portuguese. In two cases, the comparison group performed significantly better than the FFE + literacy group when students were tested in Portuguese. Taken together, these findings provide tentative evidence of the efficacy of teaching literacy in local languages. This suggests that, while more time may be needed for the effects of the intervention to translate into improved literacy on a wide scale, the first signs are promising.

In 12 of the 18 cases reported, the FFE + literacy group performed significantly better than the FFE only group overall, amongst students tested in Portuguese. This finding provides promising early evidence that the literacy intervention is having a positive impact on the emerging literacy skills of students within schools where the FFE project is operating. While more time may be needed for the effects of the intervention to translate into more consistent results, the first signs are encouraging.

When the results of FFE + literacy (Portuguese) are compared with those of FFE + literacy (local languages), only two lines show a significant difference: students tested in Portuguese performed significantly better than those tested in local languages in the oral vocabulary subtask and the reading syllables subtask.

Although the FFE + literacy students did not perform significantly better than the comparison students overall, two trends emerge clearly from this situation:

- i) Amongst students taught and tested in local languages, the FFE + literacy students performed significantly better than the comparison students in 8 of the 18 lines; the FFE + literacy students performed better than the comparison students in a further 7 of the 18 lines, although these differences were not statistically significant.
- ii) Amongst students taught and tested in Portuguese, the FFE + literacy students performed significantly better than the FFE only students in 12 of the 18 lines; the FFE + literacy students

¹⁹ This discussion considers only results where a difference observed is statistically significant. In reality, there were other results where a difference was observed which suggested the FFE + literacy group of students outperformed one of the other groups; in some cases, these results would have been statistically significant if a lower threshold had been applied.

performed better than the FFE only students in a further 5 of the 8 lines, although these differences were not significant.

One possible explanation for the fact that the FFE + literacy group did not perform significantly better than the comparison group overall might concern the characteristics of the comparison group. It had become clear at baseline that, despite the careful sampling plan drawn for this study, the comparison group consistently outperformed the two intervention groups in literacy. The baseline report speculated that the comparison group could not be considered truly comparable to the FFE + literacy and FFE only groups but differed in terms of other variables. Demographic and socio-economic factors were offered as possible explanations: two of the districts in the comparison group, Matola and Marracuene, are peri-urban, and have a higher socioeconomic profile than the four project districts. Indeed, the four districts of Maputo Province in which the project is implemented were chosen precisely because they have the lowest socioeconomic profiles and the highest rates of poverty, food insecurity and malnutrition of the districts in Maputo Province.

In order to probe further the possibility of differences in performance within the comparison group, the mean scores of the EGRA sub tasks for each district were combined in a single figure. Figure 3 below reveals that, in contrast to the assumption that students from Matola and Marracuene might be outperforming those from other districts, it is in fact students from Boane who out-perform those from other districts overall; students from Matutuine and Moamba also perform well overall on several subtasks. Magude has the lowest scores overall. Figure 4 below shows the same results, arranged to show overall EGRA scores by district (the subtask scores are still visible within the histogram columns). This allows the overall EGRA scores to be ranked by district: Boane performed best overall, followed by Moamba, Matutuine, Maracuenne, Matola, Manhiça, Namaacha, then Magude. This does not support the notion that the comparison group might be performing well overall because it includes the peri-urban districts of Matola and Marracuene, which come fourth and fifth in the overall ranking.

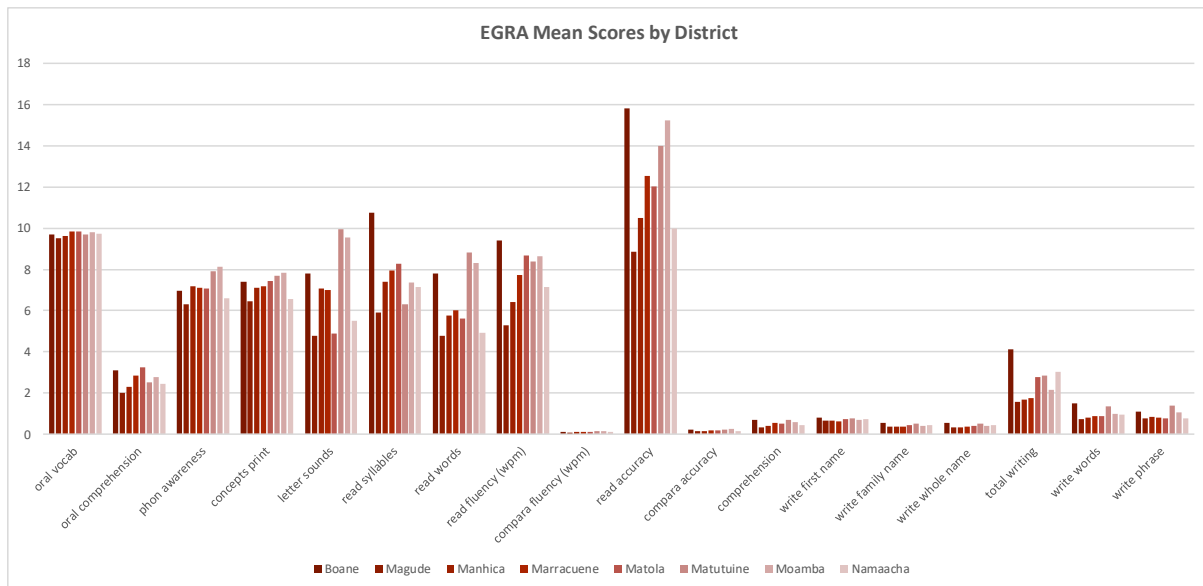


Figure 3: EGRA mean subtask scores by district

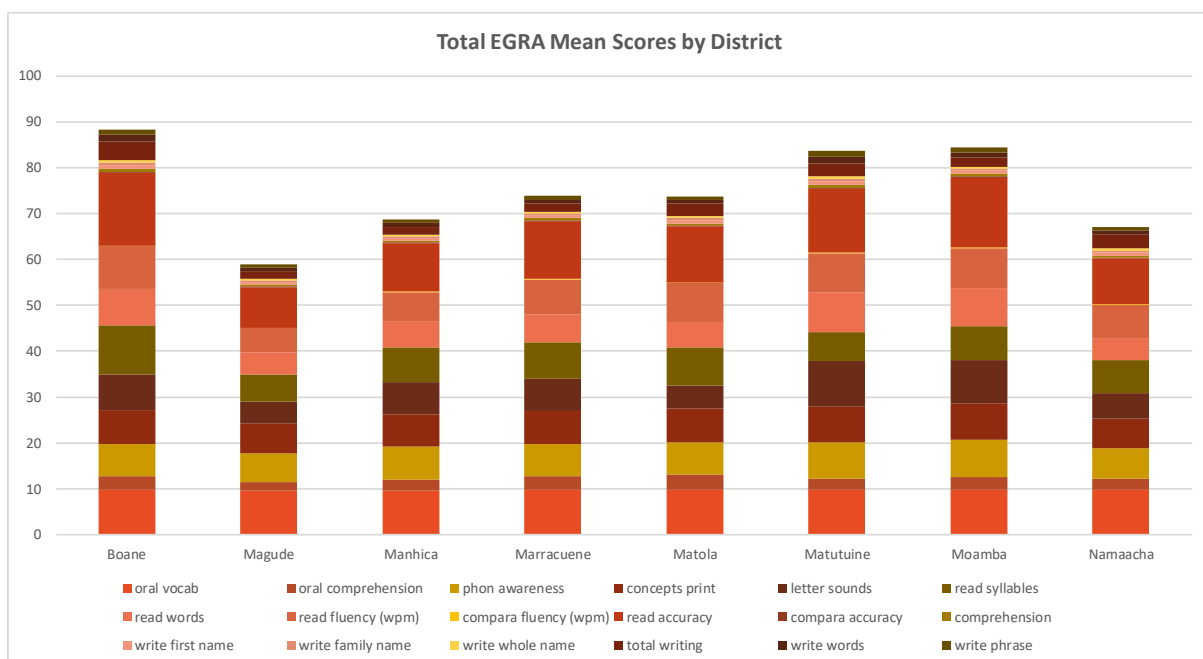


Figure 4: EGRA mean overall scores by district (subtask scores included)

The notion that the reason the FFE + literacy students did not out-perform the comparison students overall, because the comparison group performed better overall, due to extraneous variables, must then be rejected. How then to explain the fact that, despite having made progress through the literacy intervention, the FFE + intervention students are not overall performing significantly better than the comparison group? In the judgement of the external evaluator, this is likely to be due to two factors:

- The progress which the literacy intervention is beginning to show is still at an early stage and needs to be consolidated. This is not surprising, given that this is only the third year the literacy intervention has been operating in schools and such programs take time to embed;
- It is impossible to know what the EGRA outcomes of the students in the FFE + literacy schools would have been if they had not benefited from the literacy intervention. It is possible – as strongly suggested by one of the project leaders - that without the project interventions, this group of students would have educationally fallen further behind the students from the more socioeconomically advanced districts in the comparison group.

The final evaluation of FFE2 in 2020 should allow the initial achievements reported here to be tested after an additional year of the literacy intervention and cast more light on how the initial progress made translates into later literacy. In particular, for the students taught in local languages, the final evaluation will capture how literacy instruction begun in local languages then transitions into literacy in Portuguese, the national language of instruction.

Improved Quality of Literacy Instruction (MGD 1.1)

The project ToC encapsulates the assumption that the quality of literacy instruction will be improved by a) better access to school supplies and materials and b) increased skills and knowledge of teachers. The project interventions which aim to address both of these dimensions are presented below.

Access to School Supplies and Materials (MGD 1.1.2)

Indicator 2 (Output): Number of textbooks and other teaching and learning materials provided as a result of USDA assistance. Final target: 39,600
Midterm results (March 2019): 71,932
Final target 182% achieved

Number of schools receiving school supplies and materials as a result of USDA assistance. Final target: 264
Midterm target: 264
Midterm results (March 2019): 271
Final target 103% achieved

The project distributed textbooks and other instructional materials throughout phase one and continues to do so in phase two. These materials are presented as being intended to support extracurricular learning clubs (although in reality schools have discretion over how they are used). Until recently, the project clubs component organized annual training sessions in the use of these materials in the classroom for pedagogical directors and one additional teacher from each school. During these training sessions, participants were also consulted about

the type of materials they would like to receive. In recognition of the limitations of this form of training, and that those teachers who received it were not necessarily “cascading” it down to their colleagues as anticipated, in recent months the strategy has changed: training of teachers in the facilitation of learning clubs and the use of the materials supplied has been included in the in-service training for teachers offered by the project “professionals”.

During phase two, analysis of the project monitoring records shows that targets for the distribution have been surpassed. Just over 80% of teachers in project schools reported their school had received materials to support extracurricular learning clubs²⁰. Of these:

- just over 30% said they had been consulted over the choice of materials;
- 93% of teachers in FFE + literacy schools and 100% in FFE only schools said they found the materials useful
- when asked whether they had used any of the material “today or yesterday”, 32% of teachers in FFE + literacy schools and 33% in FFE only schools said that they had; teachers of bilingual classes claimed to have do so slightly more often (38%, compared with 29% in Portuguese only schools);
- almost all (100% of FFE + literacy and 94% of FFE only) described the materials as either “good” or “excellent”.

When asked whether they now had access to sufficient teaching and learning resources, 52% of FFE + literacy teachers and 44% of FFE only teachers replied that they did. Teachers in bilingual schools were more likely to say they had access to sufficient resources (54%, compared with 45% of teachers in Portuguese only schools). This difference is probably a consequence of the FFE + literacy schools, to which most of the bilingual schools belong, having benefited from extra resources through the literacy program. The project should now ensure that all teachers are aware of the existence of the materials provided and encouraged or capacitated to use these.

For more detail, see Technical Appendix 2, pp 28 – 40.

School directors expressed gratitude for the materials and claimed they made a significant difference to teachers’ ability to teach within their schools. One school director stated:

We received chalk, we received reams of paper, we received pens, pencils, colored pencils, which the children used in the clubs ... this was linked to changes in their writing... One child was full of admiration

²⁰ During analysis of the midterm data collected from teachers, it became clear that there is confusion amongst teachers between these materials provided by the project to all participating schools and the materials developed by the literacy team and provided to schools taking part in the literacy program. This is discussed further in the presentation of support to teachers below.

to see the difference between what he wrote the first day and what he wrote more or less at the end [of the year]. We gave them to the children to motivate them and to show their families at home too.

During school visits some of the materials, including syllable charts and other posters, were in evidence in some classrooms, suggesting they are indeed being used. In many cases, they can be seen in the directors' offices: this is often because this is the most secure and clean place to store objects of value (it is to be hoped that they are in fact used for teaching and learning, and not merely kept safe).

Lesson observations provided an opportunity to observe whether pupils had access to basic equipment such as textbooks, pens or pencils, exercise books and erasers. The results do not show a particular pattern of differences between FFE + literacy, FFE only and comparison group schools. During lesson observations, for the classes observed:

- 89% of FFE + literacy schools, 87% of FFE only schools and 81% comparison schools had a blackboard and chalk in acceptable condition;
- in 83% of FFE + literacy schools, 84% of in FFE only schools and 79% comparison schools there were textbooks for all pupils in acceptable condition, compared with;
- in 79% of FFE + literacy schools, 83% of FFE only schools and 81% of comparison schools, all pupils had either a pencil or a pen;
- In 90% of FFE + literacy schools, 88% of FFE only schools and 97% of comparison schools, all pupils had an exercise book;
- In 27% of FFE + literacy schools, 17% of FFE only and 21% of comparison schools, all pupils in the class observed had an eraser.

More details can be found in Technical Appendix 4, pp 66-75²¹.

The project has undoubtedly contributed to teachers and pupils having improved access to some teaching and learning materials. It is to be noted that they now appear to be at a similar level to pupils in the comparison schools, at least as far as the most basic classroom equipment and materials are concerned, as observed during the lesson observation sessions. It is not possible to know from the information gathered whether the project schools are actually better equipped than the comparison schools in terms of less basic teaching and learning materials, such as syllable charts and other books and equipment, although this is likely to be the case.

²¹ Technical Appendix 4 (classroom observation) is available in electronic form: pdf file entitled 03_ADPE_TABLES_COBSERV_PDF__c

Number of awards given to students. Final target: 30,000

Midterm target: 19,500

Midterm results (March 2019): 17,552

Midterm target 90% achieved

Number of awards given to teachers. Final target: 7,500

Midterm target: 4,875

Midterm results (March 2019): 7,845

Final target 105% achieved

Providing awards to both pupil and teachers has been used since the beginning of the project as an incentive to encourage both to make efforts in teaching and learning. During the first phase of the project, awards were presented at high profile ceremonies and district-level meetings, involving large numbers of schools, known as the “Olympiads”. Despite the success of these large-scale events in the district capitals in raising the visibility of the project and promoting the importance of academic achievement, it was felt that local communities, parents and guardians were excluded from them and that they were also costly and time consuming to organize. Recently, the project has focused on more local events and competitions, which allow local schools and families to take part and are considered more sustainable.

Rewards can vary from a pencil or a reading book to a bicycle in certain rare cases. Some of the awards encountered during fieldwork were poor quality plastic toys and school equipment, which seems regrettable, given current awareness of the environmental impact of limited use plastics and the potentially toxic composition of some low-grade plastics.

Analysis of the project monitoring records reveals that the final target for awards to teachers has been surpassed. The midterm target for awards to pupils has been 90% achieved. Of the wide-ranging subjects evoked by respondents during interviews and focus groups, awards were mentioned less frequently than during previous evaluations, suggesting that they are less of a focus than during the first phase of the project. However, they are still perceived by school directors and district education officials as being a valuable means of encouraging both teachers and students.

Before distributing further awards, it would be useful to assess their usefulness in providing real incentives and also whether the types of awards currently distributed are the most appropriate for this purpose and reflect ADPP’s wider objectives of environmental protection and sustainability, or whether these might be replaced by more sustainable alternatives.

Literacy Instructional Materials (1.1.3)

Number of grade 1-3 children receiving literacy books. Final target: 22,300

Midterm target: 16,450

Midterm results (March 2019): 22,111

Midterm target 134% achieved

Number of supplementary literacy materials produced and distributed to project schools. Final target: 159,500

Midterm target: 81,950

Midterm results (March 2019): 199,280

Final target 125% achieved

Within the scope of the literacy intervention supported by CE, a range of literacy books and materials in local languages Xichangana and Xirhonga and in Portuguese has been produced by the project literacy team, in partnership with experts from INDE and with support from Cambridge's reading experts. Targets for distribution of these have been surpassed. The production of the materials was an enormous undertaking, which mobilized the literacy team for the first months of the project. As part of this effort, the orthography of the two local languages was revised and standardized. The materials produced include textbooks, resource materials for teachers, reading books for students to take home and story books for teachers to read aloud to their class. The midterm evaluation revealed widespread praise for these innovative materials, for the quality of their content and their culturally appropriate messages. They are recognized as filling a much-needed gap in the resources available to teachers. District education official and head teachers expressed great appreciation of these materials, as did teachers and reading coaches. Government officials express gratitude that the project and CE have been able to provide these materials, which are currently beyond the scope of their stretched budgets. In particular, they are delighted that the project is promoting bilingual education, which has long been part of the MINEDH's plans but without resources being available to implement this.

On examination, the books and other materials appear reasonably durable, without being excessively expensive. They are attractively presented and contain pictures and themes which reflect the daily lives of the pupils who will use them. There were a number of orthographic errors in the first printing which have been corrected in preparation for a second printing in the future.

Skills and Knowledge of Teachers (1.1.4)

There are two major dimensions of the project interventions which target improving the skills and knowledge of teachers: i) in-service training (CPD) of primary school teachers in schools in the project districts as part of the literacy intervention coordinated and supported by CE and ii) the initial teacher education (pre-service teacher training) delivered by the 11 teacher training colleges run by ADPP in partnership with the MINEDH in each

province of Mozambique. The skills and knowledge of primary school teachers in project schools are considered first. The skills and knowledge of trainee teachers in EPFs are addressed in the following section.

In-service training in literacy of primary school teachers

Indicator 6 (Output): Number of teachers/educators/teaching assistants trained or certified (primary schools) as a result of USDA assistance. Final target: 879

Midterm target: 586

Midterm results (March 2019): 557

Midterm target 95% achieved

Number of teachers who receive in-service training as a result of USDA assistance. Final target: 293

Midterm target: 293

Midterm results (March 2019): 557

Final target 190% achieved

Indicator 5 (Outcome): Number of teachers/educators/teaching assistants in target schools who demonstrate use of new and quality teaching techniques or tools as a result of USDA assistance. Final target: 525

Midterm target: 350

Midterm results (March 2019): 451 (estimate based on 93/240 classroom observations)

Final target 129% achieved

586 teachers in project schools have received training in phonics and their use of in teaching reading and writing. The project intends to train further teachers in 2020. This is a significant innovation for Mozambique, where phonics has not been commonly used. In complement to the training, 27 reading coaches have been recruited by the project to provide ongoing support in the classroom. The reading coaches are themselves experienced primary school teachers, some retired, some on career breaks from their teaching functions, who were identified through a competitive process as possessing the necessary skills and attributes for this role. They have received 2.5 days' training prior to commencing their role and ongoing training and support from Cambridge and the literacy team. The reading coaches were considered by most informants to bring considerable added value to the program. The reading coaches themselves reported observing positive changes within the classrooms they support. One reading coach explained:

These days we have pupils who are in the second or third grade who already read perfectly and write perfectly....in Rhonga. We can see the difference through the assessments we do, because each term we do a mini-EGRA ..., also the reports we receive at the district level in the schools where the reading coaches are working it's different from the other schools, yes, the performance is better.

Informants regretted that there were not more reading coaches and supervisors since these are rather thinly spread, particularly in some districts; in particular more supervisors are needed in order to have more time in the field.

The literacy team and reading coaches reported having a good working relationship with GoM officials at all levels; this was confirmed during interviews with national, provincial and district-level officials. One demonstration of such positive relationships is the fact that the District Education Office (SDEJT) focal points for the literacy program are invited to training sessions for reading coaches and teachers, where they participate as trainers. This is a rare case of good practice in working in partnership with government staff and building capacity within government institutions. The literacy team would like district education directors to also take part in these meetings, to build their awareness of pupil-centered, phonics-based reading methods, in the hope they will then serve as champions of these at strategic levels within the education system.

The literacy component has an M&E plan and a full-time M&E officer who systematically collects monitoring information from EGRAs and mini-EGRAs regularly conducted by samples of the students in FFE + literacy schools and from observations of lessons delivered by the teachers receiving literacy coaching. Reports of these evaluations and lesson observations are produced regularly by the M&E officer who, despite being dedicated to the literacy project, is physically based in the M&E office which, at least in theory, allows integration with wider M&E activities. Given the gap in M&E management and provision over recent times, in reality she has frequently been left to work alone (this should change under the management of the newly appointed M&E coordinator). In contrast, the staff member responsible for M&E of the nutrition education project is currently based in the nutrition education team office, although the plan is to transition the full nutrition education M&E responsibility to the FFE2 M&E team, integrating the work and personnel²². Whereas M&E of the nutrition education project is has been closely supported by implementing partner WISHH, the implementing partner of the literacy intervention, CE focusses on providing expert technical support but little M&E of the intervention. Providing some robust technical support to the literacy M&E officer would be useful, given her position between two operational units and relative lack of support recently, and allow the information in her reports to be more widely exploited for program improvements.

85% of teachers surveyed declared they had received some form of training as part of the project. Overall, 94% of these considered the training to be either “good” or “excellent”. 72% of teachers from FFE + literacy schools and 64% from FFE only schools declared they had received visits to provide professional support visits whilst at

²² WISHH representatives explain that, at the beginning of FFE2, the nutrition education component attempted to transfer responsibility for M&E of its activities to the FFE2 M&E team. However, given the lack of stability within the M&E team, WISHH maintained support until the M&E team had the capacity to take on this role, continuing to support M&E of nutrition education activities for longer than originally planned. The nutrition education sustainability plan specifically addresses integration of M&E into the wider project M&E processes: this has begun, under the new M&E coordinator and team.

school; 94% of these teachers considered the professional support to be either “good” or “excellent”. 40% of teachers in FFE + literacy schools and 12% in FFE only school claim to teach in Xichangana or Xirhonga.

For more details, refer to Technical Appendix 2, pp 40-49.

The survey of project teachers included several items designed to assess the knowledge of literacy techniques of those teachers benefiting from the literacy training.

Of those teachers surveyed, 71% of those from FFE + literacy schools and 39% of those from FFE only schools stated they had received literacy training as part of the project. They overwhelmingly stated they had found the training useful (FFE + literacy 100%; FFE only 93%). It seems rather surprising that teachers from FFE only schools claimed to have received literacy training; it is possible these teachers were, in fact, referring to other forms of support such as training in the use of the materials provided for after-school clubs.

Teachers’ knowledge of some basic concepts of literacy was evaluated by asking them to give definitions of five commonly-used terms using multiple choice questions. Table 7 below presents the results of this task for teachers in FFE + literacy schools and FFE only schools at midterm. The baseline results are also presented. However, at baseline, 65% of respondents did not reply to these questions, making comparisons problematic. At midterm, in 4 out of 5 cases, the FFE + literacy teachers performed better than the FFE only teachers, in some cases by a considerable margin. Only in the case of “comprehension” did the FFE only teachers do better. Overall, the FFE + literacy teachers have a better understanding of these terms. It is striking that even the FFE + literacy teachers did not perform particularly well on these tasks: in no case did more than 62% of either group choose the correct response.

For more details, refer to Technical Appendix 2, pp 70-90.

Correct response	Midterm FFE + literacy (% of correct responses)	Midterm FFE only (% of correct responses)	Baseline: all teachers (% of correct responses)
Phonological awareness	52%	32%	8% (23% of responses)
Phonetics	42%	23%	9% (26% of responses)
Fluency	62%	52%	30% (86% of responses)
Vocabulary	57%	52%	28% (80% of responses)
Comprehension	58%	62%	31% (89% of responses)

Table 7: teachers’ ability to define basic terms of literacy

When asked about the different stages of the instructional model “*eu faço, nos fazemos, vocês fazem*” (I do, we do, you do), taught to teachers as the basis of the literacy intervention, 52% of teachers from FFE + literacy

schools were able to correctly define the stage “*eu faço*”, as opposed to 48% of those from FFE only schools (compared with 16% overall at baseline). 70% of teachers from FFE + literacy schools were able to correctly define the stage “*nos fazemos*”, compared with 58% of those from FFE only schools (22% overall at baseline). 62% of teachers from FFE + literacy schools were able to correctly define the stage “*vocês fazem*” compared with 53% of those from FFE only schools (23% overall at baseline) (see figure 5 below).

For more details, refer to Technical Appendix 2, pp 90-97.

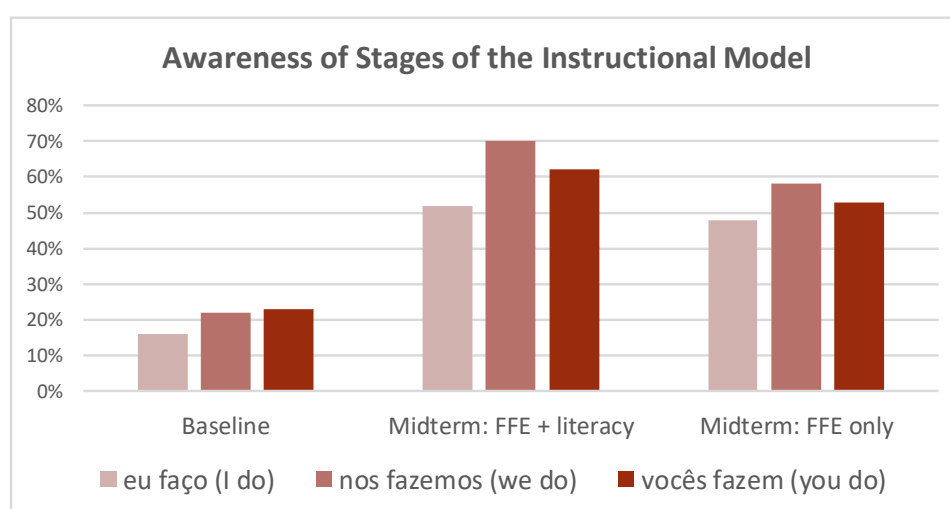


Figure 5: Teachers’ ability to define the stages of the Instructional Model at baseline and midterm

When asked about their experience of reading, 89% of teachers in FFE + literacy schools and 92% in FFE only schools claim to read every day to prepare their teaching (compared with 30% overall at baseline) whereas 9% of teachers in FFE + literacy schools and 8% in FFE only schools claim to read “about once a week” (3% overall at baseline) (see figure 6 below). 49% of teachers in FFE + literacy schools and 61% in FFE only schools claim to read for pleasure every day (compared with 24% overall at baseline), with 42% of teachers in FFE + literacy schools and 31% in FFE only schools stating they do so “about once a week” (compared to 7% at baseline) (see figure 7 below). There appears to be a large increase in teachers stating they read to prepare their teaching and for pleasure between baseline and midterm.²³

²³ It should be noted that i) only 32% of teachers surveyed responded to this question at baseline; ii) the survey was administered at midterm in the form of an interview, with the teachers’ responses captured by an enumerator on a smart phone, whereas at baseline teachers filled in a questionnaire themselves. It is probable that this difference influenced the teachers both to reply and to reply with the perceived correct or approved response at midterm.

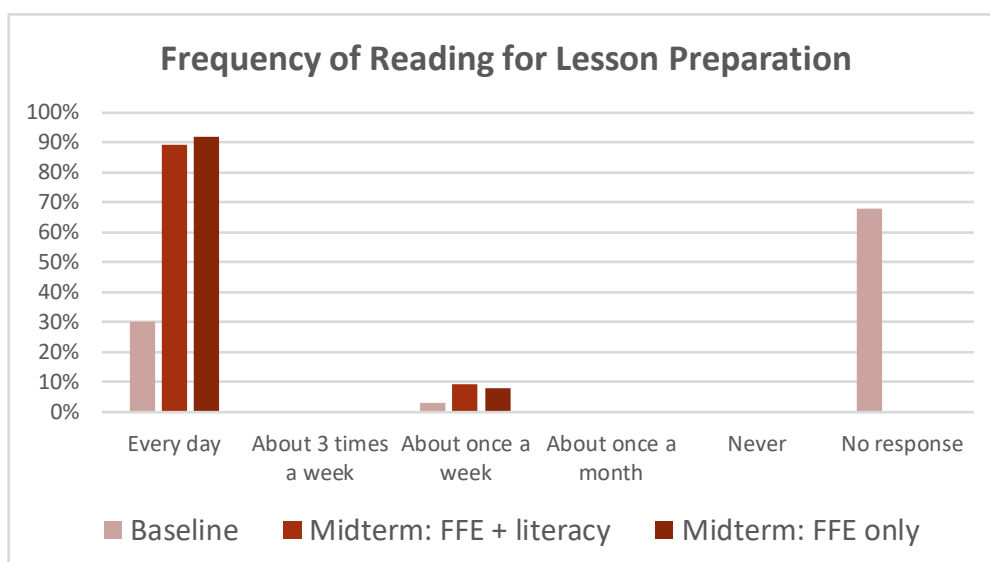


Figure 6: Teachers' reporting of reading for lesson preparation

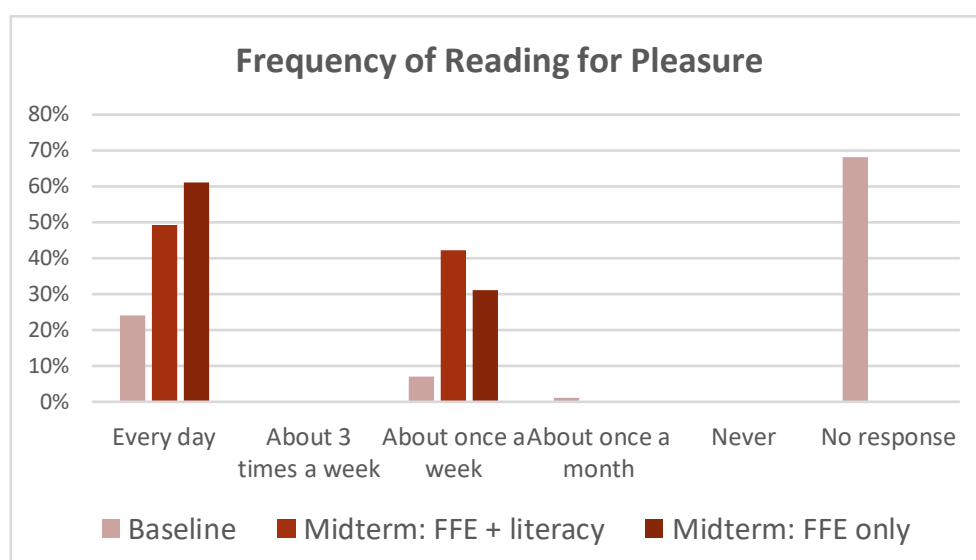


Figure 7: Teachers' reporting of reading for pleasure

When asked in which language they read for pleasure, Portuguese was the most frequently cited (80% FFE + literacy; 91% FFE only), followed by Xichangana (14% FFE + literacy; 4% FFE only) and Xirhonga (2% FFE + literacy; 1% FFE only); 4% of each group stated they read for pleasure in other languages. Even allowing for a certain degree of probable overreporting, at midterm, teachers in FFE + literacy schools are significantly more likely to read for pleasure in local languages than those in FFE only schools. At baseline, the responses were Portuguese 26%, Xichangana 4%, Xirhonga 1%, with 68% of nonresponses.

When asked to evaluate their own reading and comprehension ability in Portuguese, 64% of teachers in FFE + literacy schools and 71% in FFE only schools claim to be able to read fluently and understand everything, with 35% of teachers in FFE + literacy schools and 27% in FFE only schools stating they can read fluently and have some difficulty understanding certain words. Only 1% of teachers overall claimed to have many difficulties understanding what they are reading and 0% stated that they read rarely and had many comprehension problems (see figure 8 below). When asked to evaluate their own reading and comprehension ability in Xichangana or Xirhonga, 29% of teachers in FFE + literacy schools and 26% teachers in FFE only schools claim to be able to read fluently and understand everything, with 48% of teachers in FFE + literacy schools and 38% in FFE only schools stating they can read fluently and have some difficult understanding certain words. 13% of teachers in FFE + literacy schools and 22% of teachers in FFE only schools overall claimed to have many difficulties understanding what they are reading and 10% of teachers in FFE + literacy schools and 14% in FFE only stated that they read rarely and had many comprehension problems (see figure 9 below).

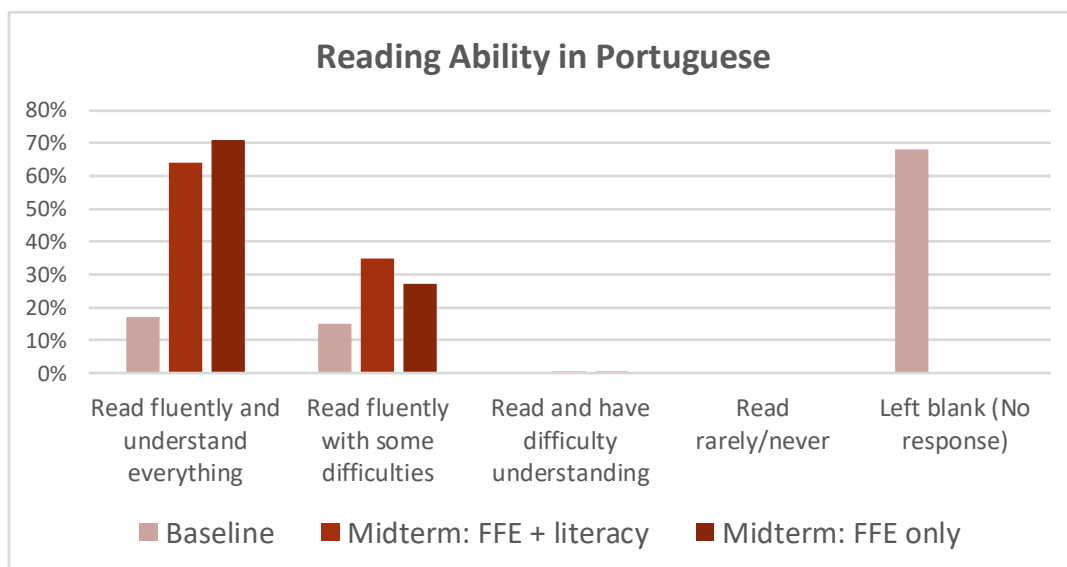


Figure 8: Teachers' reporting of own reading ability in Portuguese

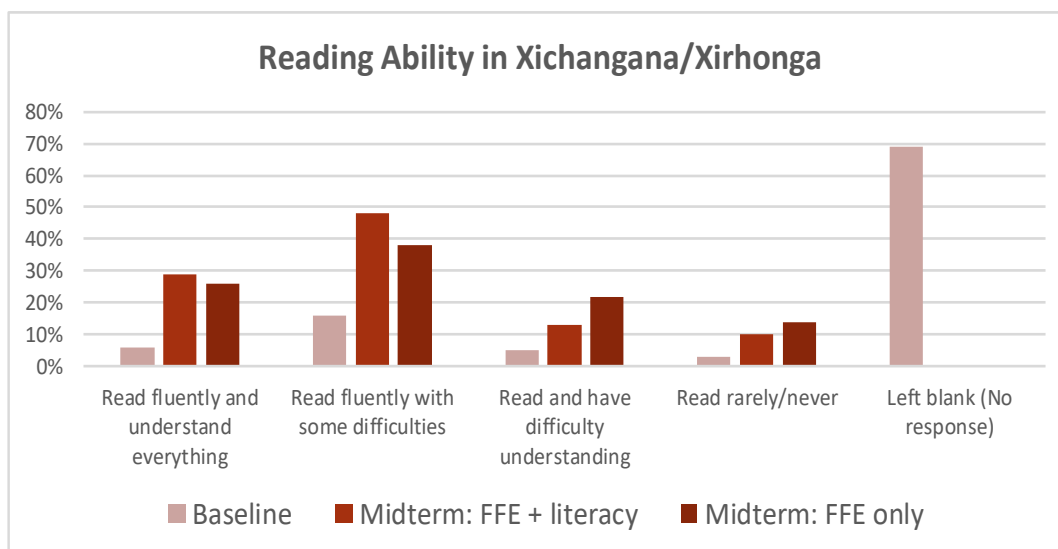


Figure 9: Teachers' reporting of own reading ability in Xichangana/Xirhonga

When asked about a variety of literacy instruction materials distributed by the project to teachers, the following responses were obtained (see figures 10, 11 and 12 below):

- Materials to improve literacy: stated to have been received by 62% of teachers in FFE + literacy schools and 77% in FFE only schools; 91% of teachers in FFE + literacy schools and 96% of teachers in FFE only schools claim to have found them useful;
- Framework to facilitate lesson planning: stated to have been received by 5% of teachers in FFE + literacy schools and 37% in FFE only schools; 95% of teachers in FFE + literacy schools and 100% of teachers in FFE only schools claim to have found them useful;
- Guide of methodological suggestions: 98% of teachers in FFE + literacy schools and 78% in FFE only schools claim to use this; 80% of teachers in FFE + literacy schools and 71% of teachers in FFE only schools claim to use it every day; 20% of teachers in FFE + literacy schools and 29% of teachers in FFE only schools claim to use it at least once a week;
- Letter- syllable- and word-cards: 88% of teachers in FFE + literacy schools and 78% in FFE only schools claim to use these in their teaching; 65% of teachers in FFE + literacy schools and 43% of teachers in FFE only schools claim to use them every day; 35% of teachers in FFE + literacy schools and 57% of teachers in FFE only schools claim to use them at least once a week;
- "Teacher read alouds" (stories for teachers to read aloud in class): 98% of teachers in FFE + literacy schools and 100% in FFE only schools claim to use these in their teaching; 73% of teachers in FFE + literacy schools and 67% of teachers in FFE only schools claim to read them aloud about 3 times a week; 24% of teachers in FFE + literacy schools and 33% of teachers in FFE only schools claim to use them at least once a week;

- Reading books for pupils: 100% of teachers in FFE + literacy schools and 100% in FFE only schools claim to use these in their teaching; 88% of teachers in FFE + literacy schools and 100% of teachers in FFE only schools claim to use them every day; 12% of teachers in FFE + literacy schools claim to use them at least once a week.

For more details refer to Technical Appendix 2, pp 113- 139

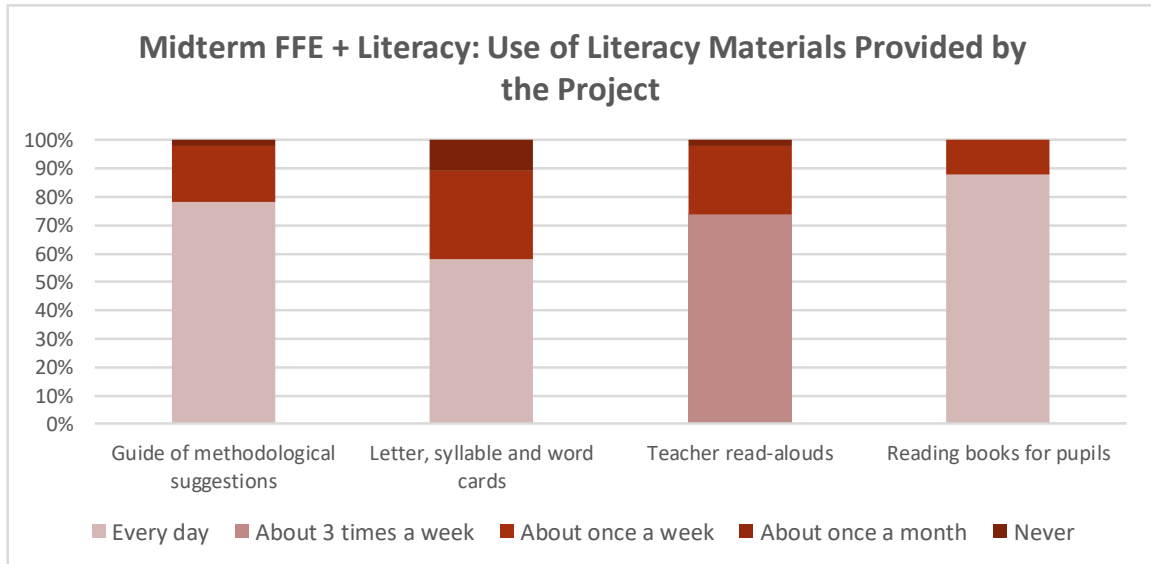


Figure 10: Use of literacy materials provided by the project (FFE + literacy at midterm)

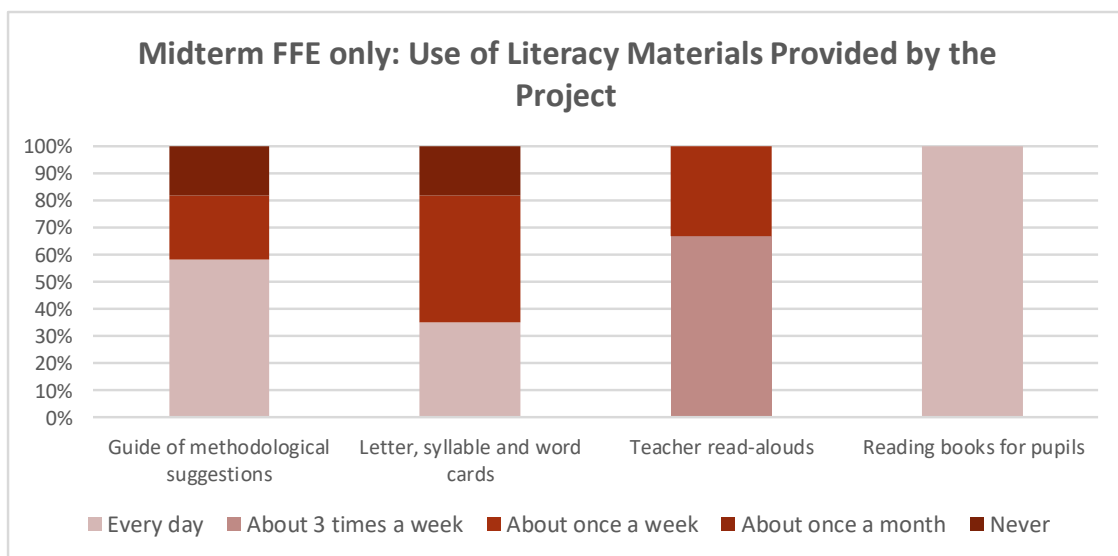


Figure 11: Use of literacy materials provided by the project (FFE only at midterm)

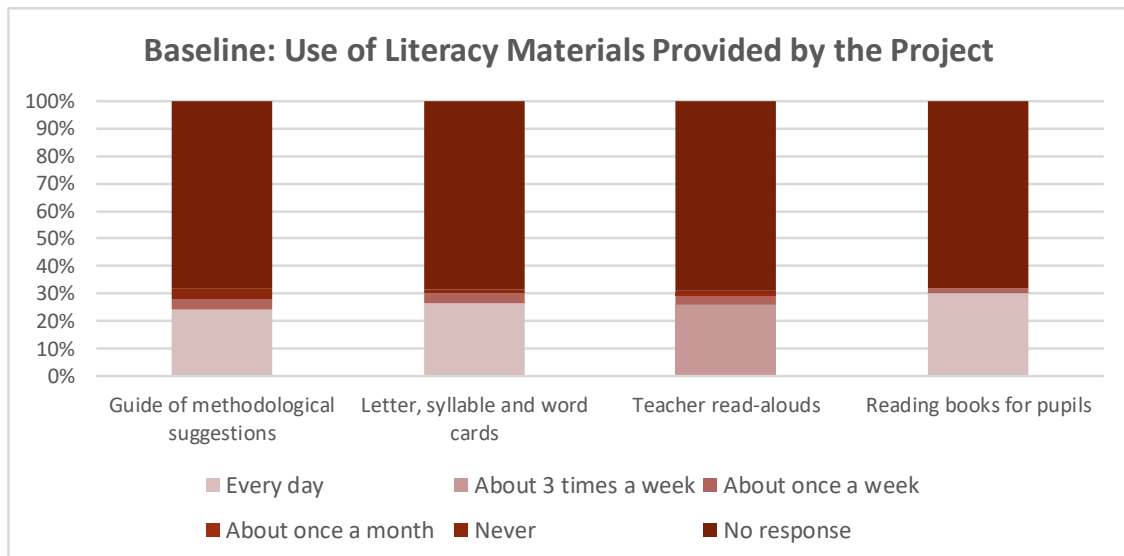


Figure 12: Use of literacy materials provided by the project (FFE + literacy and FFE only at baseline)

It should be noted that, in addition to the materials developed by the project literacy team and supplied to schools taking part in the literacy program, all project schools receive the teaching and learning materials supplied by the project to support extracurricular clubs. It is probable that, in responding to questions about materials received, teachers did not differentiate between these two types of materials; indeed, teachers in FFE only schools are likely to have assumed they were being questioned about the only materials they received, those designed to support the extracurricular learning clubs. With hindsight, the teacher questionnaire should have been designed so as to avoid this confusion (for example, by presenting images of the materials referred to in each question).

Although these responses are encouraging and suggest teachers are making good use of the materials provided, it is not possible to know for sure whether their responses referred to the materials provided by the literacy intervention or the general materials supplied as support to learning clubs.

Teachers were asked how much individual reading time they allowed their pupils in class; they were requested to select the response which best corresponded out of four options. Figure 13 shows their responses, compared with those from the baseline survey (relatively few teachers responded to this question at baseline).

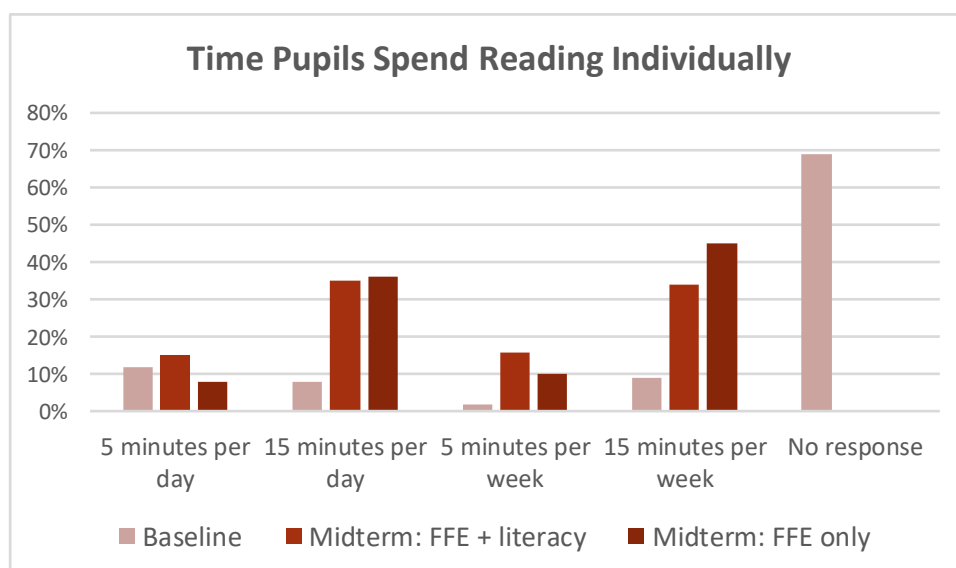


Figure 13: time pupils spend reading individually in class at baseline and midterm

No particular tendency appears to emerge from these results, except for the fact that, at midterm, “15 minutes per day or more” and “15 minutes per week or more” are by far the most frequently selected options (the range of options available may be considered not to constitute a clear continuum from more to less and may have confused some respondents).

As at baseline, teachers were presented with a series of binary options, designed to test their familiarity with modern classroom practice and approaches to reading²⁴. Their responses are presented in table 8, along with those from baseline (when, unfortunately, there was a high percentage of non-responses).

Table 8 below presents teachers’ responses at midterm and baseline. The options which are compatible with modern classroom practice and approaches (so are considered as “correct” for the present purposes) are shaded. The responses do not reveal consistent changes, either between baseline and midterm or between the teachers in FFE + literacy and FFE only schools. It is interesting that there is little change since baseline overall. The majority of teachers chose the “correct” response (modern approach) in five of the seven cases. However, they remain fairly evenly split on the questions of whether or not grade one pupils are too young to be responsible for handing out books and whether young pupils learn to read best in front of the blackboard or

²⁴ These were adapted from the Learning Gains Study used as part of the EGRA developed in Malawi by USAID Malawi and RTI, 2016.

when holding a book and at their own pace. It is notable that these “traditional” attitudes, which suggest a lack of exposure to modern pedagogical theory and practice, have not changed amongst the FFE + literacy teachers²⁵.

For more details refer to Technical Appendix 2, pp 141- 157.

Answer A	Midterm FFE + literacy	Midterm FFE only	Baseline %	Answer B	Midterm FFE + literacy	Midterm FFE only	Baseline %	N/R at baseline%
It is important to allow pupils to take their books home (baseline)/use books regularly(midterm) ²⁶	92%	92%	23% (77% of responses)	Pupils should not take schoolbooks home (baseline)/ books should be used occasionally s as to preserve them for the future (midterm).	8%	8%	7% (23% of responses)	70%
Pupils cannot practice reading at home because the majority of parents cannot read	15%	14%	2% (6% of responses)	Pupils should practice reading at home, even if their parents cannot read	85%	86%	29% (94% of responses)	69%
Grade one pupils are too young to be responsible for handing out books	47%	43%	12% (46% of responses)	Grade one pupils can learn to hand out books	53%	57%	14% (54% of responses)	74%
Pupils should be seated in rows facing the front so they can take part in the class	26%	32%	7% (24% of responses)	The way pupils are seated should be adapted to suit different activities	74%	68%	22% (76% of responses)	71%
Preparing classes in advance is necessary for good classroom management	97%	96%	27% (93% of responses)	Preparing classes in advance does not help much with classroom management	3%	4%	2% (7% of responses)	71%
Young pupils learn best when reading is done from the blackboard	56%	44%	12% (41% of responses)	Young pupils learn to read best when they are holding a book and can learn at their own pace	44%	56%	17% (59% of responses)	71%
Independent practice is not useful for young pupils because they make mistakes	25%	18%	3% (11% of responses)	Being able to make mistakes during independent practice is an important part of learning	75%	82%	25% (89% of responses)	72%

Table 8: Familiarity with modern classroom practice and approaches to reading: teachers' responses

Teachers were asked how they monitor pupils' progress in reading and how they record this, through a series of multiple-choice options; they were asked to choose all which applied. Their responses are presented in figures 14 and 15 respectively, along with those from baseline (when, unfortunately, there was a high percentage of non-responses).

²⁵ As pointed out in the baseline report, when questions are framed with focus on attitudes rather than actual practice, teachers may select what they perceive to be the desired response, rather than one which reflects their actual practice.

²⁶ The formulation of this item (both A and B) was changed at midterm, at the request of the literacy team).

For more details refer to Technical Appendix 2, pp 386-394.

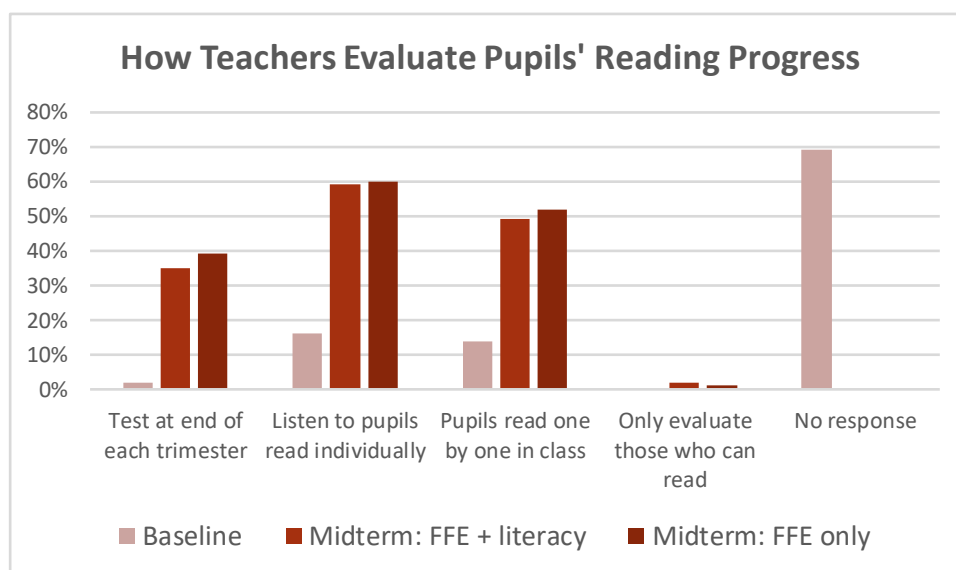


Figure 14: How teachers evaluate students' progress in reading – midterm and baseline

Regarding their monitoring of students' reading, teachers' responses at midterm were fairly similar over the two groups, with the largest number claiming they listen to pupils read individually to help them and to monitor their progress, the most modern, "correct" response, which has been the method promoted by reading coaches and fewer claiming to use a test, the more "traditional" and less effective method. It is striking that greater difference was not found between the two groups. Comparisons with the baseline results are not meaningful: at baseline, once again, a large number of teachers (69%) did not respond. Furthermore, at baseline, teachers were asked to choose only one response.

For more details refer to Technical Appendix 2, pp 394-401.

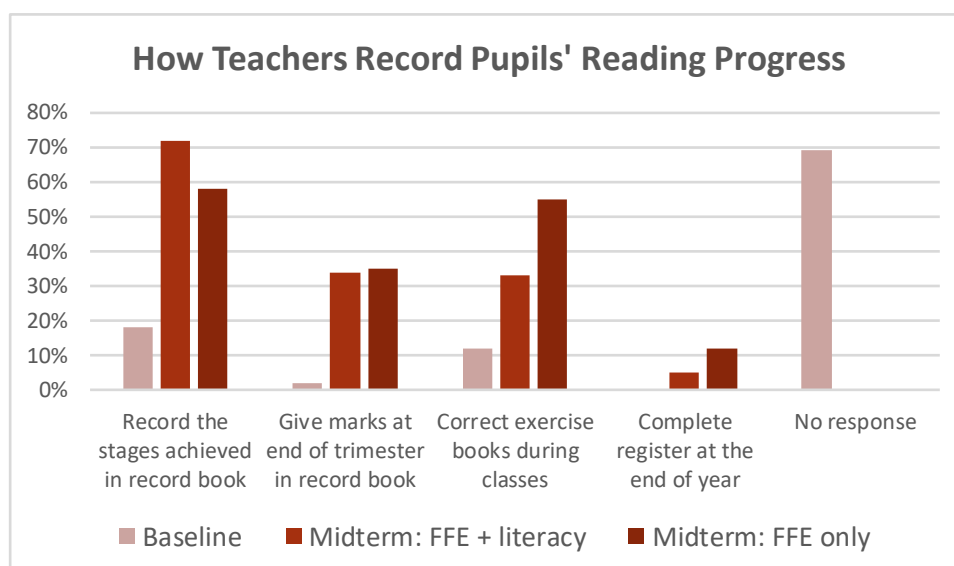


Figure 15: How teachers record students' progress in reading – midterm and baseline

Regarding their recording of students' reading progress, significantly more FFE + literacy group teachers claimed to record the stages of reading attained by individual pupils on an ongoing basis, the "correct" method promoted by the literacy coaches. They were also less likely to claim to correct exercises books during the lesson (since the question was about monitoring reading this response suggests a certain confusion amongst teachers who chose this response) or to fill in the record book at the end of the year²⁷. Taken together, these results suggest that the literacy coaching has resulted in increased knowledge of effective recording methods amongst participating teachers. Once again, comparisons with the baseline results are not meaningful, for the reasons cited above.

Lessons were observed in 240 project schools and comparison schools in order to attempt to evaluate whether teachers in project schools were using the "new and quality teaching techniques or tools" as a result of the literacy training. The observation instrument was developed by Simone Doctors with support from the project literacy team and is designed to be used by non-specialist observers, who are not teacher educators. All lesson observation was performed by supervisors, who had been trained in this activity and selected for their capacity to perform it. Several of the supervisors were themselves trained teachers. During fieldwork in FFE + literacy schools, whenever possible a class taught by a teacher who had benefited from the literacy intervention was observed²⁸. The results of the lesson observation are presented below following the structure and order of the

²⁷Given that teachers were asked to select all which apply, the interaction between these choices is not clear (filling in a record book at the end of the school year is a very inadequate form of monitoring reading progress; however, it would be a useful complement to recording ongoing progresses and stages attained). With hindsight, it would be more informative to ask teachers to select only one response.

²⁸ Enumerators were given instructions as to which teachers to observe in project and comparison schools, according to a structured protocol. Where possible, they were told to observe a teacher trained by the project

items on the observation instrument (although the data was captured using ODK on smartphones, observers also had paper copies of the observation instruments, to facilitate their navigation through the instrument).

Elements verified before the lesson began

- The classroom walls were decorated with T&L materials in the case of 75% of FFE + literacy teachers, 56% of FFE only teachers and 60% of comparison school teachers gave reading exercises during the lesson; this was the case for 74% of teachers teaching in local languages and 61% of teachers using Portuguese;
- Pupils' work was displayed on the walls in the case of 44% of FFE + literacy teachers, 26% of FFE only teachers and 23% of comparison school teachers; this was the case for 44% of teachers teaching in local languages and 28% of teachers using Portuguese;
- 96% of FFE + literacy teachers, 94% of FFE only teachers and 94% of comparison school teachers had a written lesson plan; this was the case for 89% of teachers teaching in local languages and 96% of teachers using Portuguese;
- 73% of FFE + literacy teachers, 75% of FFE only teachers and 67% of comparison school teachers explained their lesson plan to the enumerator; this was the case for 79% of teachers teaching in local languages and 70% of teachers using Portuguese;
- 78% of FFE + literacy teachers, 78% of FFE only teachers and 73% of comparison school teachers had evidence of their continuous assessment of students (in the form of record books, assessment sheets, etc.) or other; this was the case for 88% of teachers teaching in local languages and 72% of teachers using Portuguese;
- 96% of FFE + literacy teachers, 96% of FFE only teachers and 91% of comparison school teachers arrived in the classroom on time; this was the case for 98% of teachers teaching in local languages and 93% of teachers using Portuguese;

Conditions in the classroom

- The classroom was "perfectly clean" in the case of 49% of FFE + literacy teachers, 52% of FFE only teachers and 40% of comparison school teachers; this was the case for 45% of teachers teaching in local languages and 48% of teachers using Portuguese; The classroom was "more or less clean but could have been better" in the case of 37% of FFE + literacy teachers, 39% of FFE only teachers and 47% of comparison school teachers; this was the case for 39% of teachers teaching in local languages and 41% of teachers using Portuguese; The classroom was "not clean" in the case of 14% of FFE + literacy teachers, 9% of FFE only teachers and 13% of comparison school teachers; this was the case for 15% of teachers teaching in local languages and 11% of teachers using Portuguese;

teaching in local languages; where this was not possible, they should observe a teacher trained by the project teaching in Portuguese. There were a series of further instructions in case neither option was possible.

- The desks were arranged in such a way that all pupils could participate in the lesson in the case of 80% of FFE + literacy teachers, 73% of FFE only teachers and 77% of comparison school teachers; this was the case for 83% of teachers teaching in local languages and 74% of teachers using Portuguese;
- Pupils with SEN were placed appropriately in the case of 44% of FFE + literacy teachers, 32% of FFE only teachers and 53% of comparison school teachers; this was the case for 50% of teachers teaching in local languages and 40% of teachers using Portuguese;

During the lesson

- The lesson had an introduction, development and a conclusion in the case of 94% of FFE + literacy teachers, 96% of FFE only teachers and 97% of comparison school teachers; this was the case for 95% of both teachers teaching in local languages and teachers using Portuguese;
- The teacher corrected the students' homework in the case of 76% of FFE + literacy teachers, 73% of FFE only teachers and 80% of comparison schoolteachers; this was the case for 82% of teachers teaching in local languages and 74% of teachers using Portuguese;
- The teacher was considered by the observer to use the instructional model "I do, we do, you do" in the case of 91% of FFE + literacy teachers, 95% of FFE only teachers and 91% of comparison school teachers; this was the case for 92% of teachers teaching in local languages and 93% of teachers using Portuguese;²⁹
- The teacher was considered by the observers to demonstrate knowledge and mastery of the content being taught in the case of 99% of FFE + literacy teachers, 99% of FFE only teachers and 99% of comparison school teachers; this was the case for 100% of teachers teaching in local languages and 98% of teachers using Portuguese;
- 83% of FFE + literacy teachers, 78% of FFE only teachers and 30% of comparison school teachers gave oral exercises during the lesson; this was the case for 85% of teachers teaching in local languages and 59% of teachers using Portuguese;
- 77% of FFE + literacy teachers, 81% of FFE only teachers and 29% of comparison school teachers gave reading exercises during the lesson; this was the case for 67% of teachers teaching in local languages and 63% of teachers using Portuguese;

²⁹ During supervisor training in administering the classroom observation, the instructional model was explained to the supervisors, and they were told criteria to look out for in their observations. However, they are not education specialists (although many are trained teachers) and had not had formal, sustained training in this instructional model. It is therefore possible that they applied the criteria quite loosely. The fact that over 90% all three groups of teachers were categorised as using the model, whereas in principle only those in the FFE + literacy group would have had direct access to it, suggests this. This highlights a limitation of observation checklists where the observer is asked to choose between two binary responses, with no possibility of a nuanced response.

- 75% of FFE + literacy teachers, 75% of FFE only teachers and 94% of comparison school teachers gave written exercises during the lesson; this was the case for 73% of teachers teaching in local languages and 84% of teachers using Portuguese;
- The teacher monitored the students' activities to verify the lesson objective were being met in the case of 99% of FFE + literacy teachers, 96% of FFE only teachers and 76% of comparison school teachers; this was the case for 98% of teachers teaching in local languages and 92% of teachers using Portuguese;
- The majority of pupils (80%) participated actively in the lesson in the case of 89% of FFE + literacy teachers, 79% of FFE only teachers and 96% of comparison school teachers; this was the case for 88% of teachers teaching in local languages and 88% of teachers using Portuguese;
- The majority of pupils (80%) demonstrated they had understood what to do in the case of 92% of FFE + literacy teachers, 84% of FFE only teachers and 89% of comparison school teachers; this was the case for 92% of teachers teaching in local languages and 87% of teachers using Portuguese;

Communication and language

- The teacher showed mastery of Portuguese (when spoken) in the case of 99% of FFE + literacy teachers, 99% of FFE only teachers and 100% of comparison school teachers; this was the case for 98% of teachers teaching in local languages and 99% of teachers using Portuguese;
- The teacher showed mastery of Xichangana (when spoken) in the case of 96% of FFE + literacy teachers, 96% of FFE only teachers and 88% of comparison school teachers; this was the case for 97% of teachers teaching in local languages and 94% of teachers using Portuguese;
- The teacher showed mastery of Xirhonga (when spoken) in the case of 100% of FFE + literacy teachers, 100% of FFE only teachers and 89% of comparison school teachers; this was the case for 93% of teachers teaching in local languages and 100% of teachers using Portuguese;
- The teachers' language was appropriate for children, clear and precise, audible and friendly in the case of 96% of FFE + literacy teachers, 99% of FFE only teachers and 93% of comparison school teachers; this was the case for 98% of teachers teaching in local languages and 95% of teachers using Portuguese;
- The teachers refrained from using humiliating or abusive language in the classroom in the case of 97% of FFE + literacy teachers, 97% of FFE only teachers and 90% of comparison school teachers; this was the case for 100% of teachers teaching in local languages and 93% of teachers using Portuguese;
- The teacher took account of students with SEN when communicating in the case of 54% of FFE + literacy teachers, 47% of FFE only teachers and 50% of comparison school teachers; this was the case for 59% of teachers teaching in local languages and 47% of teachers using Portuguese;

Use of teaching and learning materials and resources

Some of the information about teaching and learning materials is reported in the section on Access to School Supplies and Materials above.

- 96% of FFE + literacy teachers, 87% of FFE only teachers and 87% of comparison school teachers used appropriate T&L materials to support their teaching; 98% of teachers teaching in local languages did so, compared with 87% of teachers teaching in Portuguese;
- 98% of FFE + literacy teachers, 97% of FFE only teachers and 99% of comparison school teachers wrote on the blackboard neatly and without errors; 97% of teachers teaching in local languages did so, compared with 98% of teachers teaching in Portuguese;
- 61% of FFE + literacy teachers, 58% of FFE only teachers and 56% of comparison school teachers used materials which were suitable for learnings with SEN; 68% of teachers teaching in local languages did so, compared with 55% of teachers teaching in Portuguese;

Classroom management

- 96% of FFE + literacy teachers, 92% of FFE only teachers and 89% of comparison school teachers maintained discipline while teaching without resorting to corporal punishment; 95% of teachers teaching in local languages did so, compared with 91% of teachers teaching in Portuguese;
- In the classes of 92% of FFE + literacy teachers, 75% of FFE only teachers and 93% of comparison school the majority of pupils (at least 80%) participated and was involved in the lesson; this applied in the case of 89% of teachers teaching in local languages did so, compared with 86% of teachers teaching in Portuguese;
- Classroom rules were in evidence in classes taught by 90% of FFE + literacy teachers, 90% of FFE only teachers and 87% of comparison schools; this was the case in classes taught by 91% of teachers teaching in local languages did so, compared with 89% of teachers teaching in Portuguese;
- 92% of FFE + literacy teachers, 97% of FFE only teachers and 96% of comparison school teachers were perceived to notice and respond to the learning needs of students with SEN; this was the case for both 95% of teachers teaching in local languages and in Portuguese;
- There was a balance between teacher talking time/presentation and student talking time/activities in 94% of FFE + literacy teachers, 96% of FFE only teachers and 91% of comparison school teachers were perceived to notice and respond to the learning needs of students with SEN; this was the case for 95% of teachers teaching in local languages and 93% of teachers teaching in Portuguese

Group work

- 42% of FFE + literacy teachers, 39% of FFE only teachers and 46% of comparison school teachers got students to work in groups; this was the case for 44% of teachers teaching in local languages and 41% of teachers using Portuguese;
- Of these, 90% of FFE + literacy teachers, 77% of FFE only teachers and 88% of comparison school teachers had students well positioned for group work; this was the case for 90% of teachers teaching in local languages and 83% of teachers using Portuguese;

- When used, teachers managed to monitor students' participation in group work in the case of 97% of FFE + literacy teachers, 97% of FFE only teachers and 100% of comparison school teachers ;this was the case for 100% of teachers teaching in local languages and 97% of teachers using Portuguese;
- When used, groups were well balanced in terms of gender, disability and SEN in the case of 87% of FFE + literacy teachers, 90% of FFE only teachers and 78% of comparison school teachers had students well position for group work; this was the case for 83% of teachers teaching in local languages and 86% of teachers using Portuguese;

Questions and feedback

- 89% of FFE + literacy teachers, 89% of FFE only teachers and 74% of comparison school teachers involved both boys and girls and gave them equal opportunities to participate; this was the case for 86% of teachers teaching in local languages and 83% of teachers using Portuguese;
- 87% of FFE + literacy teachers, 74% of FFE only teachers and 90% of comparison school teachers responded to students' questions; this was the case for 89% of teachers teaching in local languages and 81% of teachers using Portuguese;
- 82% of FFE + literacy teachers, 65% of FFE only teachers and 83% of comparison school teachers allowed pupils to express their own ideas in the classroom; this was the case for 80% of teachers teaching in local languages and 75% of teachers using Portuguese;
- 97% of FFE + literacy teachers, 96% of FFE only teachers and 99% of comparison school teachers gave appropriate, constructive feedback to girls and boys; this was the case for 97% of teachers teaching in both local languages and Portuguese;
- 98% of FFE + literacy teachers, 97% of FFE only teachers and 96% of comparison school teachers corrected exercises completed during the lesson; this was the case for 95% of teachers teaching in local languages and 98% of teachers using Portuguese;

Equality, equity and Special Education Needs

- 42% of FFE + literacy teachers, 38% of FFE only teachers and 59% of comparison school teachers adapted activities so as to be suitable for students with SEN; this was the case for 52% of teachers teaching in local languages and 43% of teachers using Portuguese;
- Students with SEN are able to participate actively in the case of 48% of FFE + literacy teachers, 42% of FFE only teachers and 50% of comparison school teachers; this was the case for 56% of teachers teaching in local languages and 43% of teachers using Portuguese;
- Students with SEN (including specially gifted pupils) are given differentiated tasks by 42% of FFE + literacy teachers, 31% of FFE only teachers and 39% of comparison school teachers; this was the case for 48% of teachers teaching in local languages and 33% of teachers using Portuguese;

- 96% of FFE + literacy teachers, 94% of FFE only teachers and 91% of comparison school teachers encourage girls and boys to participate equally; this was the case for 94% of teachers both teaching in local languages and using Portuguese;
- 91% of FFE + literacy teachers, 92% of FFE only teachers and 90% of comparison school teachers treat all students (boys, girls, with/without disabilities, of all linguistic groups) equally; this was the case for 94% of teachers teaching in local languages and 90% of teachers using Portuguese;

Conclusion and homework

- 78% of FFE + literacy teachers, 77% of FFE only teachers and 89% of comparison school teachers set homework involving reading and/or writing; this was the case for 82% of teachers teaching in local languages and 80% of teachers using Portuguese;
- 99% of FFE + literacy teachers, 94% of FFE only teachers and 89% of comparison school teachers finished the class on time; this was the case for 100% of teachers teaching in local languages and 92% of teachers using Portuguese.

Observation criteria	Teacher group			Language of instruction	
	FFE + literacy	FFE only	Comparison	Local languages	Portuguese
Before the lesson	%	%	%	%	%
Walls decorated with T&L materials	75	56	60	74	61
Pupils work displayed on walls	44	26	23	44	28
Written lesson plan	96	94	94	89	96
Explained lesson plan orally	73	75	67	79	70
Evidence of continuous assessment (record books, sheets, etc)	78	78	73	88	72
Teacher arrived in classroom on time	96	96	91	98	93
Conditions in the classroom					
Classroom "perfectly clean"	49	52	40	45	48
Classroom More or less clean – could be better	37	39	47	39	41
Classroom not clean	14	9	13	15	11
Desks arranged so all pupils can participate	80	73	77	83	74
Pupils with SEN placed appropriately	44	32	53	50	40
During the lesson					
Lesson had introduction, development and conclusion	94	96	97	95	95
Teacher corrected homework	76	73	80	82	74
Teacher used instructional model "I do, we do, you do"	91	95	91	92	93
Teacher showed knowledge and mastery of lesson content	99	99	99	100	98
Teacher gave students oral exercises	83	78	30	85	59
Teacher gave students reading exercises	77	81	29	67	63
Teacher gave students written exercises	75	75	94	73	84
Teacher monitored activities to see objectives being met	99	96	76	98	92
Majority of pupils (80%) participated actively in lesson	89	79	96	88	88
Majority of pupils (80%) knew what to do	92	84	89	92	87
Communication and language					
Teacher showed mastery of Portuguese (when spoken)	99	99	100	98	99
Teacher showed mastery of Xichangana (when spoken)	96	96	88	97	94
Teacher showed mastery of Xirhonga (when spoken)	100	100	89	93	100
Teachers' language appropriate for children	96	99	93	98	95
Teacher avoided humiliating or abusive language	97	97	90	100	93
Teacher took account of SEN students when communicating	54	47	50	59	47
Use of T&L materials and resources					
Teacher used appropriate T&L materials to support topic	96	87	87	98	87
Teacher wrote on blackboard neatly without errors	98	97	99	97	98
Teacher used materials suitable for learners with SEN	61	58	56	68	55
Classroom management					
Teacher maintained discipline without corporal punishment	96	92	89	95	91
Majority of pupils (80%) participated and involved in lessons	92	75	93	89	86
Classroom rules in evidence	90	90	87	91	89
Teacher perceived and responded to needs of SEN students	92	97	96	95	95
Balance of teacher presentation and student activities	94	96	91	95	93
Group work					
Teacher got students to work in groups	42	39	46	44	41
When used, students well positioned for group work	90	77	88	90	83
If used, teacher monitored student participation in groups	97	97	100	100	97
When used, groups well balanced (gender, disability, SEN)	87	90	79	83	86
Questioning and feedback					
Girls and boys involved - equal opportunities to participate	89	89	74	86	83
Teacher responded to students' questions	87	74	90	89	81
Teacher allowed students to express own ideas in classroom	82	65	83	80	75
Appropriate, constructive feedback given to girls and boys	97	96	99	97	97
Teacher corrected exercises completed during the lesson	98	97	96	95	98
Equality, equity and special needs					
Teacher adapted activities for students with SEN	42	38	59	52	43
Students with SEN able to participate actively	48	42	50	52	43
Students with SEN given appropriately differentiated tasks	42	31	39	48	33
Teacher encouraged girls and boys to participate equally	96	94	91	94	94
Teacher treated all students equally (girls, boys, disabled, SEN)	91	92	90	94	90
Conclusion and homework					
Teacher set homework involving reading and/or writing	78	77	89	82	80
Lesson finished on time	76	99	89	100	92

Table 9: Summary of classroom observation findings

Table 9 (above) presents a summary of these results³⁰. The 49 observation criteria observed, grouped in 10 sections, are considered to cover various aspects, although not necessarily an exhaustive description, of “good practice” in primary classroom teaching and to reflect a significant number of the “improved methods and tools” which the literacy training seeks to transmit. Analysis of the 49 observation criteria reveals certain trends:

- FFE + literacy teachers were more likely to exhibit these characteristics than FFE only teachers; a higher percentage of FFE + literacy teachers than of FFE only teachers demonstrated these characteristics in 57% of cases; a lower percentage of FFE + literacy teachers than of FFE only teachers demonstrated them in 20% of cases; the two groups performed equally in 22% of cases³¹.
- FFE + literacy teachers were more likely to exhibit these characteristics than comparison group teachers; a higher percentage of FFE + literacy teachers than of comparison group teachers demonstrated them in 61% of cases; a lower percentage of FFE + literacy teachers than of comparison group teachers demonstrated them in 35% of cases; the two groups performed equally in 4% of cases.
- FFE only and comparison group teachers were almost equally likely to exhibit these characteristics; FFE only teachers scored higher than comparison teachers in 49% of cases, whereas comparison group scored higher than the FFE only group in 45% of cases, with equal numbers in 6% of cases;
- Teachers observed teaching in Xichangana and Xirhonga were far more likely to demonstrate the characteristics than those teaching in Portuguese: a higher percentage of teachers using local languages demonstrated the characteristics than did the teachers teaching in Portuguese in 71% of cases; in 18% of cases, a higher percentage of teachers teaching in Portuguese demonstrated the characteristics; in 10% of cases they scored equally;
- Scores overall are surprisingly high (many over 90%). This suggests that the observers using the observation instrument applied it rather generously and tended to err on the side of positive assessments³²;
- The notable exception to the overall high scores concerns all criteria pertaining to students with special educational needs (SEN), which typically have scores closer to 50% for all three groups. This

³⁰ The greyed lines are not included in the total. The item regarding cleanliness of the classroom had a three-option response: only the first option (classroom “perfectly clean”) is included in the summary of good practice used for the present analysis.

³¹ This descriptive method of presenting the findings is not based on a formal difference-in-difference analysis and does not claim to be statistically significant.

³² This is probably due to the fact that the observers had limited experience of what good teaching using modern teaching methods looks like. It may also reflect, in the case of those who are trained teachers, a reluctance to be critical of ones peers. The author’s experience of observing classroom teaching in Mozambique does not on the whole support these optimistic assessments. However, it is hoped that, although the absolute assessment may be overstated, the comparative assessment of the difference groups still has some validity, since all groups were observed according to the same criteria and strict quality controls were applied.

suggests that awareness and practice of inclusivity are still not widely developed throughout the three groups of teachers observed.

Analysis of project records reveals that, as of the time of fieldwork in April 2019, 557 primary teachers had been trained by the project (of a final total objective of 879). Of the 240 lessons observed during lesson observations, 93 of these were taught by teachers in the FFE + literacy group. The scores obtained by these teachers on the 49 criteria assessed are presented in table 10 below, grouped according to the thematic sections used in the classroom observation instrument, and as a total overall score. It should be stressed that these scores are obtained by summing the number of teachers who were judged to fulfil the criteria, in most cases using a single “yes/no” question. Table 10 presents the total scores against the total possible scores for each section. FFE + literacy teachers observed scored most highly overall on classroom management (93%) and questioning and feedback (91%) and least highly on conditions in the classroom (58%) and equality, equity and special needs (64%). Overall the 93 FFE + literacy teachers observed scored a total of 3,471 points out of a possible total of 4,259 (81%). Considering that the 93 FFE + literacy teachers observed were a sample of the 557 teachers trained by the project to date, if this score of 81% is extrapolated the total group of 557 teachers trained to date, this would suggest that, if observed, 451 of these teachers would demonstrate the desired “new and quality teaching techniques or tools” referenced in MGD indicator 5. This estimation is therefore suggested as the midterm outcome for this indicator.

Observation criteria	Total score of FFE + literacy teachers	Potential total score	%
Before the lesson	428	558	77%
Conditions in the classroom	161	279	58%
During the lesson	815	930	88%
Communication and language	369	422	87%
Use of T&L materials and resources	237	279	85%
Classroom management	432	465	93%
Group work	146	210	70%
Questioning and feedback	421	465	91%
Equality, equity and special needs	297	465	64%
Conclusion and homework	165	186	89%
Total score of FFE + literacy teachers overall	3,471	4,259	81%

Table 10 scores of FFE + literacy teachers based on classroom observation criteria

Initial teacher education in teacher training colleges (EPFs)

Indicator 6 (Output): Number of teachers/educators/teaching assistants trained or certified (EPFs) as a result of USDA assistance. Final target: 3,704

Midterm results (March 2019): 6,059

Final target 164% achieved

Percentage of a student-teachers who improve their Portuguese language literacy skills during pre-service training. Final target: 80%

Midterm target (for year 2018): 80%

Midterm results (March 2019): 84%³³

Midterm target 105% achieved

Number of teachers in training as a result of USDA assistance. Final target: 3,512

Midterm target: 3,386 (accumulated)

Midterm results (March 2019): 6,895

Final target 204% achieved

Previous evaluations of the FFE project have underlined the important numerical and quality contribution the eleven ADPP-run teacher training colleges (EPFs) are making to Mozambique's teacher pool. At the time of the midterm evaluation, 6,059 trained teachers had graduated during the life span of the current phase of the project, exceeding the target by a considerable margin. The EPF model and curriculum are designed to train teachers able and motivated to work and teach in remote rural primary schools, to be enterprising, creative and adaptable, including finding solutions to teach in low resource environments. In addition to preparing future teachers for the reality of teaching in remote rural areas, the EPF program aims to teach practical skills, such as community development, building, small scale agriculture and husbandry. The midterm evaluation of FFE revealed that, although all EPFs are now required to teach the entirety of the national teacher training curriculum, they are managing to maintain their commitment to providing future teachers with the skills they need to function effectively in challenging environments. This is achieved thanks to an intensive program of additional input, their specific culture within which trainee teachers take responsibility for many of the practical dimensions of the EPF management, such as menu planning and routine maintenance, and their boarding system, whereby student teachers are present in the EPF during evenings and some weekends.

The Baseline report recommended that M&E of EPF activities should be more integrated in the overall M&E processes of the project. Given the low level of Portuguese language skills, including handwriting and spelling, of some trainee teachers on arrival in the EPF, remedial Portuguese support is provided in all the EPFs. Previous evaluations of the first phase of the FFE program recommended the efficacy of this remedial intervention should be consistently monitored, with the EPFs gathering data on the added value provided by the EPF remedial

³³ Based on data from 6 EPFs

Portuguese program, using a standardized diagnostic instrument in all EPFs as a pre- and post-test. The instrument had been developed by Portuguese language trainers from several of the EPFs, with support from two experts in evaluation and Portuguese language assessment respectively. This information was collected at the beginning and the end of the 2018 training year in six of the 11 EPFs: EPF Gaza, Inhambane, Nhamatanda, Tete, Nametil and Nacala. The tests were not conducted at the three EPFs operating the three-year program, where the academic year is not aligned with that of the other EPFs (the intake of new students arrives in late July). It is not clear why it was not collected in the remaining 2 colleges, although there was some suggestion that the EPFs did not consider the content of the tests to be relevant to their program. If this is the case, the EPFs collectively should devise and administer a test relevant to their needs in order to gather and use systematic monitoring information.

Table 11 below summarizes the results of the pre-and post-tests in the 6 EPFs for 2018.

EPF	Average mark out of 130 - pre-test	Average mark out of 130 - post-test	Difference between pre- and post-test	%average increase in score by EPF	% of students whose scores increased more than 10%
EPF Gaza	94.7	107.8	13.1	13.8%	60%
EPF Inhambane	79.7	83.4	3.7	4.6%	23%
EPF Nhamatanda	74.6	103.2	28.6	38.3%	96%
EPF Tete	57.9	63.2	5.3	9.2%	58%
EPF Nametil	79.75	102.3	22.6	28.3%	88%
EPF Nacala	78.3	110.9	32.6	41.6%	97%
Average	77.5	95.1	17.6	22.8%	84.4%

Table 11: Results of Portuguese language pre- and post-tests in 6 EPFs 2018

Comparison of these shows that:

- 84% of students' scores improved by more than 10% between the beginning and the end of the academic year;
- On average, the scores of trainee teachers improved by 23% between the pre- and post-tests;
- There is some variation between the EPFs, across all metrics of comparison, with Nhamatanda, Nametil and Nacala showing most improvement

These results demonstrate that the EPF teacher training is providing significant added value in Portuguese language competence, at least in the 6 colleges for which this information is available. It is striking that EPFs such as Nhamatanda, Nametil and Nacala managed to improve their scores significantly more than the three other EPFs for which data is available. Going forward, it is strongly recommended that the diagnostic tests, or, if preferred, a comparable test rigorously devised to interrogate the different areas of Portuguese language knowledge should be used each year in all EPFs to monitor improvements in Portuguese language attributable

to the teacher training program. Internal investigation should be undertaken to ascertain why some EPFs perform so much better than others in this area and devise a plan to learn from and extend this good practice across all 11 teacher education colleges.

The Baseline report presented information on the overall profiles of EPF trainers and on supply and demand for specialized trainers, which the EPFs themselves had identified as useful for their planning and operations. Updated information based on the same indicators was collected at midterm by asking the EPF directors to complete the same questionnaires. Information on the overall profiles of EPF trainers and on supply and demand for specialized trainers at midterm is compared with the same information at baseline and presented below. Due to unforeseen circumstances, this information was not made available by EPF Cabo Delgado, which was dealing with the aftermath of Hurricane Kenneth at the time of data gathering. Given that data from EPF Cabo Delgado was not available at midterm, the comparison also disregards the baseline information from that EPF.

Profiles of EPF trainers

The eleven EPFs (data from ten are presented since Cabo Delgado is not included in this presentation) have between 11 and 24 trainers, with the mean being 18; this has not changed since Baseline. The ratio of female to male trainers remains the same as at Baseline, at 23%. Given the relatively low turnover in teacher trainers and the fact that the baseline and midterm data collections took place only one academic year apart, it is not surprising that there has been little change in this area.

There has been a small improvement in the overall level of qualification of the trainers. 93% of trainers hold a *Licenciatura* as their highest qualification (compared with 90% at Baseline). 4% of trainers hold a Master's degree (compared with 3% at Baseline). Only 3% of trainers now have a lower qualification than the *Licenciatura* (compared with 8% at Baseline). Of these, four hold a *Bacharelato* and two the "Medium" level (at Baseline, there were 6 trainers with *Bacharelatos*, 8 with "Medium" level and 2 with "Basic" level), suggesting that attempts to ensure all trainers upgrade their level of qualification to at least *Licenciatura* are gradually succeeding³⁴.

42% of the trainers have a qualification in Education, Science of Education or School Management; 34% in the areas of the arts, languages or humanities and 23% in the areas of science, social sciences, mathematics or physical education; this is not significantly different from the situation at Baseline. Overall the level of

³⁴ The "Licenciatura" is the equivalent of a European or US Bachelor's Honours degree; a "Bacharelato" is approximately equivalent to an Associate Degree, following two or three years of higher education, and has been phased out in recent years in favour of a three-year "Licenciatura" in alignment with other countries; "Medium level" is the teacher training qualification earned after graduating from 12th grade of secondary school; "Basic level" is the teacher training qualification earned after graduating from 10th grade of secondary school. In principle, in order to train teachers in Mozambique, trainers are now generally required to possess at minimum the level of "Licenciatura", although this requirement has not always been consistently applied.

professional qualification of EPF trainers has improved slightly: 27% are graduates of One World University/ISSET (27% at Baseline); 44% were trained at the Pedagogical University (40% at baseline); 25% at another university, generally either *Universidade Eduardo Mondlane* or the Catholic University (25% at Baseline); the number of less qualified trainers, whose highest qualification was earned at institutions such as the state-run teacher training colleges or the EPFs has fallen to 2% from 4% at Baseline. 99% of the trainers have received psycho-pedagogical training (98% at Baseline)³⁵.

The overall levels of experience of primary teaching of the EPF trainers have increased slightly since baseline: 28% of the trainers have had two or fewer years' experience teaching at primary level (compared with 30% at Baseline); 15% have had between 2 years 1 month and 5 years (17% at Baseline); 27% between 5 years 1 month and 10 years (25% at Baseline) and 30% have had more than 10 years 1 month (28% at Baseline).

There has been more change in terms of the trainers' experience of training student teachers within an EPF setting: 11% have had two or fewer years' experience training student teachers (compared with 9% at baseline); 24% between 2 years 1 month and 5 years (30% at Baseline); 35% between 5 years 1 month and 10 years (37% at Baseline) and 30% more than 10 years 1 month (24% at Baseline).

There has been significant progress concerning the administrative categories of the EPF trainers. The percentage classified (and therefore paid) appropriately as teacher trainers (*Instructor Técnico Pedagógico - ITP*) has risen to 35% (compared with 15% at baseline). 34% are now classified in the category ITP-N1, compared with 14% at baseline; two trainers are in the category ITP-N2 (compared with one at Baseline). This is encouraging progress, on which the EPFs and the government are to be congratulated, although there is still considerable progress to be made in this respect. The most frequent category is still DN1, the highest category of "*docente*" (teacher), with 57% of teacher trainers still classified in this way (compared with 66% at baseline). The percentage of trainers classified in the lower DN categories of DN2, DN3, and DN4 is now 8%, compared with 19% at Baseline (see figure 16 below).

³⁵ As explained above, that data from EPF Cabo Delgado was not available at midterm. This comparison also excludes the baseline information from that EPF.

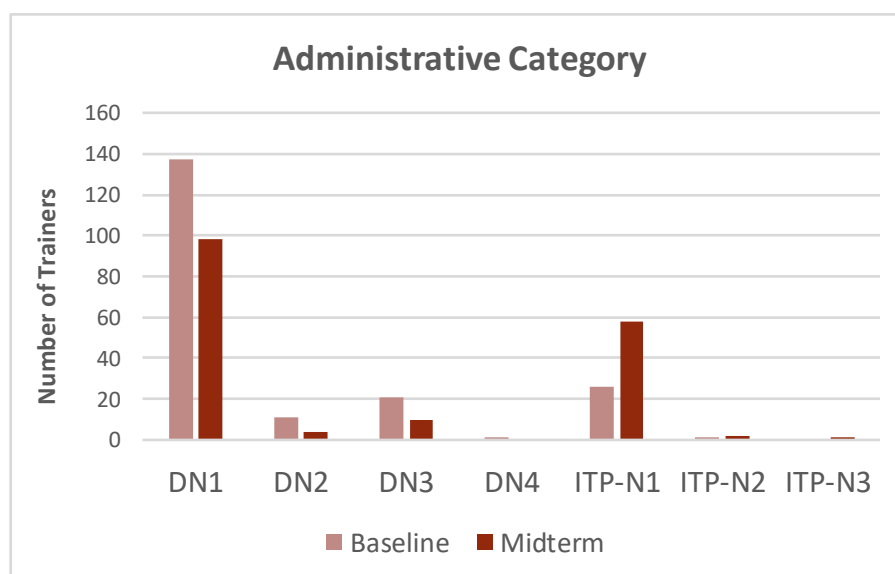


Figure 16: Administrative categories of EPF trainers

Specialized subjects: trainer supply and demand

Analysis of the overall relationship between the needs for specialized teacher trainers and the numbers of trainers employed per subject throughout the ten EPFs for which data is available reveals a situation fairly similar to that at baseline. Although the majority of positions are filled in many subjects, such as Psycho-pedagogy (16/17 posts filled, compared with 19/19 at baseline), Portuguese language teaching methods (16/17 posts filled, compared with 15/16 at baseline), Expression Techniques and Language (11/13 posts filled, compared with 13/14 at baseline), School management and Organization (12/14 posts filled, compared with 13/14 at baseline), Social Sciences teaching methods (13/14 posts filled, compared with 13/15 at baseline) and Bantu languages and bilingual education (9/10 posts filled, compared with (9/11 at baseline), there are considerable numbers of vacant posts in certain other subjects. There are a large number of unfilled posts in Research methods/ICTs (5/11 posts filled, compared with 5/12 at baseline), Visual education teaching methods (4/10 posts filled, compared with 4/11 at baseline), Building, Maintenance and School Production (4/9 posts filled, compared with 4/11 at baseline), Crafts teaching methods (2/10 posts filled; 2/11 at baseline), Music teaching methods (7/10 posts filled; 6/11 baseline) and Moral and Civic Education (5/9 posts filled; 4/11 baseline) (see figures 17 and 18 below). As pointed out in the baseline report, although the evaluation does not reveal why there should be such a gap between supply and demand in these subject areas, the situation is similar in other countries and sectors where labor market forces have led to a shortage of specialized teachers.

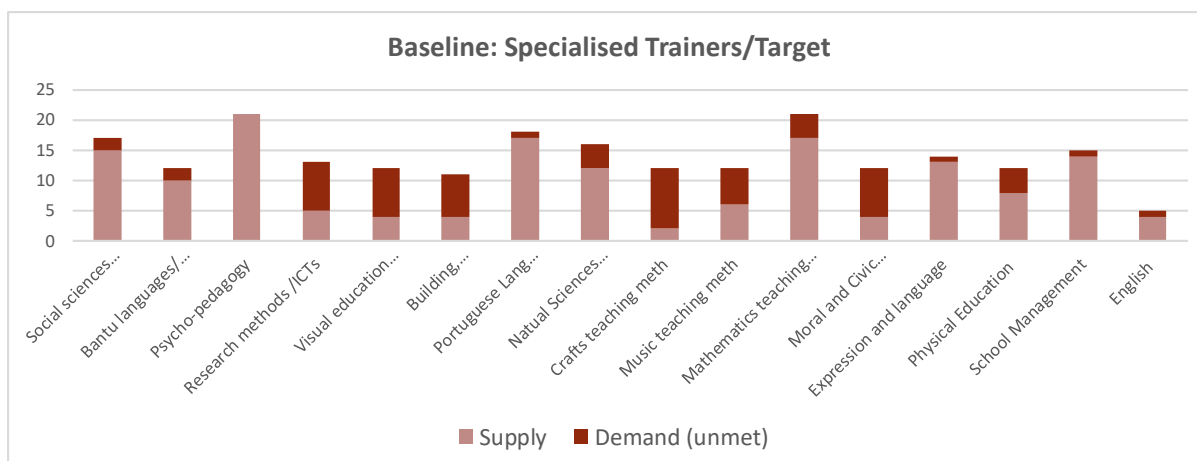


Figure 17: Specialized subjects: EPF trainer supply and demand at baseline

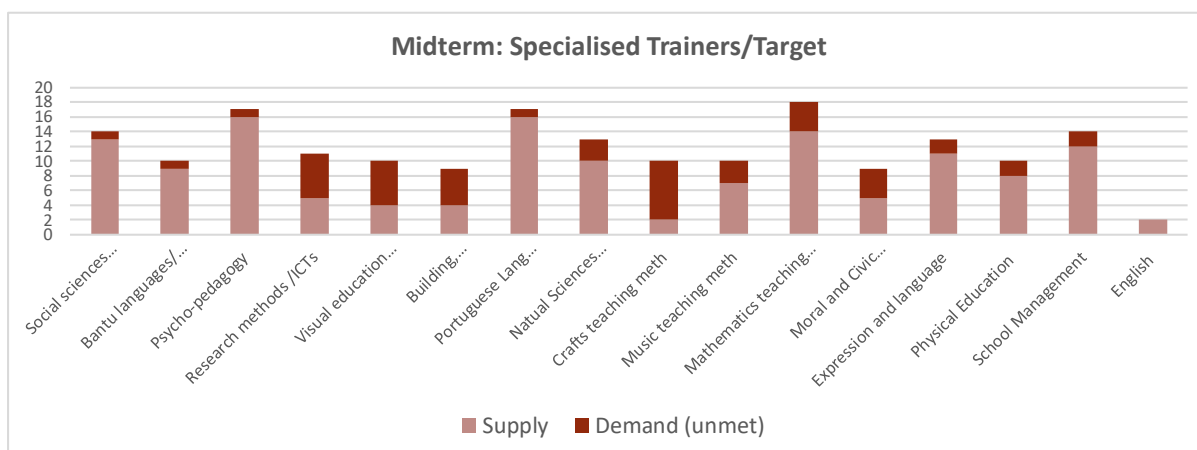


Figure 18: Specialized subjects: EPF trainer supply and demand at midterm

Finally, the percentage of the EPF trainers overall with at least five years' experience of primary teaching has increased slightly to 64%, compared with 61% at baseline. Once again, there is considerable variation between the EPFs in this regard, with EPF Maputo, Nacala, Nametil and Tete having the best ratios³⁶.

³⁶This is significant in terms of conformity with the Ministry of Education and Human Development's requirement for teacher trainers in government teacher training institutions. The requirement is based on a regulation regarding the required profile of teacher trainers, introduced in 2007 by the then Ministry of Education and Culture. Article 53 of this regulation states that the following people can apply for posts of instructor in one of the state-run teacher training colleges:

- Instructors and teachers who have a higher education qualification in the disciplinary areas required to teach the curriculum and who have a minimum experience of 5 years of basic level primary teaching;
- Professionals with a "medium" level qualification, who have undergone specific training and have the recognised professional competence and relevant professional experience of 7 years minimum.

Visits to EPFs and discussion with management, trainers and students, revealed several areas of progress since baseline were noted, along with some which should be monitored.

The EPFs are still able to provide a culture in which a “different sort of teacher” may be formed, with a focus on the more holistic aspects of education, which provides trainees with the skills, attributes and attitudes which the realities of teaching in low-resource rural schools require. However, this is made more challenging by the requirement that all EPFs follow the national curriculum for teacher education, and the fact that all but three EPFs are still obliged to offer only a one-year teacher training program: only EPF Maputo, EPF Chimoio and EPF Cabo Delgado offer the 3-year course. Students who have low levels of Portuguese language or reading or writing on arrival at the EPF are given extra support, as one EPF director explained:

There are some students who have problems in writing and even in reading.../... when they arrive, the Portuguese language trainer gives them a diagnostic test .../... the aim is not to correct it and give a mark, the aim is to evaluate their level, how does this person write, what is their handwriting like, what is their spelling like .../... to select the people who have problems and identify those who need particular support. If you noticed, our classrooms all have syllable charts so in their free time, study time, those who have handwriting problems are told to draw those letters, copy from the syllable chart. It's effective, we see changes: people who wrote rather badly when they arrived can write better by the time they graduate. They know how to form the letters well, mostly cursive letters, which are the ones they are recommended to teach.

During visits to three EPF, the external evaluator was able to observe a number of EPF students conducting teaching practice and to assess the quality of support and feedback they received. The quality of the practice lessons observed was variable, not surprisingly given that the student teachers were close to the beginning of their training, only one month into their course. On the whole, they made good use of teaching aids, including items they had evidently produced themselves. The lessons were well structured and based on a written lesson plan. Concepts and content were introduced before being practiced. There was little use of group or pair work, although one student teacher had pupils moving around the classroom and making patterns to illustrate a mathematics principle she was teaching, which was very effective and engaging for the pupils. One student teacher showed remarkable awareness of two disabled children in the class and managed to include them in the lessons, physically helping a child with reduced mobility to come to the blackboard to write an exercise and guiding the hand of a child with motor coordination challenges to hold a pencil correctly and form letters. All the student teachers had progress to make in terms of the pace of their lessons, providing students with opportunities for active practice of the concepts taught and monitoring the comprehension and participation of students in the class. However, the lessons observed were all of an acceptable level for near-novice trainee

teachers; some showed considerable promise. The quality of observation and feedback from the EPF tutors and the school-based mentors was high. Feedback was balanced between positive comments and aspects to improve and was provided in a constructive manner. Suggestions for improvement were practical and realistic. The external evaluator was impressed by the conscientious and competent support provided to the student teachers observed (of course they were aware they were themselves being observed, so this may have conditioned their behavior).

Trainee teachers in the EPFs have school-based teaching practice throughout their training, with support and feedback from their trainers and from school-based mentors. The coordination between the EPFs and practice schools appears to be good, at least in the EPFs visited.

The EPFs visited had active gardens producing a wide variety of vegetables, fruit and other produce, which is used for the meals served to students and staff, which are well-balanced and planned by students on the basis of principles learnt during nutrition education. They have also begun to raise small livestock (chickens, rabbits or ducks) for sale or consumption. Student teachers contribute to the garden work and husbandry, learning valuable skills to apply when they become teachers. The nutrition education component of the EPF curriculum includes materials on basic nutrition, balanced diet, hygiene and sanitation and the use of school gardens, in addition to a nutrition album (teaching tool), pamphlets, posters and a cookbook developed by the nutrition education team.

The infrastructure and equipment within the EPFs, at least those visited, is relatively good and has benefitted from a program of progressive upgrading over the course of FFE2.

Finally, the recent evacuation of EPF Nhamatanda following Cyclone Idai must be mentioned. This occurred in early March, during the weeks following the first and just prior to the second visit to Mozambique of the external evaluator. When the devastating cyclone struck, the students were evacuated on foot from Nhamatanda, leaving all their possessions behind. All students and staff escaped alive, though shaken. They were subsequently sent to five different EPFs who welcomed them and allowed their training to continue with only 2 weeks delay. This prompt response and the relatively small interruption in studies is a testament to the efficiency and engagement of ADPP Mozambique.

One class of the student teachers evacuated from Nhamatanda was welcomed in the EPF Gaza, where the external evaluator met them during her visit. They were still visibly affected by their experience and sat huddled in donated coats, claiming they felt cold in Gaza.

Recommendations

Due to the short duration of the one-year teacher training courses, it is vital that EPFs promote active, teacher-centered teaching methods even more vigorously. Additional transport is needed to allow trainee teachers to reach their practice schools on time and presentable to teach (many currently walk long distances, meaning they arrive dusty and less than fresh).

Systematic pre-and post-testing of students to measure the added value of the Portuguese literacy instruction should take place each year in all EPFs.

The link between nutrition education and food production and meal planning and preparation in the EPFs could be made even stronger by explicitly linking theory and practice, using nutrition education to reinforce the practical food-based activities and vice versa. Introducing some assessment of the application of nutrition education in menu and food preparation activities might reinforce the status of nutrition education within the EPFs.

Some needs for infrastructure and equipment upgrading remain. The accommodation for teaching staff at EPF Gaza is in extremely poor condition and, in some cases, unsafe for human habitation. Additional computers are needed for use by both students and trainers.

Trainer trainers should have access to continuous professional development in order to stay abreast of developments in their specialist subjects, in pedagogy and in assessment. Trainer trainers should have access to more resources and additional computers in the staff room.

Students evacuated following Cyclone Idai should continue to receive appropriate support and possibly counselling in some cases.

Skills and Knowledge of School Administrators (1.1.5)

Indicator 3 (Outcome): Number of school administrators and officials in target schools who demonstrate use of new techniques or tools as a result of USDA assistance. Final target: 210

Midterm target: 150

Midterm results (March 2019): 165

Midterm target 110% achieved

Indicator 4 (Output): Number of school administrators and officials trained or certified as a result of USDA assistance. Final target: 264

Midterm results (March 2019): 426

Final target 161% achieved

Analysis of project records reveals that the target for training school administrators and officials in management and administration of the school feeding program in target schools has been surpassed as of midterm. School directors reported this training to have been useful. During school visits, the external evaluators were impressed with the level of knowledge and commitment to the project demonstrated by school leaders as they showed the visitors the installations and activities put in place by the project.

It was not possible to formally collect data demonstrating that school leaders demonstrating use of the new techniques or tools acquired. However, two items of the school survey, which was conducted in 170 project schools are considered as a proxy for the purposes of the midterm evaluation: enumerators were asked to observe the state of cleanliness of the storeroom and of the kitchen; In 106 of the 170 schools visited (62%), the storeroom was considered by the observer to be “perfectly clean”; 51 (30%) were judged “more or less clean” and 13 (8%) “not clean”. In 104 schools surveyed (61%), the kitchen was considered to be “perfectly clean”, 58 kitchens (34%) were judged “more or less clean”, with 8 (5%) judged “not clean”. Where this infrastructure provided by the project and used continuously to support it is well maintained, this is considered as evidence that school leaders are performing their leadership role and applying the training they have received. Extrapolating from the figure of 61% of the schools sampled, it is estimated that the leaders of 61% of the total of 270 schools (165 schools) are likely to be demonstrating these characteristics, making an estimate of 165 school administrators and officials who are expected to be applying the training they have received.

Overall, the detailed responses of school directors to the school survey reveal a high degree of knowledge of and commitment to the project.

Attentiveness (1.2)

(Outcome): Percentage of teachers who report increased attentiveness of students in the classroom. (Special study indicator). Final target: N/A
Midterm target: N/A
Midterm results (March 2019): 11%

(Outcome): Percentage of students in project schools who report increased attentiveness in the classroom. (Special study indicator). Final target: N/A
Midterm target: N/A
Midterm results (March 2019): 6%

Attentiveness in the classroom is an important link in the causal chain which the project ToC seeks to address: short-term hunger leads to lack of attentiveness in the classroom leading to problems in learning. The ToC postulates that reducing short-term hunger in the classroom should lead to fewer problems of attentiveness,

leading in term to better learning. Many interview informants mentioned the improvement in students' attentiveness since school feeding began. One school director explained:

They cover large distances to come to school, some of them leave home without eating anything but when they arrive, they have the meal and they manage to be calm and attentive in their lessons.

A SDEJTs school feeding focal point was more explicit:

Because if I'm in class and I have something in my stomach, my attention is better than if my stomach's rumbling and I'm physically present but not attentive.

Previous evaluations showed the project school feeding interventions had reduced short-term hunger and problems with attentiveness; however, in the absence of improved teaching, these had not resulted in improved learning. At midterm, in order to offer some triangulation of these information obtained, both teachers and learners themselves were asked about their experience of students feeling hunger in the classroom³⁷.

When asked whether they had problems paying attention in the classroom, 42% of the pupils in intervention schools (FFE + literacy and FFE only combined) said they did, compared with 48% in comparison schools. Both groups had reported slightly lower rates at baseline (38% for pupils in intervention schools, 36% for pupils in comparison schools). Although pupils in intervention schools are less likely to report problems of attentiveness in the classroom, differences in change between baseline and midterm for the two groups are not statistically significant (see figure 19 below and Technical Appendix 1, pp 197 - 198).

³⁷ Students are the primary beneficiaries of many of the project's interventions and their voices and experiences are essential to the evaluation. In order to capture multiple experiences and mitigate the risk of bias or recall errors (or adults providing "official" responses they perceive to be correct or desired), wherever possible the same questions are asked of students and teacher.

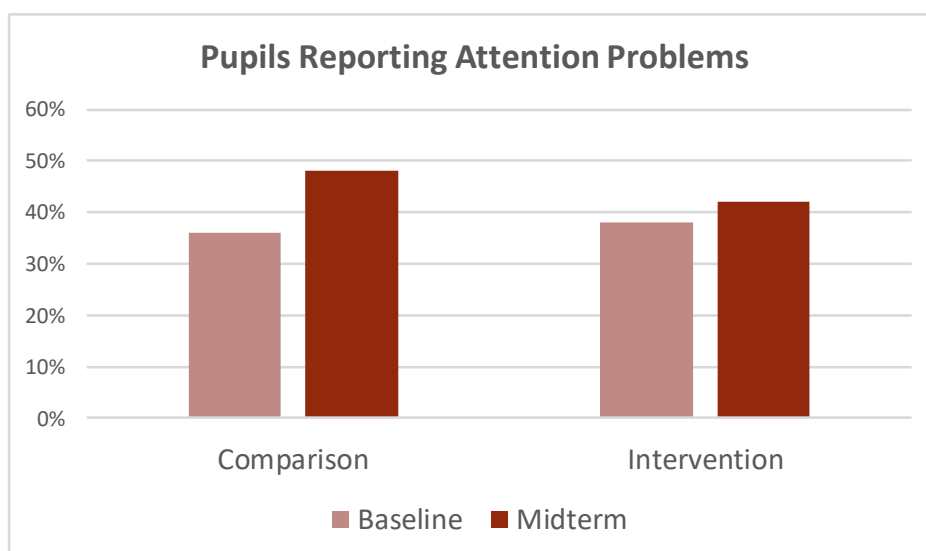


Figure 19: Pupils’ reporting of attention problems

Teachers were also asked about their experience of students’ attentiveness in the classroom. At midterm, 79% of teachers surveyed (in FFE + literacy and FFE only schools combined) reported their students “often” or “sometimes” having attention problems in the classroom (compared with 88% at baseline) and 22% that they “never” did (compared with 11% at baseline) (see figure 20 below and Technical Appendix 2, pp 52-54)³⁸.

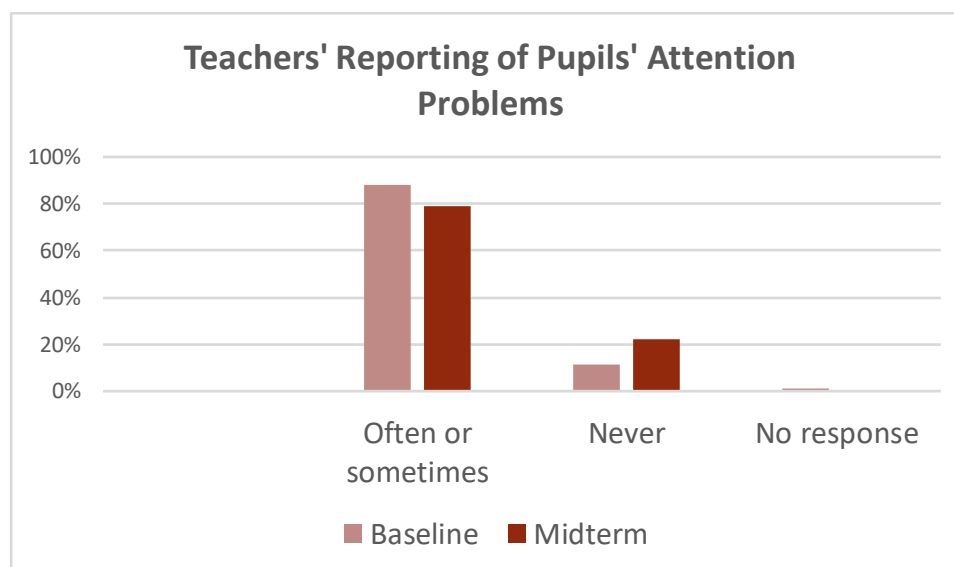


Figure 20: Teachers’ reporting of pupils’ attention problems

³⁸ The “often” and “sometimes” responses were combined, due to their imprecision, compared with the more categorical “never”.

Teachers were asked whether their pupils had difficulties learning. At midterm, 84% of teachers surveyed (in FFE + literacy and FFE only schools combined) reported their students “often” or “frequently” having attention problems in the classroom (compared with 82% at baseline); 15% reported students “never” had problems learning (16% at baseline) (see figure 21 below and Technical Appendix 2, pp 54-56).

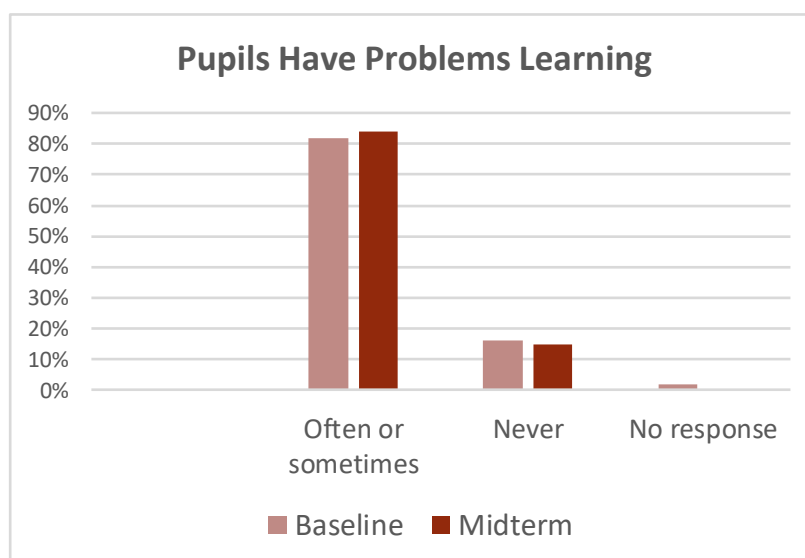


Figure 21: Teachers’ reporting of pupils’ learning problems

On the basis of teachers’ responses to these questions, it appears that improvements in students’ attention in the classroom are not yet being translated into improved learning.

Short-term hunger (1.2.1)

(Outcome): Percentage of teachers who report reduced short-term hunger of students in the classroom. Final target:N/A
Midterm target:N/A
Midterm results (March 2019): 10% increase in “never” since baseline

(Outcome): Percentage of students in project schools who report reduced short-term hunger in the classroom. Final target: N/A
Midterm target:N/A
Midterm results (March 2019): 18% decrease since baseline

When asked whether they sometimes felt hungry in the classroom, at midterm 42% of pupils in intervention schools (FFE + literacy and FFE only combined) said they did, compared with 58% of students in comparison schools. These levels of hunger are lower than the baseline levels for both groups: 60% for pupils in intervention schools, and 68% of students in comparison schools. Although considerably fewer students in project schools

said they were sometimes hungry than those in comparison schools and considerably fewer students in intervention schools said they were sometimes hungry at midterm than at baseline, the differences in change between baseline and midterm for the two groups are not statistically significant (see figure 22 below and Technical Appendix 1, pp 195 - 196).

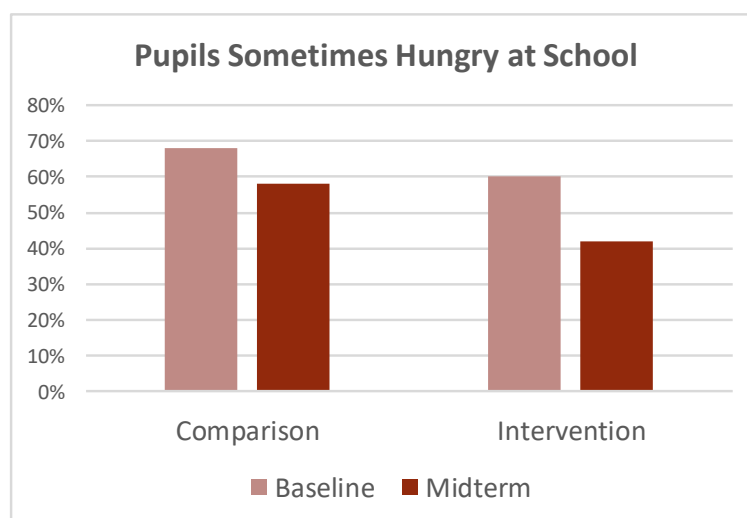


Figure 22: Pupils' reporting "sometimes" hungry at school

Students were also asked whether they felt hungry "now" i.e. at the time of the interview³⁹. At midterm 23% of pupils in intervention schools (FFE + literacy and FFE only combined) said they did, compared with 33% of students in comparison schools. Again, these levels of hunger are lower than the baseline levels for both groups: 42% for pupils in intervention schools, and 66% of students in comparison schools. Once again, although considerably fewer students in project schools said they were hungry "now" than those in comparison schools and considerably fewer students in intervention schools said they were hungry "now" at midterm than they had at baseline, the differences in change between baseline and midterm for the two groups are not statistically significant (see figure 23 below and Technical Appendix 1, pp 196 - 197).

To increase validity and allow triangulation of responses, pupils were asked two separate but related questions about their experience of hunger at school. It should be borne in mind that, due to fieldwork constraints, pupils were asked whether they felt hungry "now" at varying times of day; in some cases after they had already eaten the soya porridge, in others before. This constraint also applied during the baseline data gathering. Interview times were recorded in all cases.

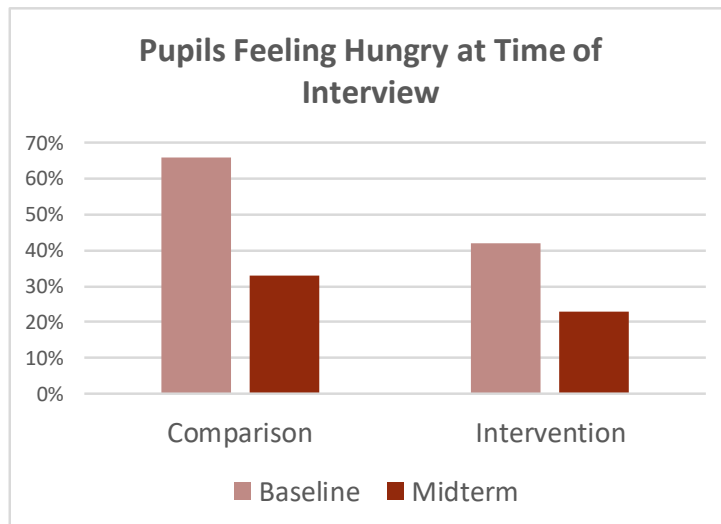


Figure 23: Pupils’ reporting hungry at time of interview

When asked about their perceptions of student’s hunger in the classroom, at midterm 67% of teachers stated that their pupils “often” or “sometimes” appeared to be hungry in class (76% at baseline) and 33% that they “never” appeared hungry (23% at baseline) (see figure 24 below and Technical Appendix 2, pp 50-52).

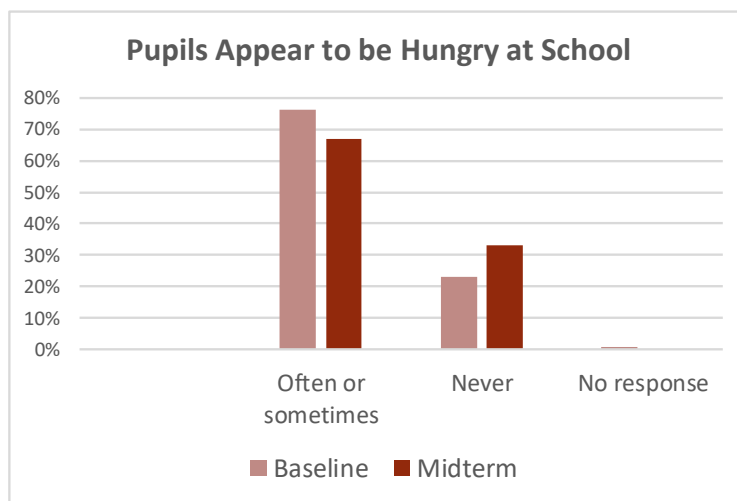


Figure 24: Teachers’ reporting pupils appear hungry at school

Access to food (school feeding) (1.2.1.1)

Indicator 15 (Output): Number of daily school meals (breakfast, snack, lunch) provided to school-age children as a result of USDA assistance. Final target: 48,100,000
Midterm target: 27,056,250
Midterm results (March 2019): 28,162,260
Midterm target 104% achieved

Indicator 16 (Output): Number of school-age children receiving daily school meals (breakfast, snack, lunch) as a result of USDA assistance. Final target: 87,000; (000 new: 13,000; continuing: 74,000; female: 42,630; male: 44,370)

Midterm target: 74,000 (new: 2,000; continuing: 72,000; female: 36,260; male: 37,740)

Midterm results (March 2019): 84,026 (new: 787; continuing: 83,239; female: 42,141; male: 41,885)

Target 114% achieved (new: 39%; continuing: 116%; female: 116%; male: 111%)

Indicator 17 (Output): Number of social assistance beneficiaries participating in productive safety nets as a result of USDA assistance. Final target: 74,000 87,000; continuing: 69,000; new: 13,000; female: 42,630; male: 44,370)

Midterm target: 74,000(continuing: 72,000; new: 2,000; female: 36,260; male: 37,740)

Midterm results (March 2019): 84,026 (new: 787; continuing: 83,239; female: 42,141; male: 41,885)

Final target 114% achieved (new: 39%; continuing: 116%; female: 116% ; male: 111%)

Analysis of project records shows the distribution of school meals to be on target. In order to triangulate this information, the midterm evaluation surveys included questions for students, teachers and school leaders about the distribution of school meals.

64% of schools surveyed reported that school feeding had been missed one or more times during the past two months. This is a significant increase since baseline, when 34% schools stated that in the past two months there had been one day or more when they had not been able to prepare the lunch. At midterm, the main reasons cited for not being able to feed children were running out of CSB+ (cited by 21% of schools who had had gaps in feeding, compared with 45% at baseline); cooks not turning up (cited by 52% of schools who had gaps in feeding, compared with 47% at baseline); lack of water (cited by 16% of schools who had gaps in feeding, compared with 23% at baseline); lack of firewood (cited by 10% of schools who had gaps in feeding, compared with 11% at baseline).

25 of the 170 schools surveyed (15%) reported running out of CSB+ during the past two months. 7 of these schools reported remaining without CSB+ for 1-3 days; 6 for 4 – 6 days; 4 for 7 – 9 days; 3 for 10-12 days and 3 for 13 – 15 days. One school reported being without CSB+ for 20 days and one for 21 days. Project representatives reported that delays in a shipment of CSB from the United States early in April accounted for the number of schools which ran out of the commodity at this time. Although it is encouraging that 88% of schools did not run out of CSB+, for those schools unable to feed their students due to interruptions in the stock, the impact on their students will be considerable. Not to receive food when they have become used to doing so is disruptive and distressing. It may also have long-term negative impact on the advances made in terms of attendance and retention, if families lose trust in school feeding and, potentially, in the school in general for not being able to provide a promised service. This risk was mentioned several times during evaluation interviews.

Analysis of project records in the commodity tracking files for March and April 2019 (see tables 12 and 13 below), the same period as the school survey was conducted, reveal:

- The principle overall reason for non-feeding is pupil absence;
- Of reasons directly attributable to the project itself, the main reasons for schools not feeding are lack of CSB+ and lack of cooks;
- In March 2019, 126 schools (46%) failed to feed at least once;
- In March 2019, 435 feeding days (8%) were lost of a total of 5,420 potential days (271 x 20);
- In April 2019, 179 schools (68%) failed to feed at least once;
- In April 2019, 989 feeding days (19%) were lost of a total of 5,149 potential days (271 x 19)⁴⁰;
- Infested CSB+ is still a problem in a small number of schools;
- No schools fail to feed because of lack of utensils;
- Only 4% of schools failed to feed due to lack of water, suggesting many schools without safe drinking water are managing to find a source of water to prepare the soya porridge (interview respondents explained that pupils bring water from home or river water is used).

	Magude (66)		Moamba (65)		Manhica (85)		Matutuine (55)		Total	
	Schools	Days lost	Schools	Days lost	Schools	Days lost	Schools	Days lost	Schools	Days lost
No CSB+	0	0	15	83	20	77	0	0	35	160
Infested CSB+	3	7	0	0	1	7	0	0	4	14
No water	1	10	7	29	1	9	2	3	11	51
No firewood	1	1	1	1	3	15	0	0	5	17
No cooks	9	31	13	38	5	21	10	23	37	113
Other	2	2	2	2	1	1	5	10	10	15
No reason given	9	24	3	9	7	18	5	14	24	65
Total	25	75	41	162	38	148	22	50	126	435

Table 12: Schools not feeding for specific reasons and feeding days lost March 2019

	Magude (66)		Moamba (65)		Manhica (85)		Matutuine (55)		Total	
	Schools	Days lost	Schools	Days lost	Schools	Days lost	Schools	Days lost	Schools	Days lost
No CSB+	18	63	26	160	59	466	9	61	112	750
Infested CSB+	0	0	0	0	0	0	0	0	0	0
No water	1	1	4	13	3	24	1	1	9	39
No firewood	0	0	0	0	2	5	0	0	2	5
No cooks	4	6	7	13	7	16	1	6	19	41
Other	2	20	1	2	3	3	2	10	8	35
No reason given	12	54	0	0	7	14	10	51	29	119
Total	37	144	38	188	81	528	23	129	179	989

Table 13: Schools not feeding for specific reasons and feeding days lost - April 2019

⁴⁰ The relatively high number of feeding days lost due to schools running out of CSB in April is attributed to a delayed shipment of CSB+, rather than to failures in the distribution of the commodity to schools.

It is clear that the two priorities in terms of improving the rate of school feeding are: i) ensuring there is CSB+ in all schools at all times and ii) ensuring there are sufficient volunteer cooks available each day.

Teachers rated the distribution of food in school to be “excellent” (27%) or “good” (67%). However, 66% of teachers surveyed stated the distribution could be improved, citing the following aspects which require improvement:

- Quality of CSB+ delivered to schools;
- Timely delivery of CSB+, so this does not run out;
- Absence / lateness of volunteer cooks;
- Lack of incentives for volunteer cooks and storeroom managers;
- Insufficient cooking utensils, bowls and spoons;
- Better organization of time so distribution of food does not excessively interrupt teaching time (for example, better planning, so as to reduce waiting time and feed pupils after/ before class or during the long break);
- Lack of water or firewood.

This list is very similar to the aspects revealed in the school survey and to those cited during previous evaluations. It is also rather concerning that there continue to be so many days without school feeding, especially where this is due to schools running out of CSB+. The final evaluation of FFE 1 had found there had been improvements to the distribution of CSB+, due to the use of a commodity tracking system (CTS) and improved processes and systems.

Students who participated in focus groups explained that the soya porridge was often too hot when distributed, so children got burned. They also claimed to only enjoy the food when they were able to add sugar, brought from home.

There has been a relatively high turnover within the project Logistics team. At the time of the midterm data collection, a new Logistics manager had recently been appointed and he appeared to have a clear vision of where changes were needed and how to implement these.

In addition to several positive aspects of the logistical organization of the project observed, some others require monitoring:

There have been improvements in unloading and Customs clearance processes and sanitary controls for CSB+ arriving from USA, so these procedures are now achieved more rapidly.

Information flow appears to have improved since previous evaluations, including greater cooperation from most storeroom managers who send information on weekly stock levels by SMS to their “professional”, who transmits

this information to the logistics manager so CSB+ delivery can be planned. There are still some flaws in the information flow between schools and the logistics manager. Some storeroom managers lack commitment or are absent during deliveries. The storeroom managers are teachers and often do not perceive the management of the storeroom as part of their responsibilities as teachers. They have not received any form of incentive for managing the storerooms since the inception of the project.

During fieldwork, the external evaluator was told several times of CSB+ in school storerooms being found to be infested. The rate of infestation of CSB+ is low overall; where it occurs, the commodity is sifted to remove weevils then repackaged.

Recommendations
Cooks should be made aware of the need to let the porridge cool down before serving, to avoid burning the students.
Adding additional sugar is not recommended, as the nutritional content of the CSB+ is already well balanced to meet the nutritional needs of growing children.
The project's internal organization of these processes is still in need of some improvement, as are other logistical processes.
Consider offering storeroom managers an expression of thanks and appreciation in the form of a token gift of phone airtime in recognition of their important role in the project in order to improve morale and commitment.
Systems must be improved so all infestation is eliminated.

Height, weight and BMI z-scores

In order to see whether school feeding impacts on pupils' physical development, two anthropometric measures: height and weight will be recorded at three time points, so that BMI z-scores can be calculated, according to WHO guidelines. At midterm, BMI z-scores for the two groups receiving school feeding (FFE + literacy and FFE only) were compared with those from the comparison group. The FFE + literacy and FFE only group had a mean score of -0.514 (SD 1.24), compared with -0.437 (SD 1.72) at baseline. The comparison group had a mean score of -0.675 (SD 1.92) compared with -0.416 (SD 2.15) at baseline. The difference in rate of change from baseline to midterm between the two groups was therefore not significant. This result suggests that those children who received school feeding did not grow significantly more than those who did not.

At baseline, calculating BMI Z-score scores proved problematic in many cases, since approximately one third of pupils did not have accurate dates of birth but a year of birth and 1 January as a default date⁴¹. It had been hoped that at midterm accurate dates of birth could be collected in order to rectify this. Enumerators were trained to probe several avenues to obtain correct dates of birth for students and the ODK data collection interface required them to indicate the source of the date of birth information. This effort was successful in yielding more dates of birth, although it was still not possible to obtain dates of birth for all students. Of the 2,882 students for whom data was available at both time points, 1,935 did not have consistent dates of birth at both time points. When midterm dates of birth were available, these were considered to be more reliable and were therefore used for both the mid-term and baseline data, to allow appropriate comparisons.

In order to probe the FFE2 midterm results further, the height and weight gain of participating students were also analyzed⁴². The rate of change from baseline to midterm was almost identical between the two groups. At midterm, the mean height of the FFE + literacy and FFE only group was 126cm (SD 7.83), compared with 114cm (SD 6.62) at baseline. The mean height of the comparison group was also 126cm (SD 8.75) compared with 114cm (SD 6.29) at baseline. The rate of increase in weight from baseline to midterm was also very similar between the intervention and comparison groups. At midterm, the mean weight of the FFE + literacy and FFE only group was 24.5 kg (SD 4.52), compared with 19.8 at baseline (SD 3.39). The mean weight of the comparison group was 24.2 kg (SD 5.10), compared with 20.0 at baseline (SD 3.62). No significant difference in either height or weight was found between the two groups.

The complete results are presented in Technical Appendix 1, pp 188 - 194.

Informants consistently commented on the physical improvements observed in students' appearance since the onset of schools feeding: students looking "healthier" and noting improvements in the condition of their skin and hair, in addition to increased energy and dynamism.

⁴¹ There are several possible explanations for this: early grades pupils may not know their date of birth; school records may be incomplete or inaccurate; the enumerators may lack the experience and expertise to be able to elicit this information.

⁴² The final evaluation of FFE1 had found that students receiving school feeding grew on average 2 cm more than students in the comparison group, although increases in their BMI Z-scores were not significantly different.

Student attendance (1.3)

Indicator 1 (Outcome): Number of students regularly (80%) attending USDA supported classrooms/schools. Final target: 59,200 (female: 29,008; male: 30,192)⁴³
Midterm target: 59,200 (female: 29,008; male: 30,192)
Midterm results (March 2019): 84,026 (female: 42,141; male: 41,885) [if 80% used: 67,220 (female: 33,712; male: 33,508)]
Final target 142% achieved (female: 145%; male: 139%) [if 80% used: 114% achieved (female: 116%; male: 111%)]

During school visits for the midterm evaluation, as during other evaluations of the project, teachers, school directors, parents, government officials and other stakeholders consistently claimed that school feeding and the other project interventions had led to increases in student attendance. One SDEJT focal point reported:

As soon as the meals were introduced, we saw fewer children missing school – before they missed a lot – and the pedagogical results improved.

A parent explained this slightly differently:

The children don't miss school anymore because as soon as they remember that they'll get food there they're motivated .../... they get there early to be sure of receiving the food, even though they know it will be served later on.

Analysis of project monitoring data reveal that targets for school attendance have been surpassed at midterm. There are however two caveats to be born in mind: firstly, the targets for 80% attendance are calculated by the project as 80% of the numbers of students receiving food. However, they ought to reflect the numbers of students who attend school at least 80% of the time (see footnote below). In an attempt to partially address this issue, calculations based on 80% of the project numbers are included in square brackets in the box above; however, without actual daily attendance data it is not possible to accurately calculate the number of students who attend 80% of the time).

Secondly, the attendance numbers reported by the project are actually based on school feeding data. Since the inception of the project, there have been challenges for the project to obtain consistent, reliable pupil-level attendance data from schools, due to lack of rigor in recording pupil attendance and teachers' reluctance to complete an additional daily attendance register for the purposes of monitoring school feeding, in addition to

⁴³The targets for this indicator as set out in the contractual Attachment E in Amendment #1 are based on 80% of the target for students benefiting from school feeding; however, results are reported based on 100% of the results of students benefiting from school feeding, which makes the target appear to have been surpassed by more than is actually the case. It is recommended that the targets and measurement of this indicator should be amended from 80% of students attending school to the numbers of student who achieve 80% school attendance, to allow the project to report on this indicator as intended.

the school register. After considerable effort, towards the end of FFE1, the project leadership managed to obtain authorization from the DPEDH to have specially designed school registers produced which included carbon paper so the register could be filled in once by the teachers producing two copies: one for the school, the other for the project. The midterm evaluation visits revealed that, although this has led in some cases to attendance information in paper form being successfully collected for the project, this information is not currently being collated electronically, due to other priorities for M&E staff and to the fact that the information received from schools is incomplete since not all schools are filling in the registers systematically. For this reason, the numbers used by the project to record pupil attendance are the same as those used by the project tracking system to track and record the number of students receiving school feeding. These are based on school level enrolment data provided at the beginning of the school year by the district education offices (SDEJT), which does not always reflect mid-trimester changes in pupil numbers. These are not therefore an accurate record of attendance rates and do not allow verification of the claim made by school directors and education officials that school attendance rates have improved since the introduction of school feeding.

Number of after-school learning clubs active in project schools. Final target: 600 Midterm target: 600 Midterm results (March 2019): 675 Final target 113% achieved

Number of school children participating in clubs. Final target: 11,300 Midterm target: 11,300 Midterm results (March 2019): 12,300 Final target 109% achieved

One of the project activities intended to promote increase school attendance is support to extra-curricular learning clubs. Of the schools surveyed at midterm, 89% reported having one or more learning club (compared with 98% at baseline). Of those schools which report having clubs, 98% report having reading clubs (97% at baseline) schools. 87% report having mathematics clubs (84% at baseline). 28% reported having clubs for “recreational arts and dance” (50% at baseline). 32% report having school garden clubs (26% at baseline). 34% report having other types of club (7% at baseline)⁴⁴.

94% schools report seeing “some differences” in the children who participate in the clubs (94% at baseline). 95% report that they are more active in regular lessons (93% at baseline). 100% schools report pupils have improved

⁴⁴ Project staff reported that between baseline and midterm, the method for counting clubs changed: only literacy numeracy, gardens, science and sanitation clubs were counted; sports, recreational arts and dance were no longer counted for reporting purposes, although these continue to be supported by the project. This change may have impacted on the responses given by headteachers to the school survey.

learning outcomes (93% at baseline). 95% report improvements in the pupils' communication and expression (87% at baseline). For more information, see Technical Appendix 3 (school survey), pp 69 – 89⁴⁵.

When asked about the challenges encountered in implementing clubs, the rate of challenges reported appears to have substantially diminished since baseline. At midterm, 44% schools report low levels of pupil participation (73% at baseline), 55% report difficulties finding time for club activities (69% at baseline), 19% report a lack of interest amongst teachers in implementing clubs (55% at baseline), 12% report lacking information about methodologies or activities which can be used in the clubs (57% at baseline), 30% report lacking the materials needed (48% at baseline). For more information, see Technical Appendix 3, pp 247 – 259.

Teachers were also asked about extra-curricular clubs. 83% of teachers reported that clubs had been created in their schools (92% at baseline). 95% of teachers reported knowing what activities are offered by the clubs (85% at baseline). 95% reported that the clubs had a positive effect on pupils' schoolwork (88% at baseline) (see Technical Appendix 2, pp 22 – 28)⁴⁶.

During the first phase of the project, in the absence of substantive support and training of teachers, the project's support to school clubs and provision of kits of educational materials was the main activity designed to improve teaching and learning. The baseline report of FFE2 pointed out that, now the project includes an intensive literacy intervention, at least in some schools, "the role of school clubs must be clarified, so clubs do not simply repeat what happens in lessons. The project needs to rethink its support to extra-curricular clubs, to allow them to make more of an impact on learning. Club leaders should receive clear guidance and training in using play-based and other alternative activities to promote learning and the value and potential of clubs should be communicated to government, school leaders and teachers". The midterm evaluation found some evidence of progress in this direction: in all project schools, pedagogical directors and teachers responsible for clubs have been trained in the facilitation of extracurricular clubs, play-based and other activities to promote learning and use of the materials provided by the project. Within the recent specialization of the project "professionals", several of these now specialize in providing support to clubs and can be called upon by schools who require more help in that area.

Unfortunately, at midterm information on club attendance was not collected from students, due to a technical error, which led to the relevant questions being excluded from the instrument during fieldwork. The baseline report noted that "The reporting of clubs' existence by schools does not necessarily mean that these are particularly active or dynamic or that pupils' attendance is regular". There is some indication from the qualitative

⁴⁵ Technical Appendix 3 (school survey) is available in electronic form: pdf file entitled 04_ADPE_TABLES_SCHOOL_PDF__c

⁴⁶ Technical Appendix 2 (teacher survey) is available in electronic form: pdf file entitled 05_ADPE_TABLES_TEACHERS_PDF__c

information gathered at midterm that this situation has improved since baseline. The project should continue to monitor the quality and activities of extra-curricular clubs and ensure that pupils are participating actively, that the activities proposed are useful, practical and well-facilitated, so that the clubs are able to actually make a substantive contribution to student learning and outcomes.

School directors reported positive innovations, such as learning clubs taking place off the school premises:

Last year we had clubs functioning in the neighborhoods, because of the distance separating the children from the school - some children live three kilometers away ... so we took advantage of the children in the neighborhoods who could already read ... and were able to teach the others. We made groups of neighbors and those older ones, we gave them materials and a monitor responsible for literacy and numeracy went to the community.

A number of positive developments were noted at midterm, along with some aspects to be monitored:

Overall, the club sector appears to be more dynamic. Some basic rules have been applied, such as the posting of attendance times for clubs.

Links with other components, such as literacy, school gardens and nutrition are stronger at midterm, in part due to the role of the “professionals” who have specialized in these areas and act as specialist focal points.

The principles and mission of clubs appears to be better understood than during previous evaluations. There is also more flexibility in their organization, so clubs can take place outside of the school premises where necessary or be facilitated by older pupils if appropriate.

Recommendations
The system for gathering evidence of club activities and criteria for measuring these needs to be strengthened: a list of participants may prove clubs exist but is not proof that they provide useful activities.
The clubs coordinator does not have a dedicated team and would benefit from more support.
Principles of active, learner-centered learning and practical, active activities need to be continually promoted, rather than traditional teacher-based methods and reciting lessons by heart.

School infrastructure (1.3.3)

Indicator 7 (Output): Number of kitchens, storerooms and firewood saving stoves maintained as a result of USDA assistance. Final target: 792 (kitchens: 264; storerooms: 264; firewood saving stoves: 264)

Midterm target: 792 (kitchens: 264; storerooms: 264; firewood saving stoves: 264)

Midterm results (March 2019): 1,002 maintenance activities (kitchens: 271; storerooms: 271; firewood saving stoves: 460)⁴⁷

Final target 127% achieved (kitchens: 103%; storerooms: 103%; firewood saving stoves: 174%)

Analysis of project records reveals that the targets for maintaining the school infrastructure provided by the project (in terms of the number of interventions to maintain kitchens, storerooms and firewood-saving stoves) have been surpassed.

The midterm school survey included inspection visits of the infrastructure to verify its state of maintenance and cleanliness. In 62% of schools visited, the storeroom was found to be “perfectly clean”; in 30% of schools visited the storeroom was found to be “more or less clean”; in 8% “not clean”. 77% of schools reported that the storeroom was cleaned “daily” (79% at baseline), 16% “weekly” (18% at baseline), 4% “twice a week” (1% at baseline), 1% “from time to time” (1% at baseline) and 2% gave another response.

In 61% of schools visited the kitchen was found to be “perfectly clean”; in 34% of schools visited the kitchen was found to be “more or less clean”; in 5% “not clean”. 88% of schools reported that the kitchen was cleaned “daily” (98% at baseline), 9% “weekly” (5% at baseline), 1% “twice a week”, 1% “from time to time” and 1% gave another response.

27% of schools report having a maintenance plan (39% at baseline). Of these, 87% say their maintenance plan is followed (see Technical Appendix 3, pp. 7 – 19). In schools where this is the case, school council members contribute to maintenance in 54% of schools, teachers and school leaders in 30%, project staff linked to the school in 57%, the director in 35%, the pedagogical director in 15% and other people in 9%.

In 85% schools, volunteer cooks are amongst those who clean the kitchens (96% at baseline); in 42% schools pupils contribute to cleaning the kitchen (3% at baseline), the teacher who is storeroom supervisor contributes in 24% schools (1% at baseline) and a member of the school council in one school. The storeroom is cleaned by volunteer cooks in 208 (77%) schools, by pupils in 39 (14%) schools, by the teacher who is the storeroom

⁴⁷ These numbers refer to maintenance interventions performed, not to the number of installations maintained (the project has a total of 270 kitchens, 268 storerooms and 271 firewood saving stoves).

supervisor in 17 (6%) schools and school council members in 4% schools (1% at baseline) (see Technical Appendix 3, pp. 97 – 131).

Recommendation

In the interests of sustainability, schools should be encouraged to put in place maintenance plans and cleaning plans and to ensure these are followed.

Enrolment (1.3.4)

Indicator 8 (Outcome): Number of students enrolled in school receiving USDA assistance. Final target: 74,000 (female: 36,260; male: 37,740)

Midterm results (March 2019): 88,100 (female: 43,180; male: 44,920)

Final target 119% achieved (female: 119%; male: 119%)

Enrolment in project schools is one of the steps considered in the project ToC as a prerequisite to attendance and therefore to improved learning. Analysis of project records reveals that the targets for enrolment of students have been surpassed. Interview and focus group informants overwhelmingly claimed that enrolment had increased, and dropouts decreased as a result of the projects' activities, particularly school feeding.

During previous evaluations, analysis of the monitoring data gathered by the DPEDH in respect of enrolments and dropouts had provided some degree of support to these claims. It had been hoped to extend this analysis for the purposes of the midterm evaluation of FFE2. Unfortunately, the DPEDH was not able to provide information of sufficient quality to allow this analysis⁴⁸.

Capacity of government institutions (1.4.1)

This indicator appears with different MGD numbers under both SO1 and SO2. For discussion of capacity of government institutions, please refer to the section Capacity of government institutions (2.7.1) below.

⁴⁸ The DPEDH did make the information requested available to the evaluation team. Unfortunately, due to inconsistencies in the way it had been compiled, this could not be used. The evaluators requested that the data be checked and corrected on several occasions. However, due to a shortage of DPEDH personnel, this was not possible.

Government support (1.4.3)

Indicator 11 (Outcome): Value of new public and private sector investments leveraged as a result of USDA assistance. Final target: 80,000

Midterm target: 40,000

Midterm results (March 2019): 0

Target 0% achieved

To date, the project has not demonstrated the leveraging of new public and private sector investments. Discussions are ongoing within the leadership and M&E teams and through the bi-annual reporting process as to which current or future project activities may be counted towards this indicator.

Engagement of Local Organizations and Community output (1.4.4)

Indicator 9 (Output): Number of Parent-Teacher Associations (PTAs) or similar “school” governance structures supported as a result of USDA assistance. Final target: 264

Midterm target: 264

Midterm results (March 2019): 271

Final target 103% achieved

Indicator 10 (Output): Number of public-private partnerships formed as a result of USDA assistance. Final target: 2

Midterm target: 2

Midterm results (March 2019): 0

Target 0% achieved

At the outset of the project, School Feeding Committees (SFC) were created in project schools. In many schools, these were extensions of the existing School Councils, which have an important role in all primary schools in Mozambique and wield considerable influence (they often have close links to other political structures). It became apparent during previous evaluations that the model of SFCs originally planned by the project had been transformed into a more locally appropriate model; this development was welcomed as a sign of local ownership of the project. Focus groups with SFCs revealed a high level of support for and approval of the project and a high level of collaboration between the SFCs and the project. The SFCs see themselves as the owners or managers of the project and the main decision makers. This was explained during a focus group with SFC members:

The school feeding committee which exists in each neighborhood decides the program of how and what will be cooked during the week and who will cook it, together with the school council members. The support the school director and the teachers to inform each neighborhood ahead of time who are the

mums who must go and cook, in coordination with the “secretários dos bairros” (neighborhood secretaries).

The project has provided training and reference manuals to SFCs.

The project has not formed public-private partnerships. The project leadership explained to the evaluator that they had initially believed that the renovation of premises owned and made available to the project by the municipality in Manhiça constituted such a partnership. However, they had subsequently discovered that this did not meet the criteria for a public-private partnership.

Use of Health, Nutrition and Dietary Practices (SO2)

The second large Strategic Objective (SO2) of the project concerns the use of health, nutrition and dietary practices and interacts directly with the first in that it feeds into Improved Student Attendance and therefore Improved Literacy of School-aged children within the project ToC.

Knowledge of Health and Hygiene Practice (2.1)

(Outcome): Percentage of students that demonstrate acceptable knowledge of health and hygiene practices. Final target: 50%
Midterm target: 31% (female 31%; male 31%)
Midterm results (March 2019): 85% (female 84%; male 85%)
Final target 170% achieved

Number of volunteer cooks receiving health and hygiene training. Final target: 8,468
Midterm target: 6,620
Midterm results (March 2019): 9,749
Final target 115% achieved

Number of people trained in good hygiene practices. Final target: 8,400
Midterm target: 6,275
Midterm results (March 2019): 9,394
Final target 112% achieved

Analysis of the project records demonstrate that the targets for training in good hygiene practice, including training volunteer cooks in health and hygiene, have been greatly surpassed. 82% of project schools reported that the volunteer cooks responsible for school feeding within that school had received training in good hygiene practices from project staff during the previous year (compared with 89% at baseline).

During a focus group with volunteer cooks, one participant explained:

Sr M taught us a lot of good things, first to sweep the floor before lighting the fire, then while the fire is catching we wash the pan before taking it to the fire, then we put the water in, and finally we put the soya in: that's all different from the way we used to do it.

At midterm of FFE 2, as during previous evaluations, pupils were asked a series of questions to assess their knowledge of good health and hygiene practices⁴⁹. At midterm, 89% of the FFE + literacy and the FFE only students combined claimed (without being prompted) to wash their hands before eating, compared with 84% of the comparison group. At baseline, 85% of the FFE + literacy students and 71% of the comparison students had claimed to wash their hands before eating (see figure 25 below). Although, at midterm, students from the intervention groups are more likely to claim to wash their hands before eating than the comparison group, the differences between the respective changes since baseline of the two groups are not significant.

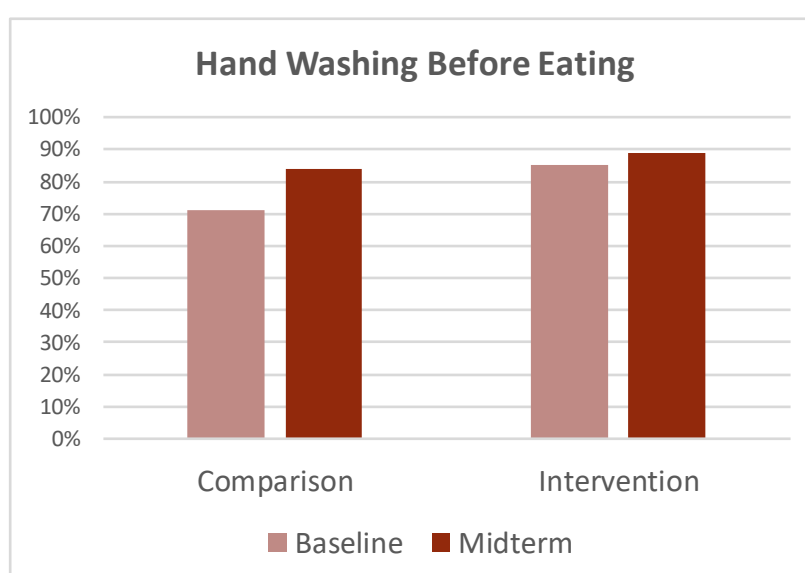


Figure 25: Students' reporting of hand washing before eating

At midterm 81% of the FFE + literacy and the FFE only students combined claimed (without being prompted) to wash their hands after going to the toilet, compared with 77% of the comparison group. At baseline, 84% of the FFE + literacy students and 66% of the comparison students had claimed to wash their hands after using the toilet (see figure 26 below). At midterm, the differences between the respective changes since baseline of the

⁴⁹ Pupils were asked what they did after using the toilet (enumerators were trained not to mention hand washing in their questioning, so as not to prompt this response). Pupils were also asked what they did before eating (likewise, enumerators were trained not to mention hand washing in their questioning, so as not to prompt this response).

two groups is significant, offering support for the intervention having impacted on the outcome, despite the intervention groups being slightly less likely to wash their hands at midterm than at baseline.

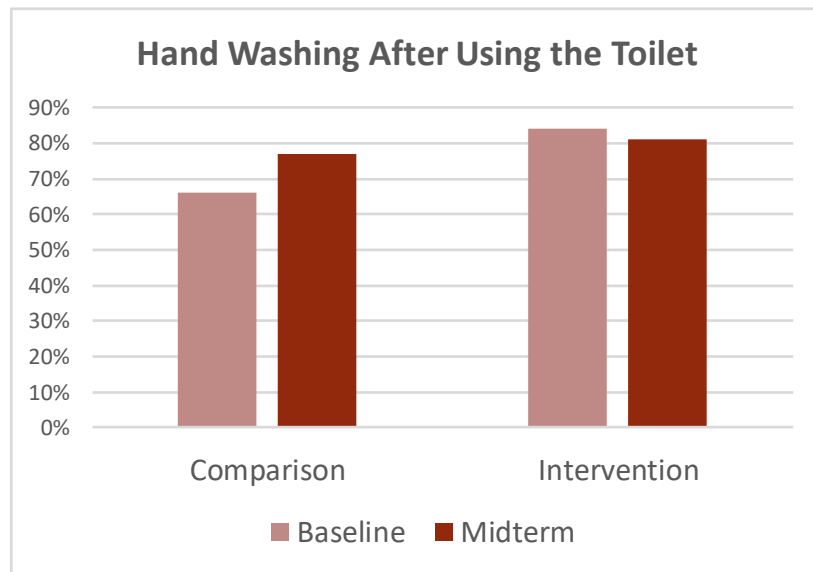


Figure 26: Students' reporting of hand washing after using the toilet

In each case, whereas previous evaluations have found significant differences between intervention and comparison students in both these respects, it seems that these are now plateauing out; for the grade 3 children in the intervention groups, since the project hygiene interventions began six years ago in most cases, their whole school career is likely to have been spent in a school where handwashing was strongly promoted, meaning that they have not experienced anything else, which would explain the relatively small change between baseline and midterm(see Technical Appendix 1 pp198 – 200 for more information). Nevertheless, taken overall, 85% of pupils in intervention schools demonstrate an acceptable knowledge of health and hygiene practices, which is a gratifying result for the project.

Teachers were asked whether they had seen differences in pupils' hygiene practices since the beginning of the project; 92% claimed that they had. Regarding their own knowledge of hygiene, teachers were asked to name three of the stages of handwashing, as taught during the hygiene training. 89% of them were able to name three stages correctly. This compares with the baseline survey where 66% of teachers were able to name three stages of hand washing correctly (see Technical Appendix 2, pp 58-60 and 366 – 77 for more detail).

Knowledge of Safe Food Preparation and Storage Practices (2.2)

Indicator 20 (Output): Number of food preparers at target schools trained in hand washing, safe food preparation and storage practices. Final target: 10,560 (female: 10,032; male: 528)
Midterm target: 8,250 (female: 7,837; male: 412)
Midterm results (March 2019): 9,749 (female: 9,588; male: 161)
Midterm target 118% achieved (female 122%; male: 39%)

Analysis of the project records demonstrate that the targets for training volunteer cooks in handwashing and safe preparation and storage of food have been surpassed. During the school survey, 82% of project schools reported that the volunteer cooks responsible for school feeding within that school had received training in handwashing and safe food preparation and storage from project staff during the previous year (compared with 89% at baseline).

Knowledge of Nutrition (2.3)

Indicator 19 (Outcome): Number of people trained in child health and nutrition as a result of USDA assistance. Final target: 10,300 (female: 7,519; male: 2,781)
Midterm target: 8,175 (female: 5,967; male: 2,207)
Midterm results (March 2019): 19,257 (female: 14,003; male: 5,254)
Final target 187% achieved (female 186; male: 189%)

Since the onset of the project, the nutrition education program has been implemented by a dedicated team managed and supported by WISHH and based in the project headquarters. Analysis of project records reveals that FFE2 targets for training individuals in child health and nutrition have been greatly surpassed.

71% of teachers reported having received nutrition training as part of the project (compared with 81% at baseline)⁵⁰. Of the modules which make up the training, 71% teachers reported receiving training in hygiene (76% at baseline); 64% in food groups (72% at baseline); 74% in Hygiene and food storage (69% at baseline); 55% in Handwashing (72% at baseline); 53% in Balanced diet (70% at baseline) (see Technical Appendix 2, pp 240 – 150).

⁵⁰ The goal (and assumed achievement) of the nutrition education of teachers was for 100% of all teachers in the 4 districts to receive training. The discrepancy between this and the 71% who reported receiving training may be accounted for by the nutrition education team failing to reach all teachers in all schools or, alternatively, by surveyed teachers failing to recall receiving nutrition education or failing to recognise the training received as nutrition education training (possibly because they consider “training” to be a more formal activity than the one they received).

Teachers were asked with whom they had shared the information learned during nutrition training. 87% claimed to have shared it with their pupils (compared with 76% at baseline); 29% with parents and guardians (43% at baseline); 41% with their own families (47% at baseline); 67% with the school council (37% at baseline); 34% with the community (34% at baseline) (see figure 27 below and Technical Appendix 2, pp 250 – 61).

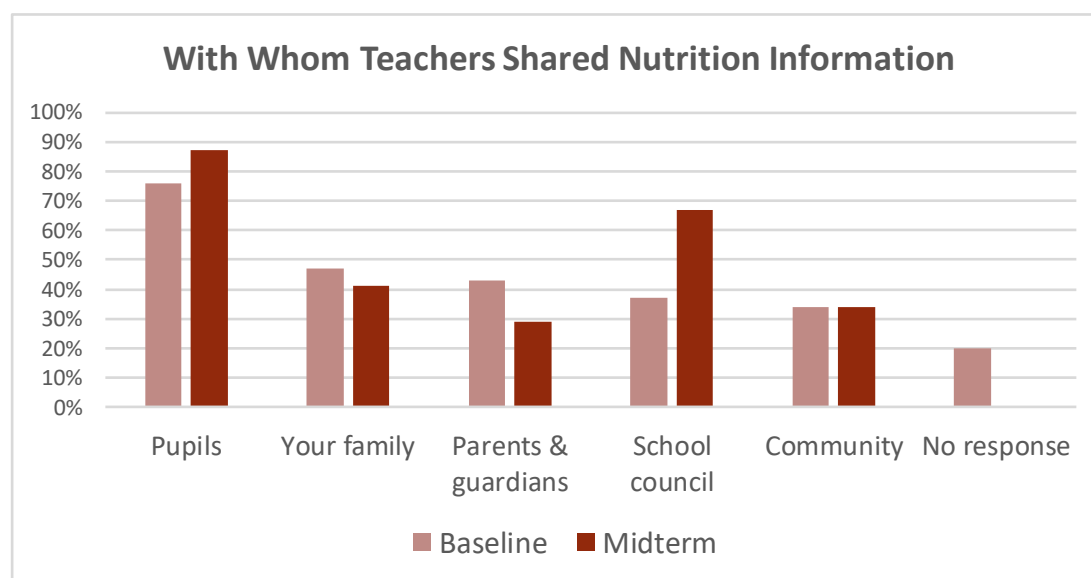


Figure 27: Teachers' reporting of sharing nutrition education

98% of those who replied to the question considered that the training had brought about changes in their own lives; 90% considered it to have brought about changes in the lives of their pupils.

89% schools (81% at baseline) reported receiving nutrition posters from the project; 81% schools (6% at baseline) reported receiving pamphlets; 87% reported receiving school feeding manuals (0% at baseline); 3% of school reported not receiving any nutrition materials. These results reflect the strong push by the nutrition team to distribute materials to schools as part of their exit strategy and to promote sustainability of the nutrition education component before their imminent withdrawal. Schools overwhelmingly reported that these materials were being used: 91% of schools reported teachers using them in their lessons (84% at baseline); 56% that they were displayed on the wall of the kitchen, storeroom or teachers' room (79% at baseline); 26% by being distributed to pupils (72% at baseline) and 10% that they were being distributed to the community (54% at baseline). These results reflect the shift in emphasis during the final stage of the nutrition education project to promoting use of materials by teachers in the classroom. During school visits, the evaluators frequently saw nutrition education materials distributed by the project either displayed on walls or being used in the classroom. The materials were of good quality, attractively presented with clear messages; they were robustly produced on laminated cardboard for durability. One item not referred to in the survey but observed during school visits was

an attractively produced recipe book containing pictures and simple instructions, giving simple, healthy, culturally appropriate recipes using recognizable, locally grown products.

The teacher survey included questions designed to evaluate teachers' knowledge of nutrition. Teachers were asked to provide examples of the different food groups used within the nutrition education program (base foods, protein-rich foods, foods rich in vitamins/minerals, energy-rich foods). Overall teachers performed well on this task, revealing a good level of assimilation of the nutrition education content. In the case of base foods, 90% of teachers were able to give two appropriate examples; 9% were able to give one appropriate example and 1% were not able to give any appropriate examples. In the case of protein-rich foods, 80% of teachers were able to give two appropriate examples, 16% to give one and 4% to give 0 appropriate examples. In the case of foods rich in vitamins/minerals, 85% of teachers were able to give two appropriate examples, 10% to give one and 4% to give 0 appropriate examples; in the case of energy-rich foods, 75% of teachers were able to give two appropriate examples, 17% one example and 8% to give 0 appropriate examples. In summary, between 75% and 90% of teachers surveyed were able to provide two examples for each food group; at baseline, this was the case in between 61% and 85% of cases (see figure 28 below). For more details, see Technical Appendix 2, pp 196 – 214 and 262 – 366.

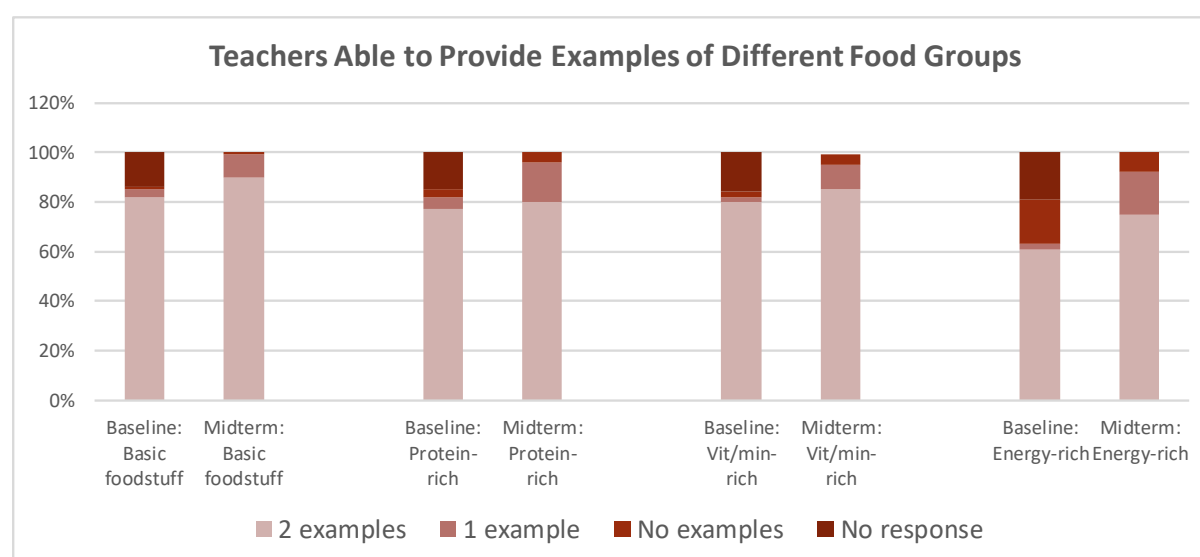


Figure 28: Teachers' ability to provide examples of different food groups

In order to triangulate the assertion by teachers that they share the nutrition training they have received with their students, pupils were asked whether their teachers talk about nutrition during school. Of the intervention group pupils, 9% claimed their teachers "often" talked about nutrition during classes (7% at baseline); 27% claimed they "sometimes" did (18% at baseline); 58% claimed they did not (53% at baseline); and 6% did not respond to this question (22% at baseline) (see figure 29 below). Although students were slightly more likely to

report their teacher talking about nutrition at midterm than at baseline, there is something of a discrepancy between these results and teachers' reporting that 87% shared nutrition education information with their pupils.

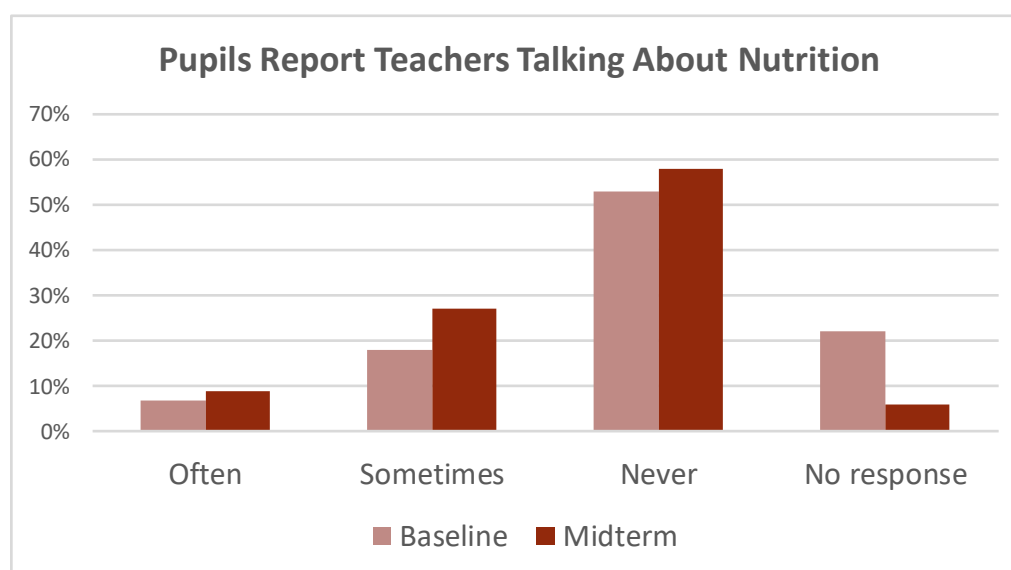


Figure 29: Pupils' reporting of teachers talking about nutrition during school

Where training content is shared, this may be rather formal and theoretical, rather than practical. One school director, when asked how the information received in nutrition training was shared replied:

First they go to the seminar and then we meet and they teach us that.../... we sit down together in the classroom and they start to teach us what they learned there.

Positive developments in the nutrition education identified in previous evaluations appear to be sustained at midterm of the second phase (better collaboration and information sharing between the nutrition education team and the rest of the project team, increased capacity, confidence, record keeping and reporting skills within the nutrition education team, more alignment between M&E of the nutrition education and the wider project). As mentioned above, the recent materials produced by the nutrition education team are of good quality and have been well received and appreciated by teachers and school directors. The midterm evaluation revealed a coherent exit strategy, with project staff, schools and communities being prepared to take responsibility for sustaining the work and achievements of the nutrition component and priorities being identified for the remaining months of the intervention, in order to maximize sustainability of the achievements of the nutrition education program once this has closed.

The school feeding focal point of one of the SDEJT described the training received in how to use the nutrition materials provided to schools:

This year we were trained in the use of the manual which has been produced, the manual to support nutrition education in primary schools. It was very interesting, it's more connected to the pedagogical area because there is a connection between the subjects in the manual and those in the curriculum. That manual can be consulted by the teacher according to the thematic unit they want to teach in class... there is a lot for the teacher to read and to reproduce, implement in their classroom, in their lessons.

Evaluation of the impact of the nutrition education program within the EPFs was rather less reassuring and there is some doubt about its sustainability, at least within some EPFs, after the end of the program. Interviews with the teacher trainers responsible for nutrition education and with the directors and pedagogical directors in the EPFs visited revealed a certain lack of dynamism, which was confirmed by the nutrition education team themselves. EPF teachers appeared to consider that, with the formal end of the nutrition education program, it will no longer be possible to continue to maintain the culture and content the program has sought to disseminate within the EPFs⁵¹. In particular, the expectation that there would be no further workshops or face-to-face contact with the nutrition team left some EPF teachers and nutrition focal points feeling abandoned and disempowered since they would no longer have resources provided by the project to work with.

This is in marked contrast to the vision expressed by WISHH, that nutrition education should hereafter be an integrated part of the EPF curriculum delivered to all trainee teachers, so that they are able to incorporate good quality nutrition education into their own classroom practice. As discussed with the nutrition education team, it is recommended that there should be a final series of face-to-face meetings with the EPF focal points and pedagogical directors with the explicit objective of transmitting methods and processes to sustain the nutrition education programs within the EPFs once the program closes. In particular, there should be a focus on practical, knowledge-based activities and methods, which do not require expensive resources to implement, and on inculcating within the EPFs a culture of promoting nutrition education, integrated with nutritional behavior, so that the theoretical and the practical dimensions reinforce one another.

⁵¹ The nutrition education team report that this issue has been addressed and the EPFs have agreed to integrate nutrition education into their teacher training programs across the board. However, this did not occur until after the midterm evaluation was completed.

Recommendations

The nutrition education program under WISHH managed team members is shortly to close, with the nutrition education program due to continue to the end of FFE2 through transfer of management to the FFE2 leadership.

A number of recommendations are offered for the time under WISHH management and thereafter:

Ensure that, in addition to knowledge of nutrition, school focal points have practical examples of how to share this in order to provoke positive changes in nutritional behavior.

Hold a final training session with EPF focal points under WISHH management aimed at ensuring continuity and ongoing transmission of information, through low-cost, knowledge-based activities and the creation of a dynamic culture. Dispel the myth that knowledge transmission requires financial resources.

Ensure that any remaining training is focused on practical application and producing change, rather than theoretical knowledge.

To promote the use of the excellent nutrition education materials designed, produced and distributed in schools, show focal points, head teachers and others how the materials can be used in schools. Give very practical advice on how they can be incorporated into lessons and the life of the school or community.

Percentage of school-age children receiving a minimum acceptable diet. Final target: 39% (female: 39%; male: 39%)
Midterm target: 34% (female: 34%; male 34%)
Midterm results (March 2019): 93% (female: 93%; male 93%)
Final target 274% achieved

As part of the student survey, students were asked what they had eaten the day before, in an attempt to establish how many can be considered to receive a minimum acceptable diet⁵². The nutrition education team were asked to code their responses as 2 “well balanced, i.e. all four food groups were mentioned”, 1 “reasonably well balanced but lacking in at least one essential food group” or 0 “not balanced/only one food group mentioned”. No students’ responses were coded as 2; 93% of students’ responses were coded as 1; 7% were

⁵² Establishing whether students are receiving a balanced diet through a simple survey is challenging. It was decided that the best way to gather information relative to this indicator was to ask the students what they ate the day before. This is not necessarily a reliable way of gathering information about diet, since recall may be poor (informants do not necessarily remember what they ate the previous day); furthermore, ideally responses should be followed up to prompt more complete recall, using follow up questions or pictures. However, in the present case, given the limited time available and the range of information to be collected, questioning was kept to the simplest form: “Can you tell me what you ate yesterday?”

coded as 0⁵³. Students whose responses were coded as “reasonably well balanced but lacking in at least one essential food group” were therefore considered to have a minimum acceptable diet. However, caveats apply (see discussion in footnotes).

Number of school gardens further developed and maintained. Final target: 60

Midterm target: 60

Midterm results (March 2019): 60

Final target 100% achieved

Number of school children benefiting from school gardens. Final target: 20,000

Midterm target: 20,000

Midterm results (March 2019): 19,929

Final target 100% achieved

School gardens have been encouraged and supported since the inception of the project. In addition to the 60 school gardens officially supported by the project plan, other project schools have school gardens.

The school survey asked schools whether they had a school garden or larger “machamba” without differentiating. Overall, the responses show a small but fairly consistent degree of progress since baseline (assuming that the sample of 170 schools surveyed is representative of the total number within the project). 72 of the 170 schools surveyed (42%) reported having either a school garden or “machamba.” Of these, 31% (16% at baseline) report growing food once a year, 10% (7% at baseline) twice a year, 14% (10% at baseline) all year around and 4% (1% at baseline) three times a year. 42% (23% at baseline) report only growing food during the rainy season. The productive area varies: 4% of schools (1% at baseline) report having between 3 and 5 hectares, 13% (3% at baseline) having between 1 and 2 hectares, 51% of schools (20% at baseline) having between 0.25 and one hectare and 32% (29% at baseline) having less than 0.25 hectares.

In some cases, the food grown regularly contributed to the school feeding. 40% (46% at baseline) of schools reported being able to show evidence of how the food produced was consumed; 49% (54% at baseline) report

⁵³During nutrition education, participants are taught that a balanced diet should contain items from each of four essential food groups. Students’ responses were only coded as 2, if they mentioned an item from each of the four food groups. A flaw in this system is that only three of the four food groups are “foods” per se (base – grains/starches, protein sources, and micronutrient sources – fruits/vegetables). The fourth food group is “energy” and is considered extra as it represents oil, other fats, and sugars. Unsurprisingly, students recalling what they ate the previous day did not mention items in the fourth group, meaning that none of them were coded as having an “adequate diet” in this scheme: no students scored as “2” because none mentioned 4 food groups. It is recommended that at end point, a more realistic coding system would be: 2 for responses that mention 3 food groups; 1 for responses mentioning only 2 food groups; 0 for responses mentioning zero or one food group. If this system is used at end point, the midterm responses will need to be recoded to ensure comparability over the two time points.

they have no evidence and 11% do not know. Where evidence exists it may be in the form of: a register of school production (38%, compared with 38% at baseline), a document provided by the SDEJT (10%, compared with 33% at baseline), it may be registered in the CSB+ distribution report (21%, compared with 6% at baseline), in a garden record book (10%, 0% at baseline) or another form (21%, compared with 10% at baseline).

Of the schools with gardens, 61% report receiving seeds from the project (62% at baseline). Of these, 16% received one type of seed, 16% received two types, 25% received three types and 43% received more than three types. 60% of schools with gardens (42% at baseline) reported receiving garden tools, such as hoes, rakes and machetes: 12% report receiving one type of tool, 40% two types, 30% three types and 19% report receiving more than three types of garden tool.

7% of the gardens are irrigated using river water (18% at baseline); 33% using rainwater (34% at baseline); 49% using water from a water fountain (42% at baseline) and 11% from another type of water source, such as a manual pump or tap water (7% at baseline). In 65% of cases this water source is available all year round (52% at baseline); in the other 35% it is seasonal (48% at baseline). Of the schools where water is only available in some seasons, 80% only cultivate food during those seasons: 20% manage to grow food all year around. 61% of the school gardens are fenced (compared with 46% at baseline).

Only 13% (compared with 6% at baseline) of schools with gardens have so far obtained a DUAT, the administrative document authorizing them to cultivate the land.

As far as community involvement in school gardens is concerned, in 31% of cases the community helps regularly (at least once weekly) with preparation of the ground and production of food (24% at baseline); in 42% of cases the community helps several times per year when there are large-scale activities (39% at baseline) ; in 26% of cases the community is not involved (37% at baseline). (See Technical Appendix 3, pp 34 -65).

51% of schools report growing legumes (31% at baseline), 78% report growing vegetables (41% at baseline), 49% report growing tubers and root crops (29% at baseline), 7% growing oil seeds (4% at baseline), 25% growing cereals (25% at baseline) and 13% fruit trees (10% at baseline), with 7% reporting other crops are grown.

(See Technical Appendix 3, pp 205 - 219).

The school survey included questions about the types and quantities of food grown in the past two years. The responses are summarized in table 14 below and compared with those reported at baseline⁵⁴. Although at

⁵⁴ Since the questions asked about food grown “in the past two years” it is possible that there may be some double reporting of food grown: the baseline survey covered years 2016 and 2017, whereas for the midterm data collection, which was conducted in April 2019, crops grown during the second half of 2017 may have been counted again. This does not necessarily greatly affect the patterns observed, since all schools were surveyed under the same conditions, and responded to the same questions.

midterm only 170 schools were surveyed, 72 of which reported having school gardens (compared with 156 of the total of 270 surveyed at baseline) comparison of the two time points suggests that:

- more schools are growing a more diverse range of products;
- the average quantities of these products grown by schools is increasing.

Although the number of schools producing quantities of food which can make some contribution to school feeding and diversification of the menu, at least for occasional meals is increasing, school gardens do not have the capacity to contribute to daily school feeding on a regular and sustainable basis.

	Schools report growing (%)	Schools report growing (%)	Mean quantity (kg)	Mean quantity (kg)	Number of schools producing 1-50 kg		Number of schools producing 51-100 kg		Number of schools producing 101-500 kg	
	Midterm	Baseline	Midterm	Baseline	Midterm	Baseline	Midterm	Baseline	Midterm	Baseline
Legumes	51%	31%	102 kg	52kg	31%	40%	10%	7%	11%	7%
Vegetables	78%	41%	109 kg	70kg	47%	40%	7%	20%	24%	10%
Tubers / roots	49%	29%	102 kg	75kg	26%	36%	13%	5%	10%	8%
Oil seeds	7%	4%	55 kg	47kg	4%	7%	1%	0%	1%	1%
Cereals	25%	25%	95 kg	70kg	15%	25%	6%	10%	4%	5%
Fruit trees	13%	10%	68 kg	114kg	8%	7%	1%	4%	3%	6%

Table 14: school reporting of produce grown in the past two years at Midterm and Baseline

Visits to schools during field work confirmed that school gardens are extremely dynamic in some cases, contributing to school feeding, nutrition education and other aspects of education such as natural sciences and mathematics. Those schools with active school gardens invariably have a director, pedagogical director or named teachers responsible for the garden activity, who is passionate about food production and gardening and highly motivated. One such school director explained that the secret of having a successful school garden in to involve everyone in both the work and the consumption:

When we began our little garden, the things we grew were for everyone, for us all. We sat down together, and we said we're going to increase our production. Everybody is going to be involved. When we have a harvest it's not only for the leadership, it's for everyone. The teachers got something to eat, everyone got something to eat, and they all felt involved. That was what motivated them, inviting everyone to participate.

These individuals reported that produce from the garden is used to supplement the CSB+ to improve the meals given to students, particularly on special occasions and festivals, when festive meals are prepared for the whole school or for the school and wider community. In a small number of cases they reported being able to regularly use manioc or sweet potatoes instead of CSB+, for example once a week. One school director explained:

Yes, as well as the soya we have cornmeal porridge (xima), from time to time with curry, we have, now we're going to have manioc with lettuce salad, yes, we've cooked maize, the mums are going to pound the maize to make cornmeal porridge.

They also enthusiastically reported receiving seeds from the larger HGSFGs and technical advice from the technicians attached to these. These visits confirmed the role school gardens can play in school feeding and nutrition education; however, it is clear that the relatively small-scale gardens in most schools, cultivated by teachers and pupils, do not have the potential to produce enough food to feed students regularly.

Number of Home-Grown School Feeding Gardens established. Final target: 8

Midterm target: 5

Midterm results (March 2019): 7

Midterm target 140% achieved

Number of school children benefitting from Home-Grown School Feeding Gardens. Final target: 6,600

Midterm target: 2,750

Midterm results (March 2019): 1,792

Midterm target 65% achieved

Home Grown School Feeding Gardens (HGSFG) are an innovation of FFE2 and aim directly to address the question of sustainability of school feeding into the future, after the end of the project activities. They are based on a vision of local production on a sufficiently large scale to support up to four schools and the principle that the surplus production can be sold to finance technical help and the purchase of foodstuffs such as salt and cooking oil needed to produce meals from the food produced. They aim to combine production of staple foods such as manioc and sweet potato with legumes, vegetables, salad and fruit; each HGSFG raises small animals (currently ducks).

At the time of midterm field visits, 7 HGSFGs were in operation; the external evaluator visited 6 of these⁵⁵. It had originally been intended that all eight HGSFGs should be operational by 2019. However, their establishment was more challenging than anticipated. One HGSFG had had to be moved from its original location due to deviation of the river on which it depended for irrigation by unregulated extraction of sand from the river bank, with the result that it was no longer possible to irrigate the plot. Having to concede that the garden was not viable was a great blow to the school and community which had invested in its initial preparation; at the time of fieldwork the motor pump and other resources were being removed to be installed in another site linked to another school.

⁵⁵ As of October 2019, the evaluator is informed that all 8 HGSFGs are now in operation.

During field visits, the evaluator was impressed by the scale of the HGSFGs and the extent of progress already achieved. Considerable areas have been cleared and planted, with the help of community members, who contribute to production for the school and in return receive an area benefiting from the same irrigation system to cultivate for their own consumption. At the time of field visits (April 2019), diverse crops were being produced: base foods, such as manioc, maize, plantains and sweet potatoes; different types of bean, legumes and tubers; vegetables and salads: lettuce, tomatoes, onions, cabbage, peppers; and fruit trees: mango, oranges, grapefruit, papaya, bananas. Each HGSFG visited also had a recently arrived flock of ducks, which were being carefully raised and tended.

Each HGSFG is based in a school and is the primary responsibility of the head teacher, with the technical work being supervised and coordinated by agricultural technicians. In each HGSFG visited, the school director and the technician were interviewed. The external evaluator was extremely impressed by the knowledge, commitment and dynamism of the technicians. These individuals had previously worked as agricultural extensionists in other ADPP projects, such as farmers' clubs. They demonstrated extensive knowledge of both agricultural techniques and community development. They work cooperatively alongside the community volunteers, showing them improved agricultural methods and practices but also being prepared to use traditional methods where these are preferred. One technician showed how he was engaged in a friendly competition with the community volunteers: the community had planted several rows of vegetables using their traditional approach and he had planted several rows alongside them using a modern method assumed to be more efficient; the harvest would demonstrate which was the winner. These knowledgeable individuals work long hours alongside community workers, leading by example, in addition to planning, keeping records, and liaising with the school. Their role appears key to motivating and engaging the community volunteers. In each HGSFG they are training one community member in particular to become the future coordinator. The technicians are also supporting neighboring school gardens, providing seeds from their production and technical advice.

The school directors demonstrated the record keeping methods used to record and account for the production. The records appear well kept, although it is not clear whether they include all the food produced. This raises one important issue regarding the management and governance of the HGSFGs. Head teachers, who have been trained to manage a school (in the best of cases - not all school directors in Mozambique have actually been trained as such) are not necessarily well qualified and equipped to manage fairly large scale agricultural production and to account for it rigorously. This requires significant skills, time, probity and influence within the community. It is not always easy to account for food produced; when large quantities of, say, lettuces are produced, as observed during field visits, who is to say who can take a lettuce (or several) home? This question is all the more pertinent in a context of relative food insecurity, where many families struggle to eat, let alone to eat a balanced diet. It is, however, a question which will need to be addressed if the HGSFGs are to succeed and maintain community adhesion and support. The evaluator learned of one school where a crop of potatoes

cultivated by the community and the school had been taken to market and sold without the profits being returned to the school. This had demotivated the community members who had worked to grow them, who were no longer participating in the work of the HGSFG. It is clear that rules and systems governing the utilization and consumption of the produce need to be strengthened and the recording of this needs to be improved.

Evaluation fieldwork revealed many positive aspects of the new HGSFGs, as well as some which will need to be closely monitored if the initiative is to be successful in the long term.

The creation of large-scale, completely fenced market gardens, with motor pumps, irrigation systems and a reliable water source is a huge achievement in a challenging environment, as is the fact that these are already yielding considerable quantities of produce and vegetables and, in some cases, contributing to the diversification of school feeding.

Recommendations
In the future, once the project is no longer distributing CSB+, an alternative base food, possibly sweet potatoes, will need to be produced or procured, in sufficient quantities to continue daily school feeding.
The existing plans and systems need to be reinforced to ensure the HGSFGs are well managed and sustainable. The food produced needs to be consistently recorded and accounted for, using transparent processes.
The agricultural technicians who currently play such an active role in coordinating and doing much of the work will either need to be retained and paid or their place taken by the community members currently being prepared to replace them. In this case, these individuals will need to develop considerable competence, including in production management, maintenance of the motor pump and other equipment, and the ability to mobilize their fellow community members and coordinate their activities.
Current levels of commitment from schools and communities will need to be sustained and, in some cases, increased.
The inclusion of small livestock rearing in the HGSFGs is an impressive innovation and offers the potential for greater diversification of diet and income generation to finance some overheads, such as fuel for motor pumps, fencing, tools, salaries, etc. Those responsible for the livestock should be made aware of principles of animal welfare. Ducks, for example, need constant access to water.

Overall, at present, the HGSFGs are evidently making a significant contribution to the diversification and sustainability of school feeding, at least in the school where they are based and which they support. Once the support from the project finishes, they will need to be self-sustaining. This will be a considerable challenge, one which is certainly not impossible to meet, but which will require considerable leadership and good management in order to succeed.

Access to Clean Water and Sanitation Services (2.4)

The project recently divided the water and sanitation component into two; the sanitation activity now comes under the responsibility of the construction coordinator. This leaves the water team available to address the challenges and demands of this delicate domain. The report first addresses the activities of the water sector, before reporting on sanitation activities.

Access to Clean Water

Indicator 22 (Output): Number of schools using an improved water source. Final target: 264
Midterm target: 226
Midterm results (March 2019): 270
Final target 102% achieved

Number of school children benefitting from clean water sources. Final target: 54,000
Midterm target: 54,500
Midterm results (March 2019): 87,847
Final target 163% achieved

Number of water tanks secured. Final target: 330
Midterm target: 330
Midterm results (March 2019): 295
Final target 89% achieved

Access to safe water continues to be a condition for school feeding, local production of food, health and wellbeing; despite the water team's best efforts, this also continues to be a challenge for some project schools, due to the depth of the water table, the saline quality of much of the subterranean water and the distance of many communities from the existing municipal piped water network. In many schools, the absence of safe water some or all of the time is a major problem. One school director interviewed painted a desperate picture of a school without water, where children and teachers cannot drink during the day, there is no water for handwashing after using the latrines and teachers who spend the night at the school cannot wash before work. Of course, the preparation of the soya porridge also requires water. One school director explained:

The big challenge is water, because children have to bring water from home [to make the soya porridge]. Often they don't have water at home either.

The water team is now stable, compared to phase one of the project where it was difficult to recruit and retain water technicians; all four water technician posts are now filled, and the same committed and competent coordinator leads the team. In partnership with the State, some schools who can afford to pay water bills are being connected to the piped water network. Progress has been made in decentralizing funds for small water interventions, which are now administered at district level, allowing more rapid repairs and maintenance. At

this stage in the project, with earlier organizational problems resolved, it seems that the challenges to do with ensuring schools have access to clean water are more to do with the intrinsic difficulties of finding technical solutions in the challenging environment in which the project operates. Unfortunately, a number of contracts for drilling boreholes have not been honored, meaning new contractors need to be identified and new contracts issued. The water coordinator explained that some of the theoretical solutions to providing schools with water, such as connecting them to the mains supply or delivering water by tanker trucks to fill up the water tanks which all project schools have received are not possible within the current budget. Furthermore, having water standing in tanks for long periods of time is a health and hygiene risk.

There continues to be a disconnect within the reporting of water team activities between the numbers of interventions reported and the number of functioning water systems. This is confirmed by comparing the figures from project records, which show targets as surpassed with the number of respondents to the midterm school survey who stated their school did not have a water system or did not always have safe water. Recommendations of earlier evaluations to develop clearer criteria for considering water installations as completed and therefore “countable” and more focus on outcomes (functioning water systems) have not led to this reporting discrepancy being completely resolved.

The FF2 baseline report carried out a thorough inventory of the different types of water installation at each school, which it was not considered necessary to repeat at midterm. Of the schools surveyed at midterm, 74% report having have some sort of system to supply water: piped mains water, electric water pumps, manual water pumps and/or rainwater harvesting (compared with 94% at baseline); 26% do not (6% at baseline). Of those that reported having a water system, 73% reported the system functioned “every day this year” 15% that it functioned “most of the time”, 7% that it functions “from time to time”, and 4% that it “doesn’t work”.

57% of these schools reported having a plan for the routine maintenance and cleaning of their water system (58% at baseline). This plan was known variously to the school council (54%), named teachers (61%), project staff not connected to the school (31%), the school director (38%), the pedagogical director (24%) and/or other people (24%).

68% reported having a water committee or designated people responsible for looking after the water system (69% at baseline). The composition of the water committee included the teacher responsible for the storeroom (43%), teachers from the school (33%), project staff not connected to the school (19%), the school director (33%), the pedagogical director (20%) and/or other people (38%). Those who actually performed the maintenance included: school council members (47%), members of the neighboring community through the school council (37%), designated teachers (37%), project staff not connected with the school (29%), pupils from the school (10%) and/or others (10%). For more details, see Technical Appendix 3, pp - 131 - 172.

When asked how often the water system was cleaned, responses varied between daily (25%, compared with 27% at baseline), weekly (17%, compared with 29% at baseline), twice weekly (6%, compared with 10% at baseline), “from time to time” (40%, compared with 35% at baseline) and “other” (13%).

A source of clean water continues to have a transformative effect on the schools and communities to which it is brought. Unfortunately, several schools still do not have such a water source or only have water during part of the year when rainwater harvesting is possible. The way the indicator is formulated may hide this reality.

Access to Sanitation Services

Indicator 23 (Output): Number of schools with improved sanitation facilities. Final target: 264
Midterm target: 264
Midterm results (March 2019): 246
Final target 93% achieved

Number of hand washing facilities rehabilitated. Final target: 171
Midterm target: 171
Midterm results (March 2019): 140
Final target 82% achieved

Number of latrines and hand washing facilities constructed. Final target: 94
Midterm target: 94
Midterm results (March 2019): 108
Final target 115% achieved

Number of educational facilities (i.e. school buildings, classrooms and latrines) rehabilitated/constructed as a result of USDA assistance. Final target: 808 (kitchens 264; storerooms 264; latrines 280 [94 constructed + 186 renovated])
Midterm results (March 2019): 862 (kitchens 194; storerooms 410; latrines 235 [108 constructed + 127 renovated])
Final target 107% achieved

Number of educational facilities (i.e. school buildings, classrooms and latrines) rehabilitated/constructed as a result of USDA assistance. Final target: latrines 94 constructed + 186 renovated = 280
Midterm results (March 2019): latrines 235 [108 constructed and 127 renovated]
Target 84% achieved

Number of school children benefiting from latrines and hand washing facilities. Final target: 48,000
Midterm target: 48,000
Midterm results (March 2019): 78,311
Final target 163% achieved

During FFE2, the installations which were constructed in all schools during the first phase (kitchens with firewood saving stoves and storerooms) are being maintained and renovated, including improving their ventilation. A large program to build and renovate improved models of latrines, all with handwashing facilities is also underway.

Analysis of project records reveals that the target for construction of handwashing facilities and of latrines has already been reached, and the rehabilitation of existing handwashing facilities and latrines is slightly behind schedule. The building component which is now responsible for construction and rehabilitation of latrines and handwashing facilities was without a manager for some months, which led to delays in implementation of the plan. At the time of fieldwork, a new manager had just been appointed a week previously and was still learning the details of the component. He had previously been the coordinator of one of the project districts, so had good knowledge of the component but was still learning some of the management processes and aspects such as procurement. Like the other new appointments, he will need to be closely supported and mentored in this role initially and possibly receive specific training. With a new manager in place, it should be possible to catch up with the delays in latrine renovation and meet the targets without too much difficulty by the end of the project.

A major priority for the remainder of the project is to ensure communities and schools are able to maintain the infrastructure after the close of the project:

We have always had close contact with the school council and the school leadership and we remind them that the assets need to be looked after and maintained, because that infrastructure doesn't belong to ADPP but to the school and the community for the benefit of the children. There is a continuous mobilization to get the message across that after the project finishes, it is they who will need to continue and maintain the infrastructure and everyone needs to play their part.

An example of this need was encountered during visits to some schools, where the external evaluator was told by school directors that some of the latrines renovated by the project several years previously were now full and could no longer be used. This is a good example of the need to look beyond targets and respond to situations encountered.

Recommendation
A plan is needed to identify those latrines which are likely to become full in the near future and work with the schools involved to take steps to either cover them over and replace them with others or to empty them where this is possible. Even where latrines will not become full for some time, it would be good practice as part of the sustainability plan to raise this issue with school councils and provide training so that they can manage the situation when it does arise.

Access to preventative health interventions (2.5)

Indicator 24 (Output): Number of students receiving deworming medication(s). Final target: 74,000
Midterm target: 74,000
Midterm results (March 2019): 91,103
Final target 123% achieved

Number of cooks tested for TB. Final target: 5,280
Midterm target: 5,280
Midterm results (March 2019): 8,402
Final target 159% achieved

The project has provided logistical support to the local health services to provide de-worming tablets to all pupils once annually. Analysis of project records reveals that this target has been surpassed as of midterm.

When asked whether they had undergone deworming during the current school year, 65% of students in project schools reported that they had (compared with 77% of the cohort at baseline). This lower result at midterm may be due to the fact that the midterm survey was administered towards the beginning of the school year (March – April 2019). 37% of comparison group students reported having received deworming tablets during the current school year (compared with 63% at baseline) (see figure 30 below). The lower figure at midterm than at baseline for the comparison group may also be due to the timing of the midterm evaluation. However, given the significant difference between the rates reported by students in project schools and in comparison schools, it is likely that this is impacted by the fact that the deworming activity is subsidized in project districts, by contributions made by the project to the fuel costs; in some cases, project staff also accompany the deworming visits and promote participation in this activity to students and their families.

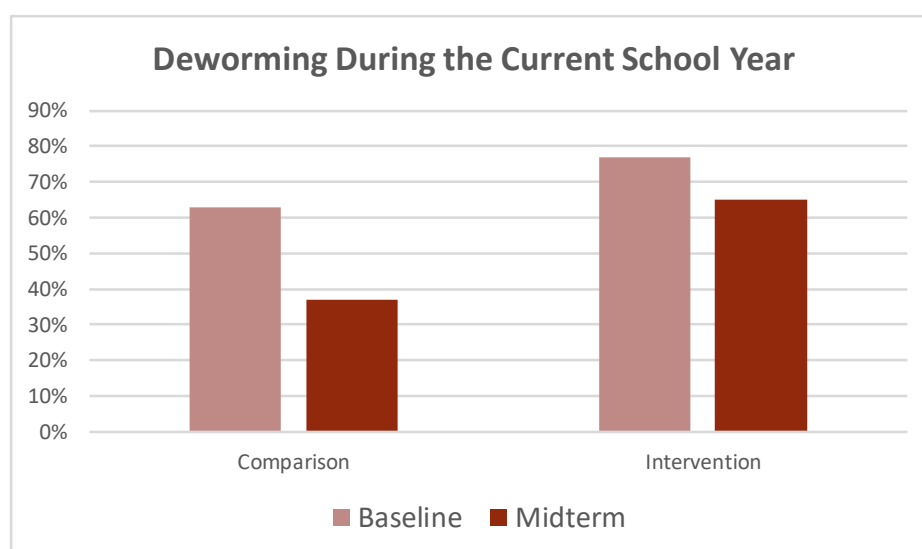


Figure 30: students' reporting of deworming during the current school year

The testing of volunteer cooks for tuberculosis was introduced during the second phase of the project in order to protect students from transmission of the disease. Final targets for this activity have been exceeded as of midterm.

Health officials from the SDSMAS (district health services) welcomed the collaboration between the health services and the project and, in particular, the fact that volunteer cooks were undergoing testing for TB.

Access to requisite food preparation and storage tools and equipment (2.6)

Number of community volunteers receiving laundry soap. Final target: 5,280
Midterm target: 5,280
Midterm results (March 2019): 5,420
Final target 103% achieved

Number of schools receiving dish washing soap. Final target: 264
Midterm target: 264
Midterm results (March 2019): 271
Final target 103% achieved

Distribution of laundry soap to volunteers was a recommendation made by evaluations of FFE1, following repeated requests from volunteers for soap to help them keep their uniforms clean. Likewise, volunteers requested dish washing soap to improve the cleanliness and hygiene of the CSB+ preparation and subsequent washing of dishes and utensils. Analysis of project records reveals that the targets for these activities have been met. During fieldwork, volunteer cooks expressed satisfaction at having been given soap, which is seen as more than an incentive and rather as a necessity for their work and therefore a recognition of their status within the project and contribution to the task of school feeding. The motivation and commitment which result from this relatively small incentive can be considered good value for money.

Capacity of government institutions (2.7.1)

Number of government officials trained in nutrition. Final target: 100
Midterm target: 65
Midterm results (March 2019): 302
Final target 302% achieved

Number of participants in school feeding conferences (district, provincial and national). Final target: 180
Midterm target: 102
Midterm results (March 2019): 1,208
Final target 671% achieved

Number of seminars and meetings conducted at the local, regional and national level. Final target: 4
Midterm target: 2
Midterm results (March 2019): 18
Final target 450% achieved

As part of the commitment to build local capacity and to sustain school feeding and other activities after the closure of the project, the project has trained government officials at national, provincial and district levels in school feeding. The target for this activity has been greatly surpassed. Government officials interviewed during fieldwork were enthusiastic about the training received and stated it had helped them play their role effectively. The school feeding focal point of one SDEJT explained:

The trainings are good, they are good because after receiving the training we feel a difference, we compare the before and after and we feel when we go into schools that something is improving thanks to the training or the seminar.

The project has also held a number of meetings and conferences to allow actors involved in school feeding in Mozambique, including the national school feeding program (PRONAE) and various other initiatives supported by NGOs and other actors, to meet and exchange experiences.

Engagement of Local Organizations and Community output (2.7.4)

Number of local leaders and school council members attending school feeding project management training. Final target: 2,000 (female: 500; male: 1,500)
Midterm target: 1886 (female: 471; male: 1,414)
Midterm results (March 2019): 2,498 (female: 689; male: 1,320)
Final target 125% achieved (female: 138%; male: 88%)

Since its inception the project has made efforts to train local leaders and school council members. During focus groups with community leaders and volunteers, participants stated they had been trained and appreciated the value of that training:

We've all received training.

Thank you for teaching us and helping us prepare food for the children. We're very grateful. Come back and teach us more any time!

This is important to i) promote ownership of the project by the communities in which it operates ; ii) ensure interventions are culturally appropriate and correspond to local needs and objectives; and iii) foster sustainability, by ensuring that, after the close of the formal project interventions, local actors have the capacity and vision to continue with those activities judged necessary and appropriate. At midterm, analysis of project records reveals that targets for training local leaders and school council members have been surpassed. During evaluation fieldwork, evaluators were impressed by the level of ownership of the project by local leadership structures, who had adopted their own systems and methods for selecting and organizing the rosters of volunteer cooks, water committee members and volunteers to work in school gardens and HGSFGs. Interviews and informal discussions with community leaders showed a high level of commitment to the project and its activities, a determination that it should continue and considerable knowledge and understanding of the technical methods needed. Although they do not of themselves guaranty sustainability, these factors are among the prerequisites to ensuring the project activities continue after the formal closure of the project.

Transversal themes

In addition to the presentation structured according to the MGD indicators, a number of transversal themes deserve attention; these cut across the different components and activities and concern the overall operations, strategy and vision of the project. They are: human resources, capacity, collaboration and ownership; transport; administrative and financial systems and procedures; monitoring and evaluation; and sustainability and relevance to the local and national school feeding policy and program environment. Within each of these themes, the evaluation revealed both positive aspects and areas of progress since the FFE2 baseline and areas which are still evolving, or which should be monitored and given additional support as needed.

Human Resources, capacity, collaboration and ownership

With the FFE program now in its seventh year of functioning, some staff now have significant experience and have demonstrated significant professional development; many staff have been promoted to functions with more responsibility. In particular, several of the “professionals” have been promoted to positions of district leaders, component manager or leadership. These welcome changes will need to be monitored and accompanied to ensure young, newly-promoted staff receive appropriate support and training.

Progress has been made in terms of gender balance within the project team. Three of the four district coordinators are now women; some components are headed by female managers, notably M&E and literacy. Some components still lack gender balance; in particular the leadership team is composed of four men, following the departure in 2018 of the only female member, running the risk of giving the impression of a “glass ceiling” to project staff.

Overall, the “professionals” have greatly grown in capacity, competence and professionalism. In many cases the novice professionals recruited seven years ago have grown into impressive, experienced professionals. Various “professionals” have developed areas of specialization and are called to apply these in different schools. Their work has been reorganized so that, amongst the “professionals” there are focal points for specific areas, such as school clubs and school gardens, leading to more efficient and skilled support to schools.

It is clear, however, that not all “professionals” have the same level of commitment to their work; the evaluator was told that some “professionals” do not regularly visit the schools. A well-planned, sensitive performance management system might be helpful to motivate the demotivated and spread the workload more equitably.

The “professionals” personal and professional situation has changed over the past seven years⁵⁶. They have families to support and many have to maintain a household at distance, only returning some weekends to see their spouses and children. Over the same period, the cost of living has spiraled in Mozambique. Many of the “professionals” explained that they are struggling to feed and clothe their children and wider families. Their salaries and conditions should be reviewed and improved to reflect this reality, in combination with a performance management scheme, so they are able to focus on their professional commitments in the knowledge their dependents are properly taken care of.

Communication and coordination between project staff at all levels (from leadership to “professionals”) with the GoM at all levels (national, provincial and district) and with the local communities is excellent. During evaluation interviews, GoM and community informants consistently praised the quality of the relationship, communication and collaboration with the project, comparing it favorably to that with other partners and projects. However, communication and coordination within the project team could be improved in some cases; alignment of workplans and working practices, defining work system and procedures and better communication between components would improve the efficiency and quality of some operations.

The level of commitment, engagement and effectiveness of the volunteers has improved overall. In general, volunteers appear to have accepted the culture of not providing monetary incentives and embraced the opportunity to improve their children’s diet and education. At the same time, innovations such as providing dish soap and laundry soap have been appreciated and taken as an expression of dignity and a recognition of the important role played by volunteers. To date, the teachers who play the role of storeroom managers have not received any such incentives.

In order to have as much information as possible about the school feeding environment in Mozambique, the evaluation sought to include other organizations currently working in school feeding in the country. One model

⁵⁶ A remarkably high number of the “professionals” have been in that role since the beginning of the project. This indicative of both the supportive culture of ADPP and the project and the barren employment prospects in Mozambique for many young professionals.

being implemented by JAM (Joint Aid Management), in the spirit of their slogan: “Helping Africans to Help Themselves” provides an interesting example and food for thought, particularly in light of the discussion of ownership and incentives above. JAM is implementing a school feeding program in 48 schools in Sofala and Inhambane provinces, which also uses CSB+ as the base food. The 48 schools develop school gardens to teach students about nutrition and grow produce to diversify their diet. The biggest difference with the FFE2 project is that JAM’s project functions as a value chain. Corn and soya grown by farmers according to JAM’s protocols and following training from JAM are sold to a factory in Dondo, Sofala which processes it to produce CSB+ (fortified with a premix imported from Holland) which is in turn sold to schools for school feeding. In the initial stages, the schools are assisted by JAM with the funds to purchase the CSB+; however, in the long term, it is planned that schools should be able to fund these purchases using the proceeds of other income generating activities, including the school gardens. There are also plans to sell CSB+ in bags of one kilogram on the open market for non-school use. The project is in the early stages but there are plans to scale up to include more schools. JAM also provides water and sanitation using solar panels or hand pumps, mobilizes parents in PTA structures and nutrition and food security training activities. JAM runs a model farm and learning center for farmers, researchers and extensionists and also provides training in financial management. According to the project representative interviewed, both the farms growing the corn and soya and the factory transforming them into CSB+ are “commercially sustainable”, employing their staff and making a profit. This model of a commercial activity, initially supported and subsidized by a supportive partner, offers answers to some of the challenges encountered by the FFE2 project, such as the tension between the principle of working with volunteers and the very real need of families living in poverty to receive income in exchange for their work and time.

Recommendations
Newly promoted staff need to receive appropriate support, training, professional development and orientation.
All appointments and promotions should use equitable processes which promote diversity. When making appointments to the leadership team, efforts should be made to identify and encourage female candidates.
Look at the possibility of introducing a well-designed performance management system for the “professionals” to increase motivation and spread the workload more equitably.
Review and the salaries and conditions of the “professionals” with a view to improving these in conjunction with a performance management scheme.
Improve communication and coordination within the project team, including alignment of workplans and working practices, better definition of work system and procedures and communication between components.

To increase commitment and motivation amongst storeroom managers, they should receive an expression of thanks and some recognition for the central role they play. For example, they could be provided with phone credit as compensation for the weekly SMS they send to report the weekly stock levels of CSB+.

Transport

The acquisition of additional vehicles and motorbikes and the strategy of providing transport to the “professionals” and supporting them to obtain driving licenses recommended in previous evaluations has been successful. Most of the “professionals” now have driving licenses and their performance and productivity has increased accordingly.

The project still does not have sufficient, appropriate vehicles for all its activities. The components responsible for CSB+ delivery, water and construction activities have to juggle their activities in order to share the available vehicles, leading to inefficiencies. Hired transporters are unreliable and often their vehicles are not fit for purpose.

The project drivers have a good knowledge of the schools in the four districts and the roads between these and play a valuable role in supporting logistical operations. Delays in processing and paying drivers’ *per diems* and expenses puts drivers in difficult situations; where they go out for the day without cash to cover their incidental costs, such as road tolls, and meals, they have to resort to borrowing money from their passengers.

Recommendations

The project needs to acquire additional vehicles in order to undertake its activities in due form.

Systems to process and pay drivers’ *per diems* and expenses should be improved to avoid putting them in embarrassing situations.

Administrative and financial systems and procedures

The influence of a strategically focused financial manager with a vision of how systems and procedures can be improved has led to improvements in administrative and financial systems and processes and streamlined some previously onerous workplace tasks. For example, the creation of a procurement office/purchasing service has led to all purchases being processed by dedicated staff. The introduction of a reception desk within the Administrative Section through which all administrative and financial requests have to pass, has reduced waiting time and made processing requests more efficient. Recruitment and training of additional staff has allowed more specialization of activities. Administrative systems could be made even more efficient by further improving processes, recruiting additional staff and training existing staff. Staff require continual training to improve work

quality and reduce errors. Additional recruitment would be desirable, for example the project would benefit from an HR manager based in the Manhiça headquarters.

The standard and quality of documentation has improved as staff better understand the need for consistency and rigor in preparing and processing documents and that all expenses or requests for funds must be supported by the correct documentation. One dedicated member of staff now checks all documentation, leading to more consistency. Still more improvements are needed, so that preparing and submitting documentation which is acceptable the first time and does not need to be sent back for improvement becomes the norm.

Some administrative processes and the management of some funds have been decentralized to district office level, leading to greater flexibility and responsiveness and reducing delays and bureaucratic bottlenecks. For example, the district water technicians have funds for small parts and repairs, so these do not have to be processed and authorized centrally, leading to more rapid interventions and flexibility. Staff in the districts required ongoing support, supervision and professional development in order for these decentralized systems to work effectively. Internal communication and coordination still need improvement; for example, workplans need to be aligned and district or school visits coordinated.

Recommendations
Look for ways to further improve administrative processes, recruit additional staff and train existing staff, for example consider employing an HR manager based in the Manhiça headquarters.
Continue to improve the standard and quality of documentation.
Provide ongoing support, supervision and professional development to staff based in the districts.
Continue to improve internal communication and coordination for example, by aligning workplans and coordinating school visits.

Monitoring and Evaluation (M&E)

High turnover of M&E management and staff throughout the life of the project to date have led to a lack of overall vision and continuity in M&E processes and procedures, some gaps in data and a lack of institutional memory of previous M&E situations and steps taken. This challenging M&E context has been mitigated to some extent by continuous support from the project leadership team and from the distance advisor based at Humana People to People Headquarters in Zimbabwe, Stanley Kudzibatira, particularly in preparing semi-annual reports and supporting evaluations, and by support from the consultant Cade Fields-Gardner. As of March 2019, a new

M&E coordinator and a new M&E manager have taken their functions; although it is early days there is every sign that these individuals have the necessary skills and attributes lead the M&E department effectively and gradually make up for the gaps in provision over the years.

Some M&E activities have been decentralized to the districts; each district now has a data clerk and an administrative assistant to perform them. The information flow is now better defined and clearer: information from schools is transmitted by the “professionals” to the district staff for data entry. This information is included in district coordinators’ reports and sent by them to the project M&E department and project leadership simultaneously. Data received from the four districts is processed as necessary by the M&E department and transmitted as necessary to Planet Aid; Planet Aid submits semi-annual reports to USDA.

Despite these improved processes, some of the data collected and entered are still not being used. For example, school attendance data is collected from schools using the carbon copies from the specially designed school registers; it is then digitized by the district data clerks. However, as of the midterm field visits, this information was not being analyzed or used in any way: project school feeding information was being used to report on attendance.

There is still a lack of clarity in the measurement and/or formulation of some indicators. For example, as indicated on page 122 above, there is a mismatch between the indicators measuring water system installations and those measuring schools with access to safe water systems.

The imminent arrival of the long-awaited database – the second database commissioned by the project, which has been subject to successive delays - should allow better management and vision of information overall. It is imperative to invest the necessary time and human resources to get the database up and running as quickly as possible, so that this investment can begin to bring real benefits before the end of the project.

Field visits revealed coherent, orderly and systematic practices of filing and archiving written information / paper records by members of the M&E team, who were able to locate and produce information and records as requested by the external evaluator as part of her series of spot checks of M&E records.

In addition to this commendable rigorous adherence to procedures, M&E staff need to have a vision of and understand the whole process: they should know why information is being collected, what it will be used for and how. Component managers and staff also need to have a wider vision of the M&E of their activities: they need to know where they are position in relation to project indicators and targets and to be able to use this information to plan and prioritize their operational tasks. The project leadership team should also, of course, have an overall vision of the project’s real-time situation in relation to the different indicators and targets.

Recommendations
All data collected and digitized should be used in some way; the project should stop collecting and processing information which is not being used. Review indicators and how these are measured to ensure they accurately capture the information intended. As a priority, ensure the new database is up and running as quickly as possible. Ensure M&E staff have a vision and understanding of the whole M&E process (why information is being collected, what it will be used for and how); component managers and staff have a wider vision of the M&E of their activities (where they are position in relation to project indicators and targets) and can use this information to plan and prioritize their operational tasks; project leadership team have an overall vision of the project's real-time situation in relation to indicators and targets.

Sustainability and relevance to the local and national school feeding policy and program environment

So close to the closure of the project, the theme of sustainability was a priority: it informed every aspect of the midterm evaluation and was an integral part of all instruments used for qualitative data collection and all interviews, focus groups and conversations with informants at all levels. The question of sustainability of the FFE2 project is intimately connected with the national school feeding program (PRONAE) and also with the overall political, economic and social context in Mozambique.

The National School Feeding Program (PRONAE), which was piloted in a small number of schools Gaza, Manhica, Tete and Nampula from 2013, has since been expanded to 70 schools, with support from the WFP. A debt from Mozambique to Russia was recently converted into funding for development, within which the PRONAE will expand to 300 schools by 2021. This is a welcome development but still only covers a fraction of the national demand. The MINEDH representatives responsible for the PRONAE express great appreciation of the FFE2 project and all it is achieving; they also welcome the openness of the project leadership and staff to working in partnership to plan future expansion and reinforcement of the PRONAE and have welcomed the opportunities for sharing of experiences and learning provided by the school feeding conferences and workshops organized by the FFE2 project. It is clear, however that there are many challenges associated with the local procurement model embraced by the PRONAE, including the technical and bureaucratic obstacles to local production, procurement and transport of food; the difficulty of incentivizing volunteers to work to produce food for school feeding without some corresponding reward; and the scale of the food production needed in a context of increasingly unpredictable climate, by populations who are ill equipped to adapt to increasing climate variability.

Despite the fact that the duration of the current FFE2 project, due to finish at the end of 2020, is known to all project partners, during evaluation interviews, representatives of SDEJTs, the DPEDH and national government expressed dismay at the prospect of the project coming to an end and invariably asked whether there was any possibility it could continue. While this response might be construed as an unwillingness to take responsibility and ownership of a program which has always been destined to be handed over to the GoM and the local schools and communities, it also reflects a hard reality. At this moment in time, in addition to the challenges already mentioned, with international cooperation aid reduced and high levels of national debt, there is no obvious source of funding for a national school feeding program⁵⁷.

At the level of the project itself, the question of sustainability is also complex. On the one hand, the activities planned to increase capacity and hand over a series of functioning activities to be continued by schools, communities and government institutions are largely on track and have been successful. On the other hand, the gaps which will be left once the significant financial contributions and donations of CSB+ by USDA terminate will be very hard to fill in the current context. Initiatives such as the HGSFGs are still at an early stage and, while this model may in the future make a substantial contribution to the food required to feed pupils in 271 schools once each school day, even the most optimistic informants predicted that the current HGSFGs might be able to feed children in a small percentage of the project schools and at best two or three times per week. The HGSFGs may in the future, with good management, generate enough income from the sale of excess crops to fund the purchase of large quantities of a base food, such as sweet potato or manioc, to replace the corn soya blend currently playing this role, at least for the schools in their immediate vicinity. However, this income will also be needed to pay overheads and possibly salaries. It is clear that, while the HGSFGs may be part of the solution, they will not constitute the whole solution.

Likewise, the literacy program, which has been successful and achieved promising results within the limited scope of the initial two- year intervention, is not likely either to be robust enough to be taken over by the MINEDH by the end of 2020, nor to be sustainable beyond that period if the financial gap left by the closure of the project is not filled, so that the reading coaches can continue to be employed and the other costs of the program supported. Despite definite achievements and advances made, other aspects of the project, such as the food preparation by volunteers also appear precarious: given the number of feeding days currently lost through the absence of cooks, one can only speculate how many of the existing volunteer cooks would continue to volunteer regularly, without the regular small incentives and other support the project provides.

⁵⁷ Mozambique was obliged to borrow from the IMF to fund the response to the recent cyclones; the IMF did not donate funds for this purpose, given the ongoing stalemate surrounding the illegal debt crisis (Hanlon, 2019). Furthermore, the government's ability to respond to the recent Cyclone Idai was a reminder of the limited capacity of the GoM to respond to the urgent needs of its citizens.

The picture is therefore complex. On the one hand, the project is proving successful in meeting its targets and performing the activities as planned and is producing visible and needed change. On the other hand, this in itself no longer appears to be sufficient to achieve sustainability beyond the end of the current project contract. Given this reality, two scenarios must be considered: an extension to the current project with an even greater emphasis on sustainability, which from the beginning works with the future implementers to ensure a staggered handover, including identifying future funding sources and a costed business model able to be self-sustaining ; or an exit plan which uses the remaining sixteen months of the project to work closely with the future implementers of all components to establish priorities and determine what can realistically be continued and put plans in place to achieve this, working within greatly reduced financial parameters.

Conclusions

Internal validity (did the project do what it said it would do?)

The midterm evaluation revealed that the majority of activities planned have been implemented as intended, often in the face of considerable obstacles, and most midterm targets have been met or surpassed, with a number of final targets already met a midterm. This is a testimony to the hard work and organizational ability of a large number of staff and leaders from Planet Aid, ADPP Mozambique and the FFE2 project. In the relatively small percentage of cases where targets have not been met, the report analyzes why and, in most cases, suggests that these are likely to be met by the end point. The one exception to this is the target concerning the percentage of school children meeting the literacy benchmark; it may be that the target set was just too ambitious to be achieved within such a short time, given the low baseline levels. Implementing literacy programs requires a wholesale change of culture and practices within a school system; such systemic changes are possible, but they take time to embed.

Like previous evaluations of FFE1 and FFE2, the current evaluation found that, unlike many projects operating in the extremely challenging context of the Mozambican education system, FFE2 is largely managing to do what it set out to do within the planned timeframe. In terms of targets and indicators then, most of the boxes placed at the beginning of each section of the Results chapter of this report make for positive reading. However, as the external evaluator pointed out to project staff and leadership at the stakeholder review workshop, activities which concern human lives and children's education cannot be evaluated only on the basis of activities, targets and indicators: it is important that project staff and leaders (and evaluators) maintain and promote awareness of the reasons why given activities are conducted, what the positive change they are intended to bring and how they are experienced by all those involved in their implementation, including but not limited to the beneficiaries. At times, the fieldwork revealed a preoccupation with numerical targets, sometimes obscuring the question of

whether these are translating into positive change (this was commented on by the project leader at the final validation workshop). A project such as this, which is well organized enough and sufficiently embedded to be meeting the majority of its targets should be capable of reflecting on whether the activities are making a difference and whether they are the right things to do, whether they could be done differently and what can be learned from the experience.

Initial impact (has it made a difference?)

This report finds that many of the activities presented are beginning to make a difference: teachers are beginning to use improved methods in their teaching; students are beginning to read better; community volunteers in the HGSFGs are beginning to apply new methods to produce large amounts of food; thanks to this produce, schools are beginning to diversify the menus consumed by students; school water committees are beginning to take responsibility for maintenance of their water systems. However, in all these cases, more time is needed for these initial changes to be considered embedded and irreversible. Demonstrating a behavior in which one has recently been trained does not constitute changing the underlying system of reflexes so that behavioral change takes place. Phenomena such as classroom practice, literacy, school culture, agricultural practices and community decision making are all complex and deeply culturally embedded, so that changes occur incrementally over time and rarely in a consistent, linear manner. Furthermore, formal training may transmit information effectively; however, there are further steps necessary for this information to become knowledge or skills, which are a prerequisite to real capacity increase (and behavior change). This is all the more necessary in a context such as Mozambique, where levels of poverty and disempowerment and poor-quality education constitute systemic barriers to change.

The overall finding of the midterm evaluation, then, is that the project interventions are beginning to make a difference but that more time is needed in order for this initial change to be consolidated and integrated into behavioral changes. The project has 14 months still to run and this time will be helpful in consolidating the initial changes achieved. However, in order to achieve long-term, truly sustainable change, it is probable that both the project activities and the technical and expert support the project is providing to those implementing them would need to be sustained over a longer period.

Strategic relevance regarding effectiveness, efficiency, impact and sustainability (were these the right things to do? what can be learned?)

The project applies many of the principles of good international development practice, such as: working in partnership with both national and local government and local communities at every stage, building capacity rather than just “doing”, building commitment from beneficiaries and partners through working together rather than by paying large per diems and other incentives. It is providing services and skills which are highly relevant

to both the beneficiary population and the country as a whole. It has a good record and robust processes in place in terms of efficiency and value-for-money and uses the resources donated by the taxpayers of a partner country wisely and responsibly. It is beginning to produce impact, in the form of positive change, at national, district, school and individual level; this change cannot yet be considered embedded, however, and requires consolidation. Certain aspects of the project activities, such as school feeding, have become part of daily life and would be badly missed were they to cease: the current cohort of primary school children in the project schools has never known school without the daily soya porridge. And yet, major questions remain about the sustainability of these achievements in the current context in Mozambique.

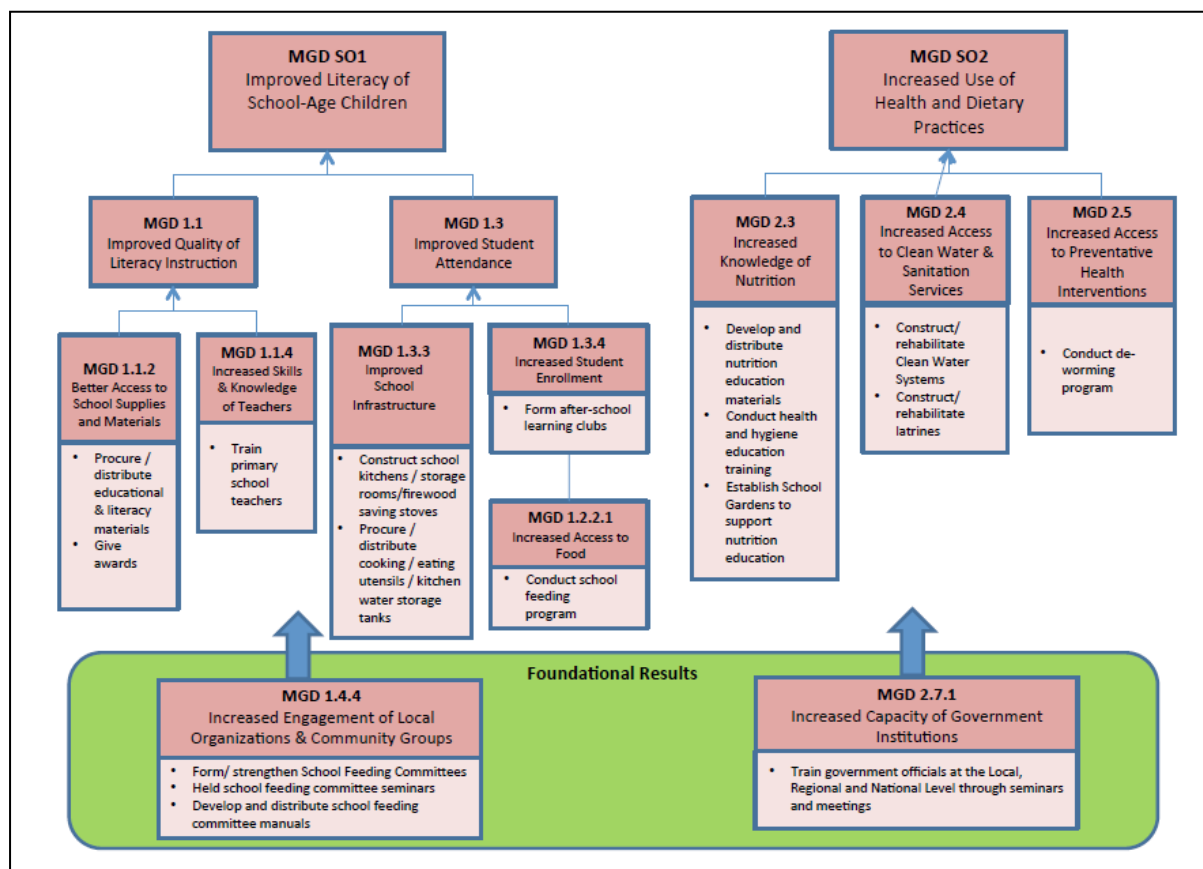
The evaluation revealed significant questions about sustainability, in terms of both government- and community-/school-level ability to sustain the activities after the end of the project. Other school feeding projects, such as the model followed by JAM (see page 129 - 30 above) may offer lessons to help the current FFE2 project be more sustainable. However, the bottom line is that school feeding (and the other activities developed by the project) requires resources, whether these are generated by project activities or provided by government, development partners, communities or families, or a combination of these. There is, as yet, no clear indication of where the necessary resources will come from to allow the project activities to continue.

External evaluator's concluding remarks

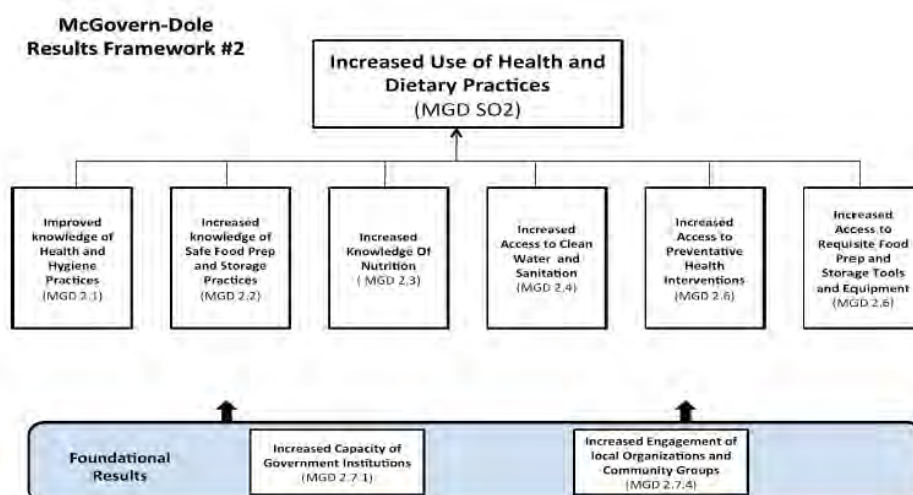
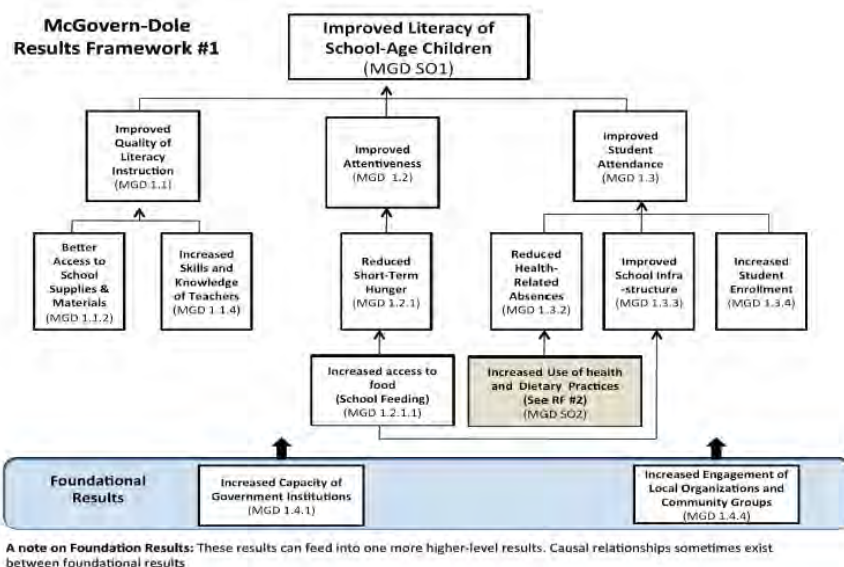
During the evaluation field visits, the external evaluator was very aware of the approaching close of the current phase of the project and that many of the achievements and positive changes observed may not be able to be maintained. In particular, the prospect that school children who are currently receiving food in school and benefiting from this, in terms of their health and education, might soon no longer receive food, is sobering. The context of Mozambique has deteriorated considerably since the external evaluator was first involved with the baseline of FFE1 in 2012. If that had not been the case, this report might be concluding on a more optimistic note, celebrating the achievements and noting their prospects of long-term sustainability. However, in the current context, it must end by celebrating the achievements and noting that, in the current reality, more time is needed for these to be sustained, although there is currently little prospect of the necessary resources being made available to allow this. The external evaluator considers that this is a valuable project, which is being well run and is largely achieving its objectives and beginning to produce the hoped-for changes. In her professional judgement, the FFE2 project merits being prolonged in order to have the opportunity to capitalize on all the investment made so far, both human and material.

Annexes

Annex 1: Results Framework of the Food for Education Project in Mozambique, capturing the Theory of Change upon which the project is based



Annex 2: McGovern-Dole Results Frameworks



Annex 3: Technical appendix 1

FFE: Impact Evaluation

Analysis including substitute students

NFER (Jose Liht)

Contents

1	EGRA Numerical Variables	5
1.1	EGRA_ST1: Oral vocabulary	5
1.1.1	Contrast 1: FFE + lit (all) vs Comparison (all)	5
1.1.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	8
1.1.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	10
1.1.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	12
1.1.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	14
1.2	EGRA_ST2: Oral comprehension	16
1.2.1	Contrast 1: FFE + lit (all) vs Comparison (all)	16
1.2.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	18
1.2.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	20
1.2.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	22
1.2.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	24
1.3	EGRA_ST3: Phonological awareness	26
1.3.1	Contrast 1: FFE + lit (all) vs Comparison (all)	26
1.3.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	28
1.3.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	30
1.3.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	32
1.3.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	34
1.4	EGRA_ST4: Concepts of print	36
1.4.1	Contrast 1: FFE + lit (all) vs Comparison (all)	36
1.4.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	38
1.4.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	40
1.4.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	42
1.4.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	44
1.5	EGRA_ST5: Knowledge of letter sounds	46
1.5.1	Contrast 1: FFE + lit (all) vs Comparison (all)	46
1.5.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	48
1.5.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	50

1.5.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	52
1.5.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	54
1.6	EGRA_ST6: Knowledge of syllables	56
1.6.1	Contrast 1: FFE + lit (all) vs Comparison (all)	56
1.6.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	58
1.6.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	60
1.6.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	62
1.6.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	64
1.7	EGRA_ST7: Reading of words	66
1.7.1	Contrast 1: FFE + lit (all) vs Comparison (all)	66
1.7.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	68
1.7.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	70
1.7.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	72
1.7.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	74
1.8	EGRA_ST8_1A: Reading fluency	76
1.8.1	Contrast 1: FFE + lit (all) vs Comparison (all)	76
1.8.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	78
1.8.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	80
1.8.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	82
1.8.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	84
1.9	EGRA_ST8_1AS: Comparative reading fluency	86
1.9.1	Contrast 1: FFE + lit (all) vs Comparison (all)	86
1.9.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	88
1.9.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	90
1.9.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	92
1.9.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	94
1.10	EGRA_ST8_1B: Overall reading ability	96
1.10.1	Contrast 1: FFE + lit (all) vs Comparison (all)	96
1.10.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	97
1.10.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	99
1.10.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	101
1.10.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	103
1.11	EGRA_ST8_1BS: Comparative overall reading ability	105
1.11.1	Contrast 1: FFE + lit (all) vs Comparison (all)	105
1.11.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	107
1.11.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	109

1.11.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	111
1.11.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	113
1.12	EGRA_ST8_2: Comprehension of reading	115
1.12.1	Contrast 1: FFE + lit (all) vs Comparison (all)	115
1.12.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	118
1.12.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	120
1.12.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	122
1.12.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	124
1.13	EGRA_ST9_2A: Total writing	126
1.13.1	Contrast 1: FFE + lit (all) vs Comparison (all)	126
1.13.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	128
1.13.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	130
1.13.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	132
1.13.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	134
1.14	EGRA_ST9_2B: Writing words	136
1.14.1	Contrast 1: FFE + lit (all) vs Comparison (all)	136
1.14.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	137
1.14.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	139
1.14.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	141
1.14.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	143
1.15	EGRA_ST9_2C: Writing phrase	145
1.15.1	Contrast 1: FFE + lit (all) vs Comparison (all)	145
1.15.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	147
1.15.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	149
1.15.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	151
1.15.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	153
2	EGRA Categorical Variables	155
2.1	EGRA_ST9_1A: Writing first name	155
2.1.1	Contrast 1: FFE + lit (all) vs Comparison (all)	155
2.1.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	157
2.1.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	158
2.1.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	159
2.1.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	160
2.2	EGRA_ST9_1B: Writing family name	161
2.2.1	Contrast 1: FFE + lit (all) vs Comparison (all)	161
2.2.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	163

2.2.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	164
2.2.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	165
2.2.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	166
2.3	EGRA_ST9_1C: Writing name overall	167
2.3.1	Contrast 1: FFE + lit (all) vs Comparison (all)	167
2.3.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	169
2.3.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	170
2.3.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	171
2.3.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	172
3	Literacy Benchmark	173
3.1	Literacy Benchmark: Boys and Girls	173
3.1.1	Contrast 1: FFE + lit (all) vs Comparison (all)	173
3.1.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	174
3.1.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	175
3.1.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	176
3.1.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	177
3.2	Literacy Benchmark: Boys only	178
3.2.1	Contrast 1: FFE + lit (all) vs Comparison (all)	178
3.2.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	179
3.2.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	180
3.2.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	181
3.2.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	182
3.3	Literacy Benchmark: Girls only	183
3.3.1	Contrast 1: FFE + lit (all) vs Comparison (all)	183
3.3.2	Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)	184
3.3.3	Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)	185
3.3.4	Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)	186
3.3.5	Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)	187
4	BMI	188
4.1	Height	188
4.2	Weight	191
4.3	BMI Z Score	193

5	Questions regarding hunger, attention and hygiene	195
5.1	HUNGER: Hungry in school	195
5.2	HUNGERNOW: Hungry now	196
5.3	ATTENTION: Attention problems	197
5.4	TOILET: Wash after toilet	198
5.5	EATING: Wash before eating	199
6	Teacher questions regarding hunger, attention and difficulty in learning	201
6.1	Students are hungry	201
6.2	Students have attention problems	202
6.3	Students have learning problems	203

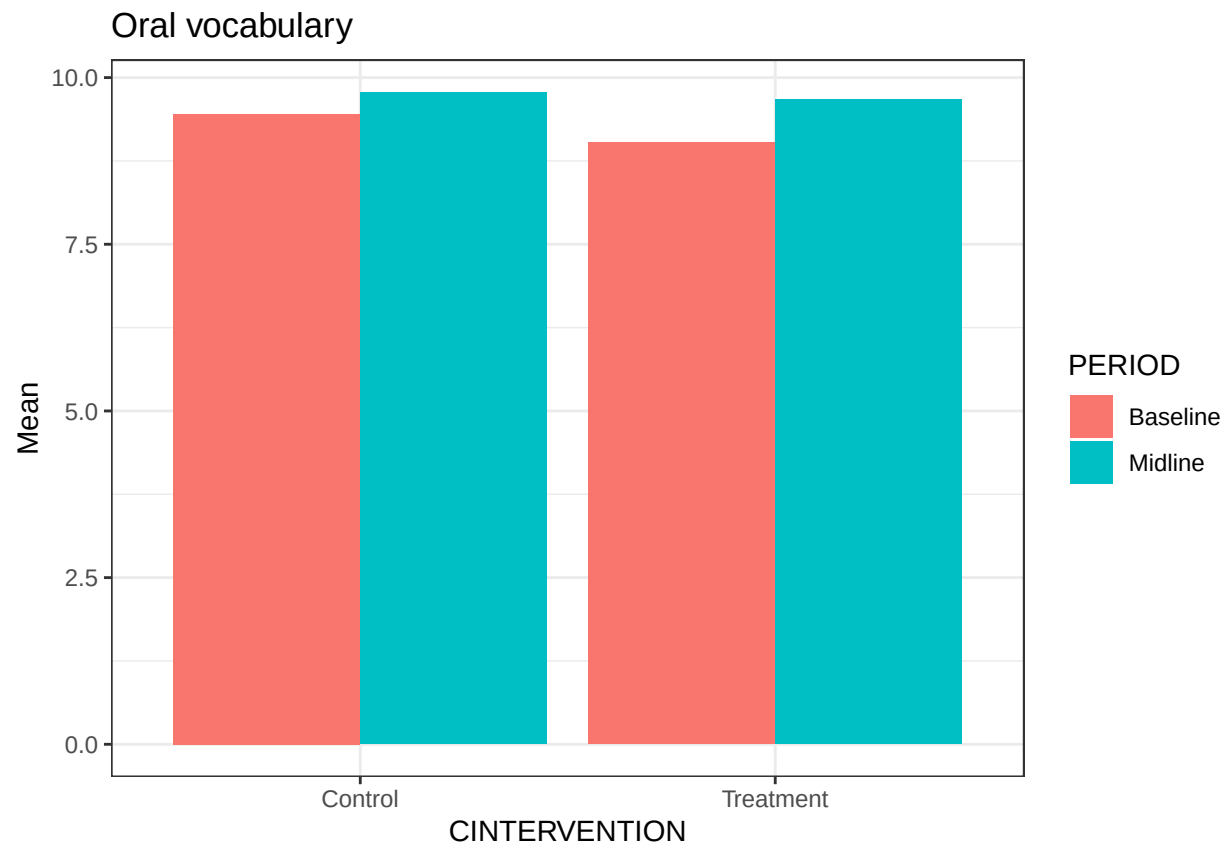
1 EGRA Numerical Variables

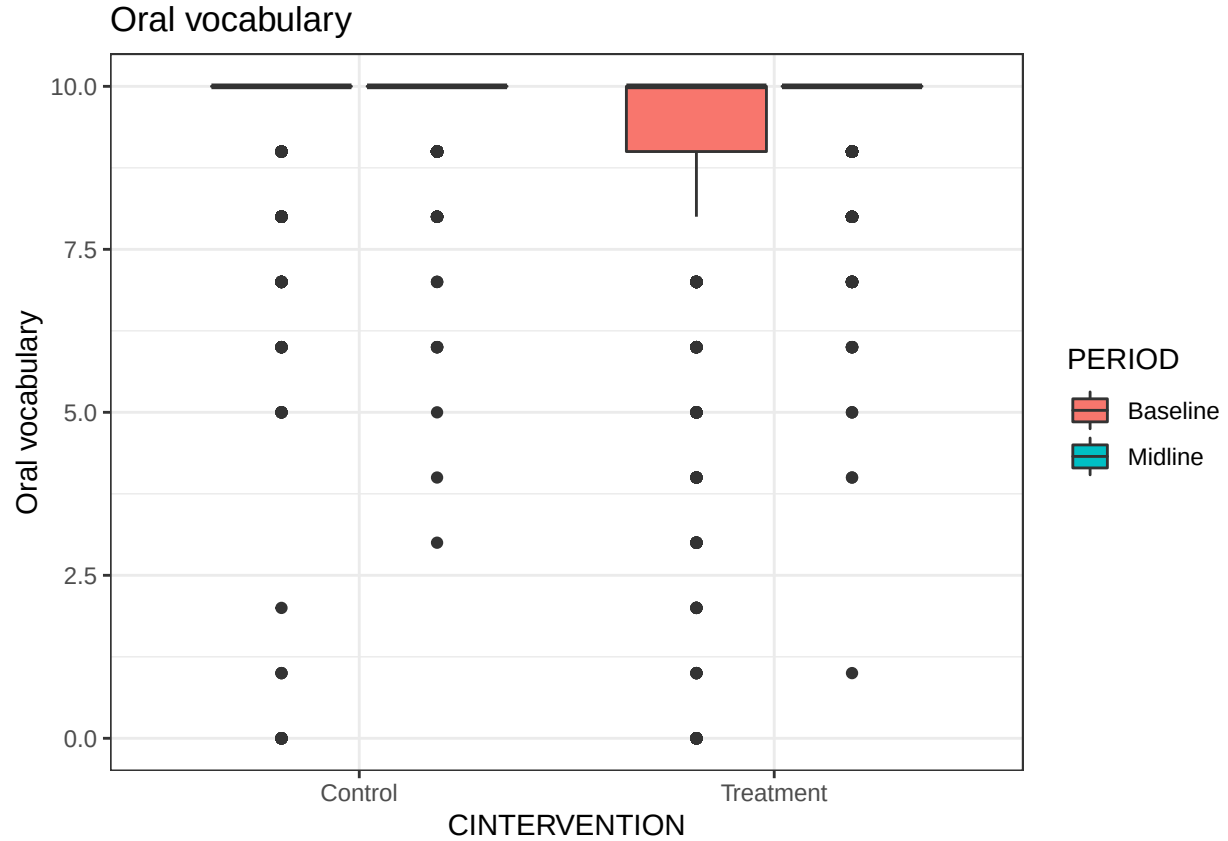
1.1 EGRA_ST1: Oral vocabulary

1.1.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 1: Oral vocabulary

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	9.450	1.636	1136	0	10	9.779	0.621	1081	3	10
Treatment	9.026	1.967	1578	0	10	9.672	0.747	1808	1	10



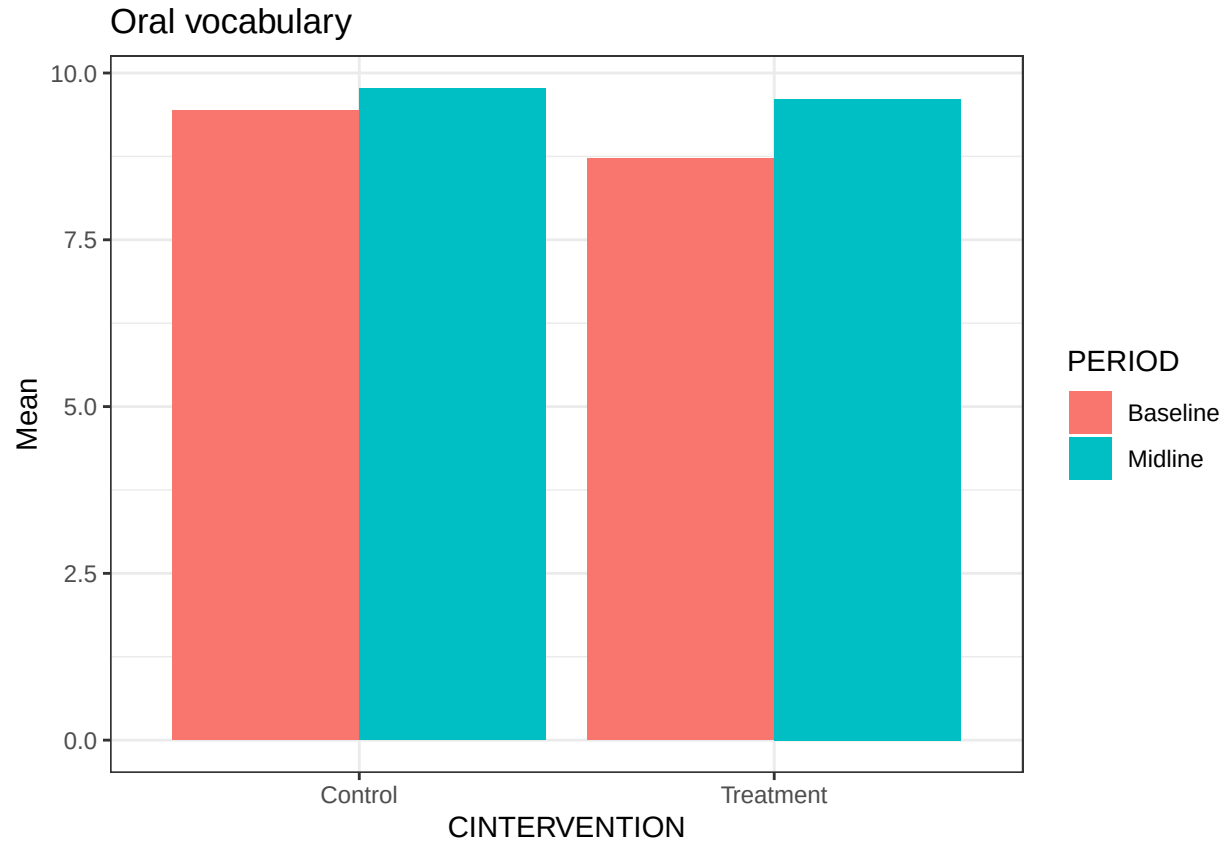


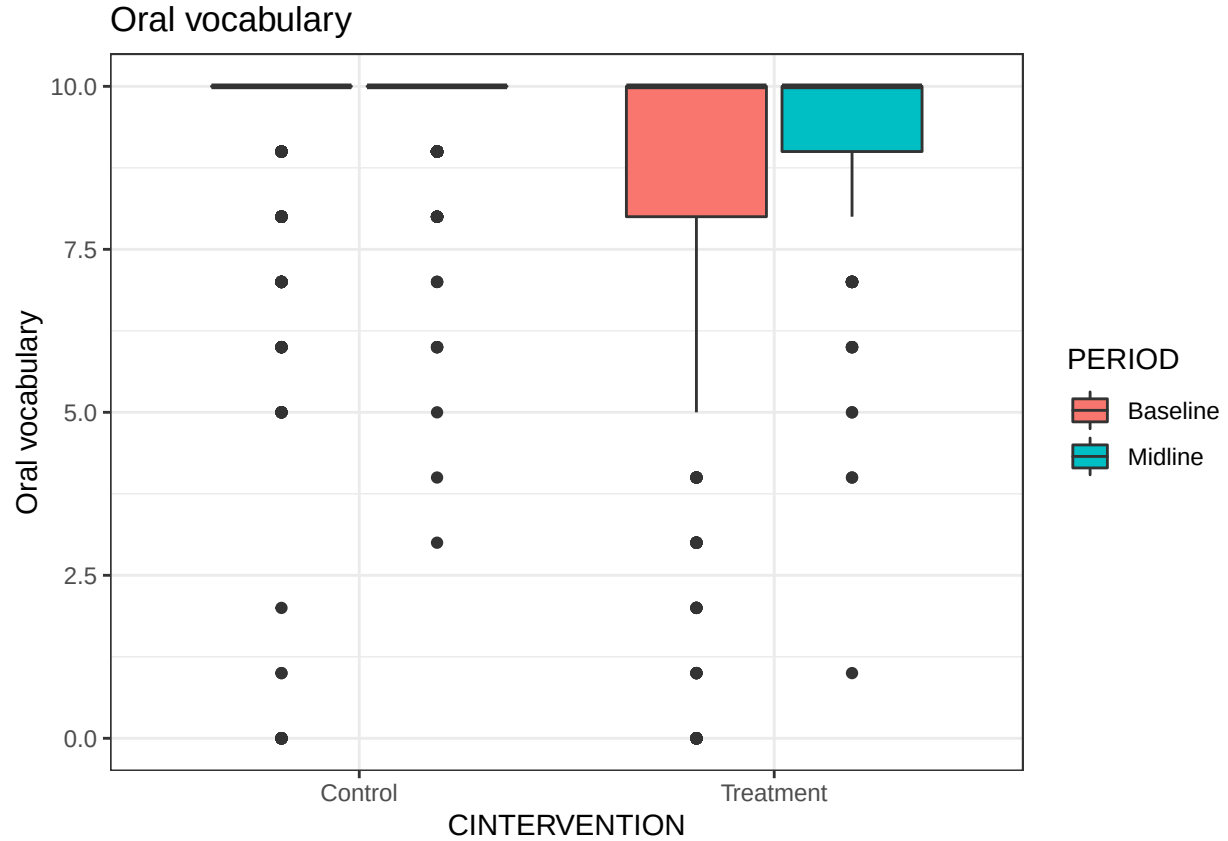
As shown in the table above, for the the Oral vocabulary EGRA subtask, the mean for the Control (Comparison (all)) condition at baseline was 9.449824 (SD = 1.636355) and the mean for the Treatment (FFE + lit (all)) condition at baseline was 9.025982 (SD = 1.967384). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at baseline was thus of 0.4238417 points. The p-value for this difference was 0.0003841511. The mean for the Control (Comparison (all)) condition at midline was 9.778908 (SD = 0.6207516) and the mean for the Treatment (FFE + lit (all)) condition at midline was 9.672013 (SD = 0.7470673). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at midline was thus of 0.1068951 points. The p-value for this difference was 0.00774526. The change from the baseline to the midline of 0.3290845 points for the Control (Comparison (all)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (all)) can be assessed against. The change for the Treatment (FFE + lit (all)) from the baseline to the midline was of 0.646031 points. Consequently, the change for the Control (Comparison (all)) relative to the change for the Treatment (FFE + lit (all)) condition was of 0.3169465 points. The p-value for this difference was 0.007771889. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (all)) when compared to the Control (Comparison (all)) condition. This provides evidence that the Treatment (FFE + lit (all)) (or some other unobserved process) impacted on the Oral vocabulary EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

1.1.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 2: Oral vocabulary

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	9.441	1.665	1047	0	10	9.773	0.629	972	3	10
Treatment	8.721	2.215	1040	0	10	9.611	0.816	1047	1	10



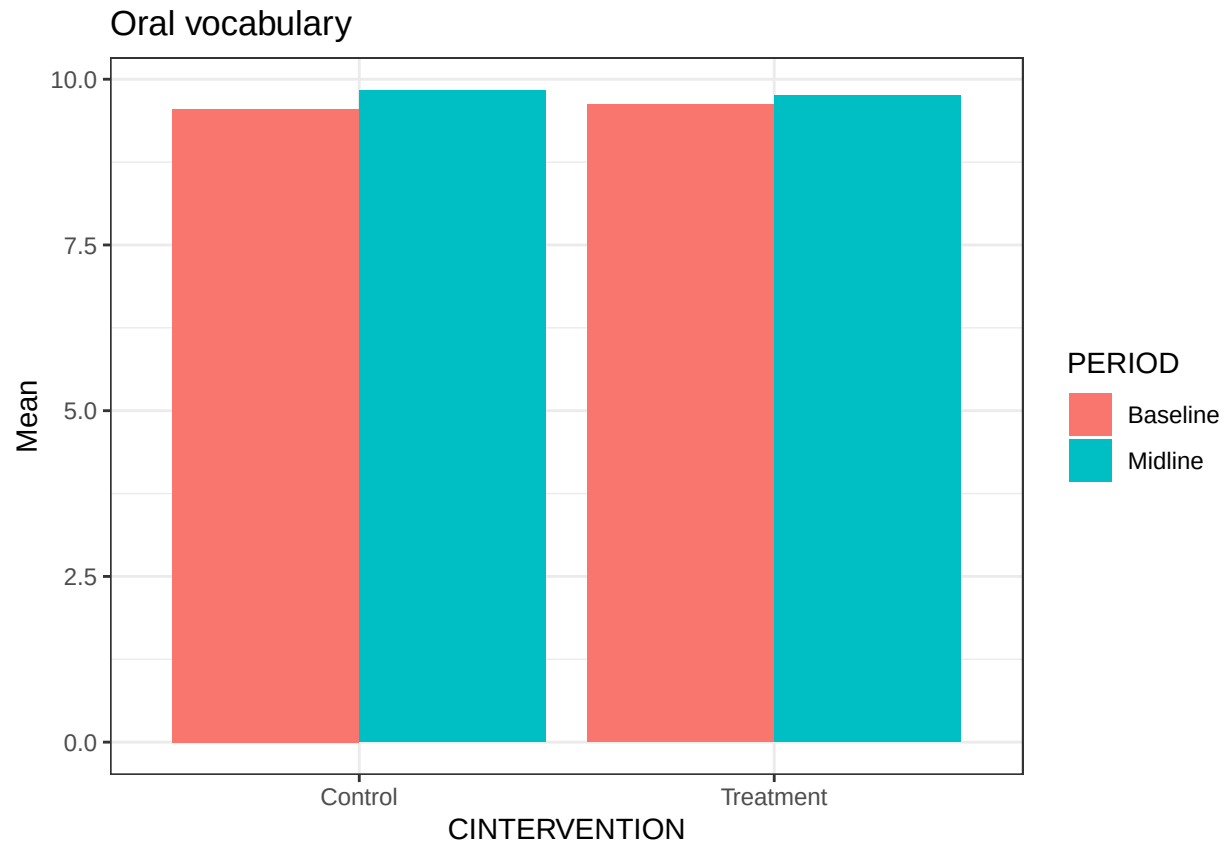


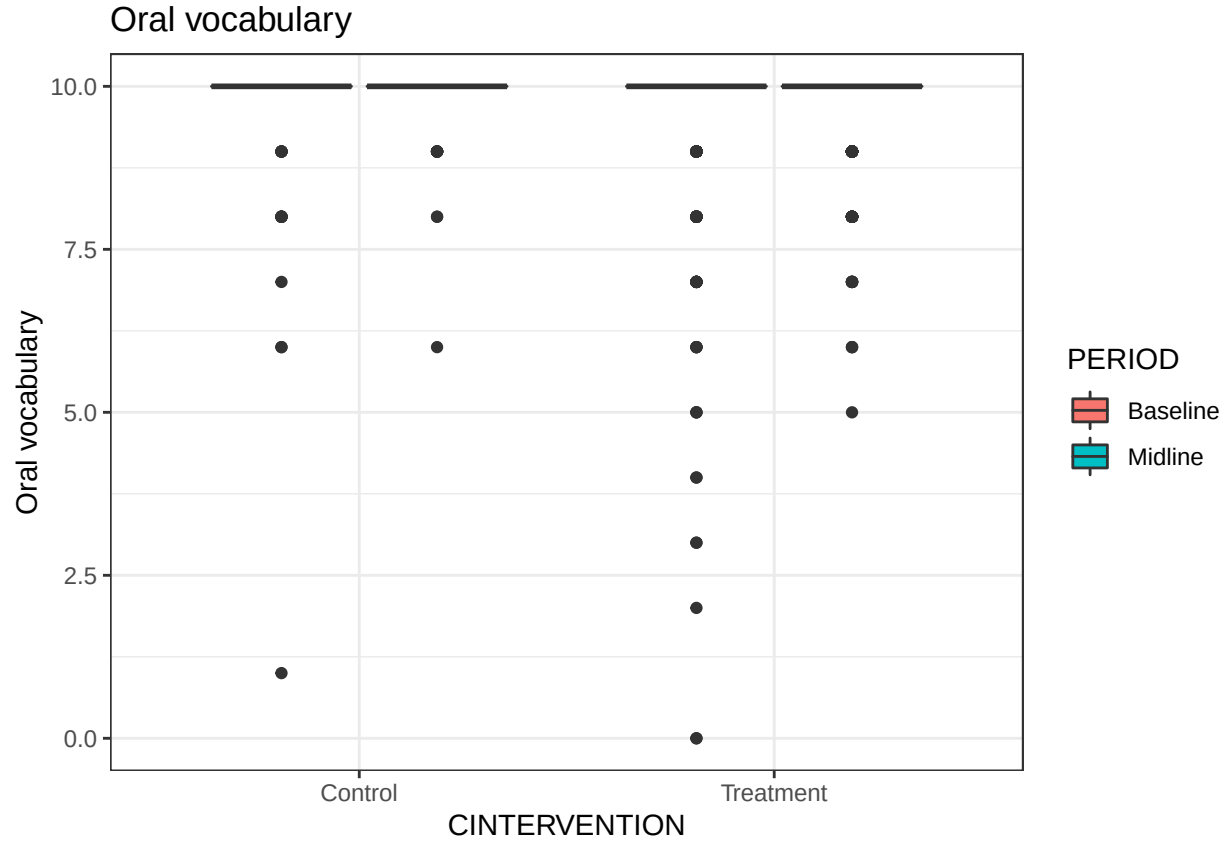
As shown in the table above, for the the Oral vocabulary EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at baseline was 9.441261 (SD = 1.665112) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 8.721154 (SD = 2.214906). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.7201069 points. The p-value for this difference was 1.973785e-06. The mean for the Control (Comparison (Portuguese)) condition at midline was 9.772634 (SD = 0.6294784) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 9.61127 (SD = 0.8161767). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.1613634 points. The p-value for this difference was 0.0008852814. The change from the baseline to the midline of 0.331373 points for the Control (Comparison (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 0.8901164 points. Consequently, the change for the Control (Comparison (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 0.5587435 points. The p-value for this difference was 0.0002102456. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Portuguese)) when compared to the Control (Comparison (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Portuguese)) (or some other unobserved process) impacted on the Oral vocabulary EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

1.1.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 3: Oral vocabulary

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	9.551	1.252	89	1	10	9.835	0.536	109	6	10
Treatment	9.615	1.161	538	0	10	9.756	0.631	761	5	10



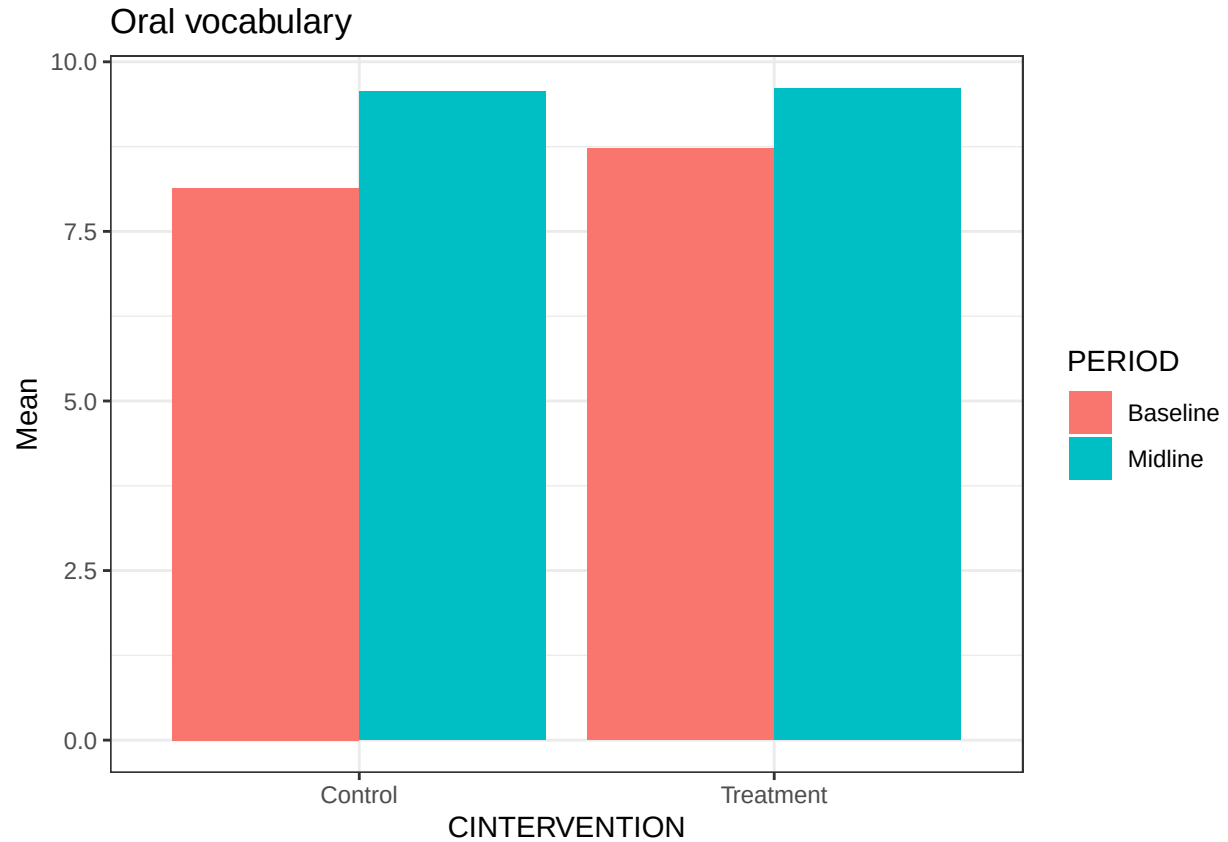


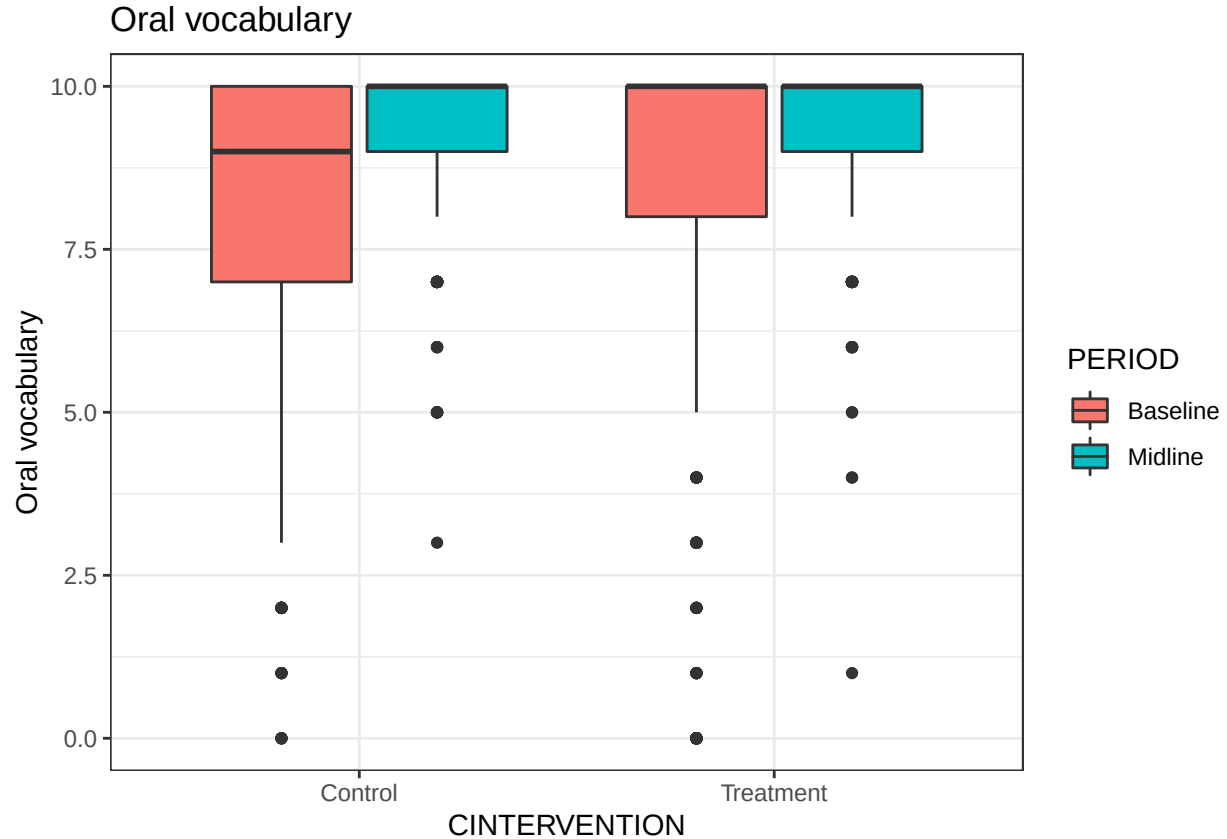
As shown in the table above, for the the Oral vocabulary EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at baseline was 9.550562 (SD = 1.252373) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 9.615242 (SD = 1.160616). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.06467984 points. The p-value for this difference was 0.7338745. The mean for the Control (Comparison (Bilingual)) condition at midline was 9.834862 (SD = 0.5359962) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 9.755585 (SD = 0.6309337). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.07927763 points. The p-value for this difference was 0.1083091. The change from the baseline to the midline of 0.2843006 points for the Control (Comparison (Bilingual)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 0.1403431 points. Consequently, the change for the Control (Comparison (Bilingual)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of -0.1439575 points. The p-value for this difference was 0.4222522. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Oral vocabulary EGRA subtask across the Control (Comparison (Bilingual)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

1.1.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 4: Oral vocabulary

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	8.141	2.212	1007	0	10	9.558	0.868	952	3	10
Treatment	8.721	2.215	1040	0	10	9.611	0.816	1047	1	10



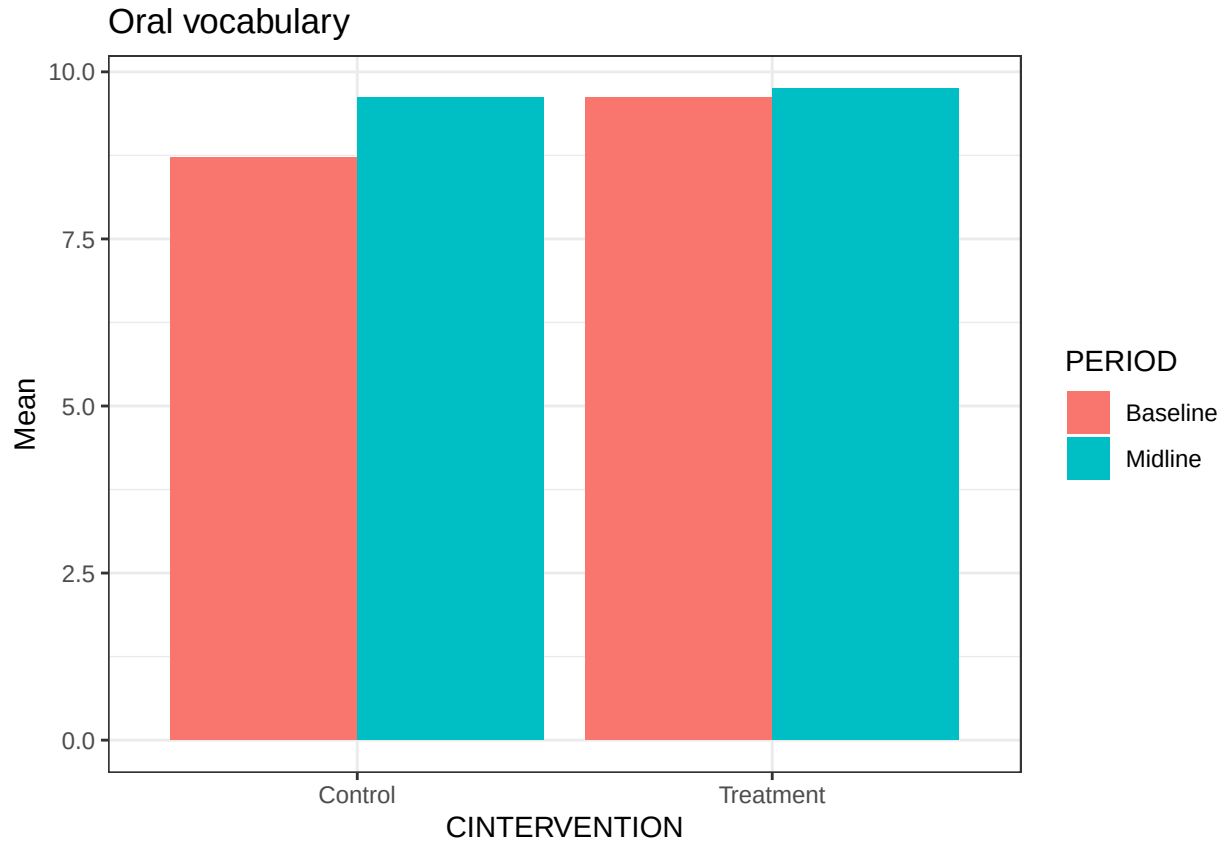


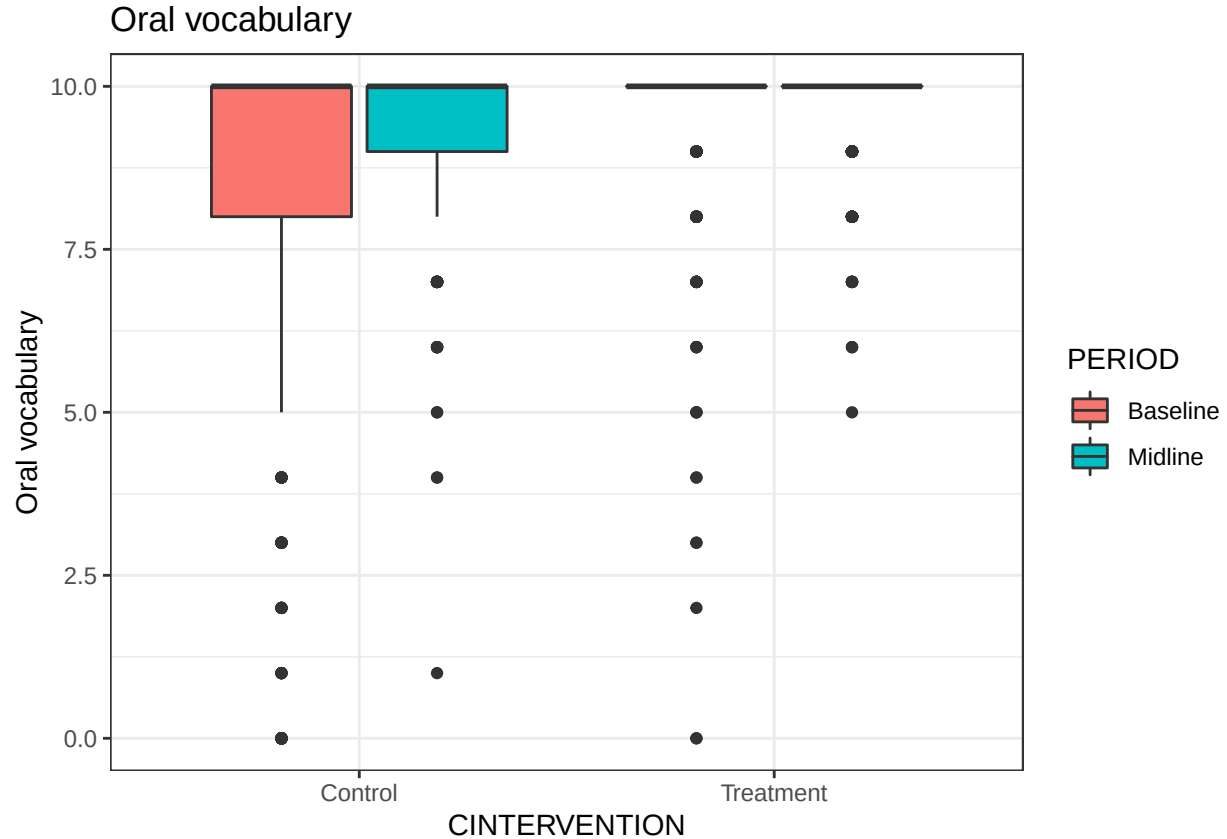
As shown in the table above, for the the Oral vocabulary EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at baseline was 8.141013 (SD = 2.211927) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 8.721154 (SD = 2.214906). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.5801409 points. The p-value for this difference was 0.002241314. The mean for the Control (FFE only (Portuguese)) condition at midline was 9.557773 (SD = 0.8681915) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 9.61127 (SD = 0.8161767). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.05349719 points. The p-value for this difference was 0.3538011. The change from the baseline to the midline of 1.41676 points for the Control (FFE only (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 0.8901164 points. Consequently, the change for the Control (FFE only (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of -0.5266437 points. The p-value for this difference was 0.006323686. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Portuguese)) when compared to the Control (FFE only (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Portuguese)) (or some other unobserved process) impacted on the Oral vocabulary EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed. It should be pointed out that the difference in the rate of change of the Treatment (FFE + lit (Portuguese)) relative to the Control (FFE only (Portuguese)) was negative and thus suggests that students in the Control (FFE only (Portuguese)) schools performed significantly better than those in the Treatment (FFE + lit (Portuguese)) .

1.1.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 5: Oral vocabulary

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	8.721	2.215	1040	0	10	9.611	0.816	1047	1	10
Treatment	9.615	1.161	538	0	10	9.756	0.631	761	5	10





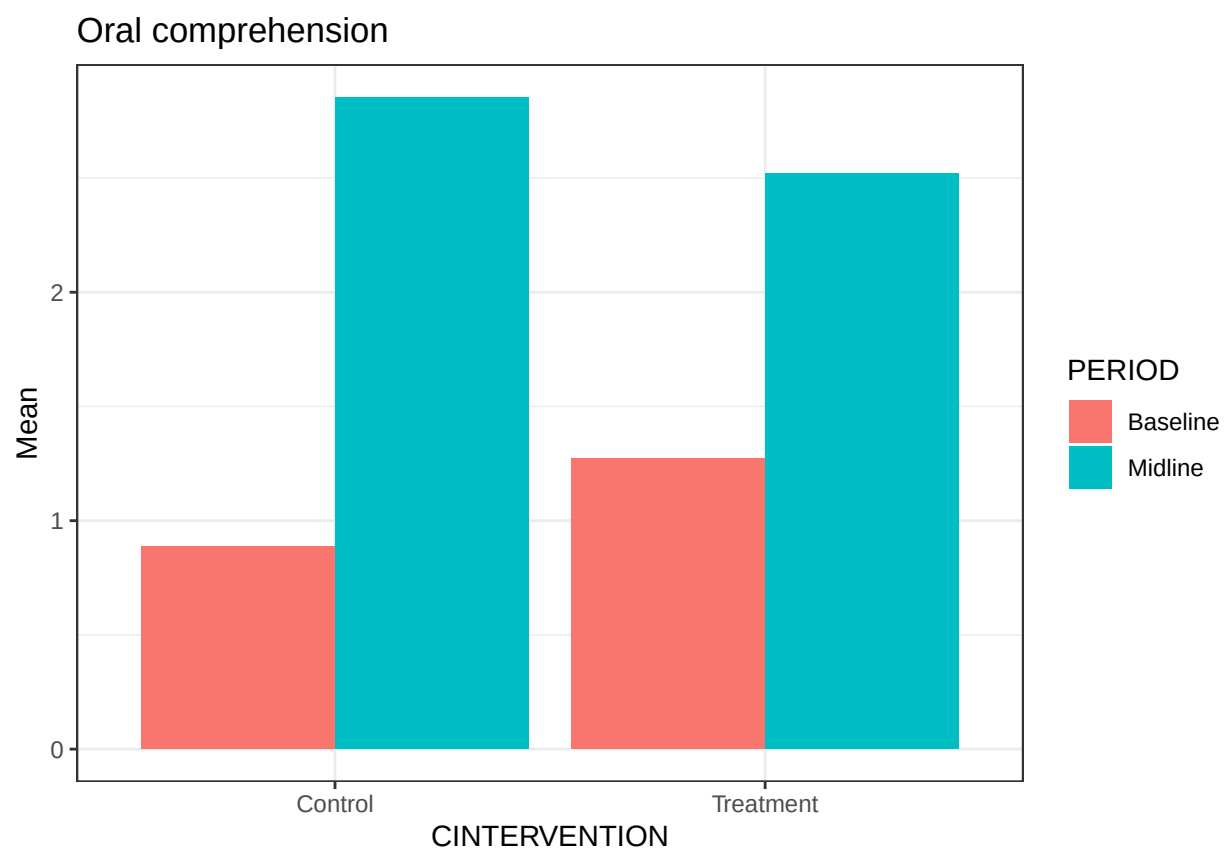
As shown in the table above, for the the Oral vocabulary EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at baseline was 8.721154 (SD = 2.214906) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 9.615242 (SD = 1.160616). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.8940878 points. The p-value for this difference was 3.994351e-09. The mean for the Control (FFE + lit (Portuguese)) condition at midline was 9.61127 (SD = 0.8161767) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 9.755585 (SD = 0.6309337). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.1443145 points. The p-value for this difference was 0.002332906. The change from the baseline to the midline of 0.8901164 points for the Control (FFE + lit (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 0.1403431 points. Consequently, the change for the Control (FFE + lit (Portuguese)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of -0.7497733 points. The p-value for this difference was 3.744629e-07. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Bilingual)) when compared to the Control (FFE + lit (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Bilingual)) (or some other unobserved process) impacted on the Oral vocabulary EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed. It should be pointed out that the difference in the rate of change of the Treatment (FFE + lit (Bilingual)) relative to the Control (FFE + lit (Portuguese)) was negative and thus suggests that students in the Control (FFE + lit (Portuguese)) schools performed significantly better than those in the Treatment (FFE + lit (Bilingual)) .

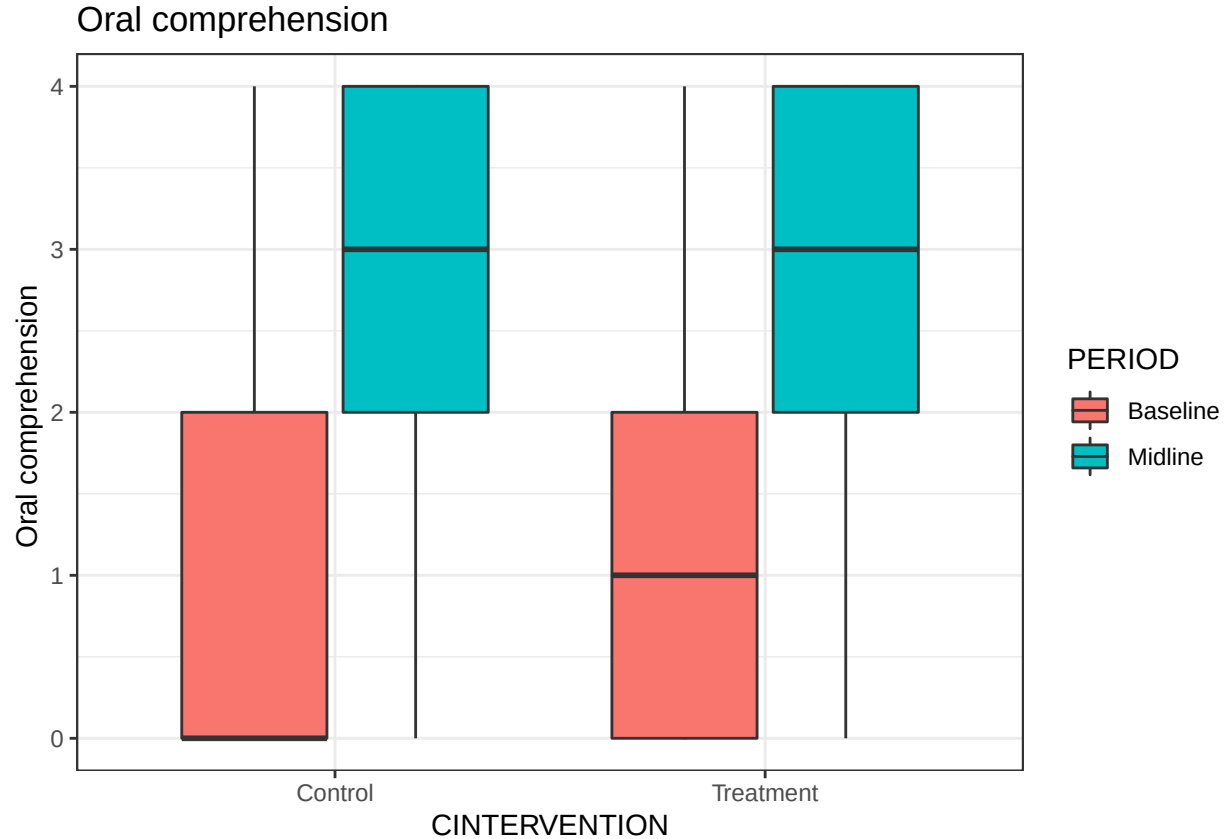
1.2 EGRA_ST2: Oral comprehension

1.2.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 6: Oral comprehension

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.888	1.159	1136	0	4	2.854	1.245	1081	0	4
Treatment	1.271	1.438	1578	0	4	2.521	1.255	1808	0	4



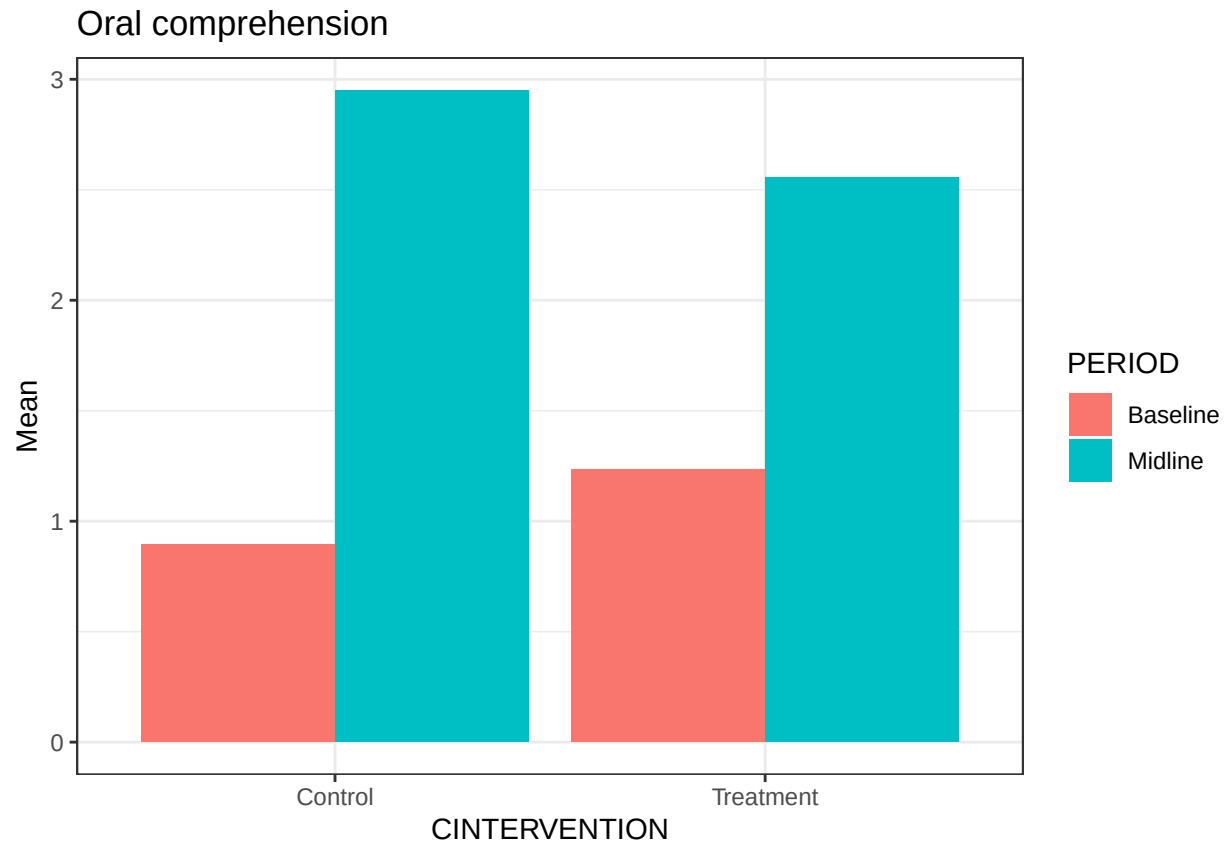


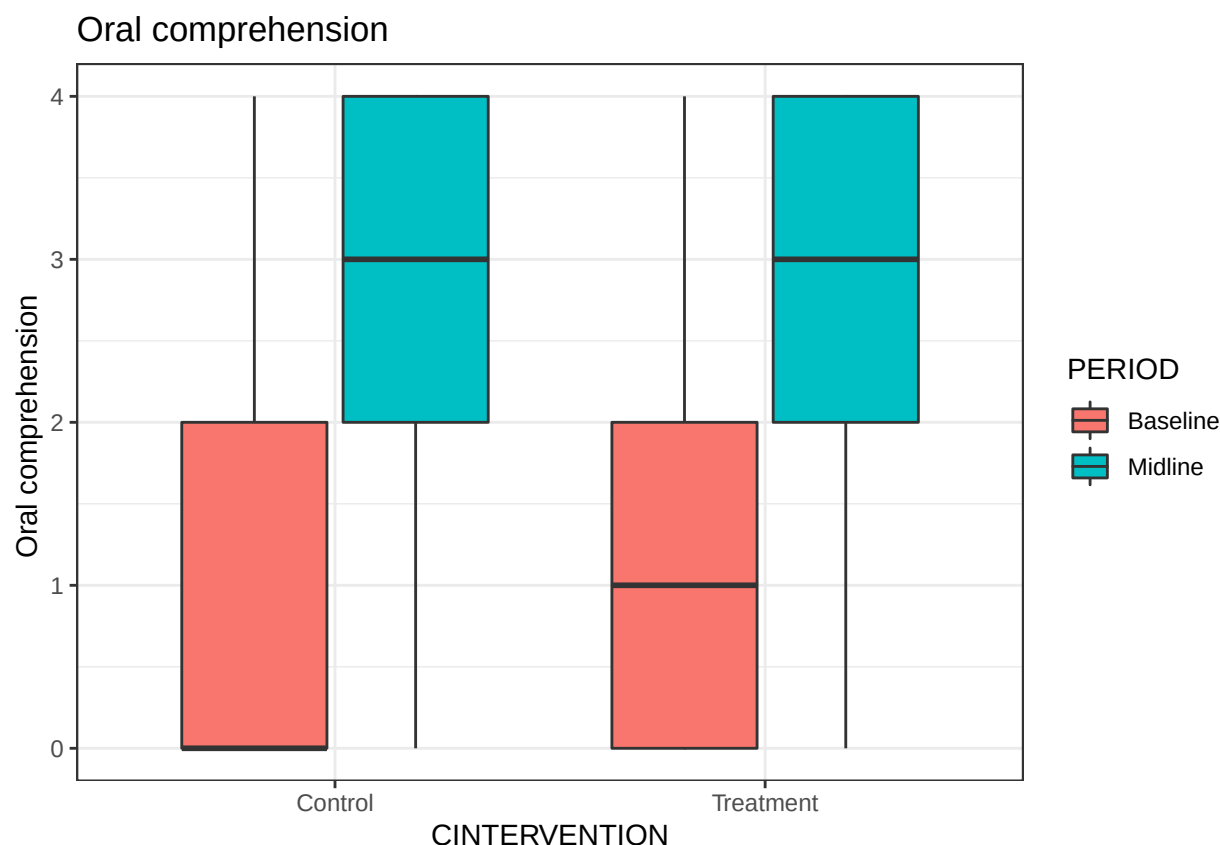
As shown in the table above, for the the Oral comprehension EGRA subtask, the mean for the Control (Comparison (all)) condition at baseline was 0.8882042 (SD = 1.159068) and the mean for the Treatment (FFE + lit (all)) condition at baseline was 1.271229 (SD = 1.437763). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at baseline was thus of 0.3830252 points. The p-value for this difference was 0.001128986. The mean for the Control (Comparison (all)) condition at midline was 2.853839 (SD = 1.244583) and the mean for the Treatment (FFE + lit (all)) condition at midline was 2.521018 (SD = 1.255249). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at midline was thus of 0.3328213 points. The p-value for this difference was 0.0009810953. The change from the baseline to the midline of 1.965635 points for the Control (Comparison (all)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (all)) can be assessed against. The change for the Treatment (FFE + lit (all)) from the baseline to the midline was of 1.249788 points. Consequently, the change for the Control (Comparison (all)) relative to the change for the Treatment (FFE + lit (all)) condition was of -0.7158465 points. The p-value for this difference was 2.260209e-06. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (all)) when compared to the Control (Comparison (all)) condition. This provides evidence that the Treatment (FFE + lit (all)) (or some other unobserved process) impacted on the Oral comprehension EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed. It should be pointed out that the difference in the rate of change of the Treatment (FFE + lit (all)) relative to the Control (Comparison (all)) was negative and thus suggests that students in the Control (Comparison (all)) schools performed significantly better than those in the Treatment (FFE + lit (all)) .

1.2.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 7: Oral comprehension

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.895	1.165	1047	0	4	2.951	1.221	972	0	4
Treatment	1.235	1.426	1040	0	4	2.555	1.309	1047	0	4



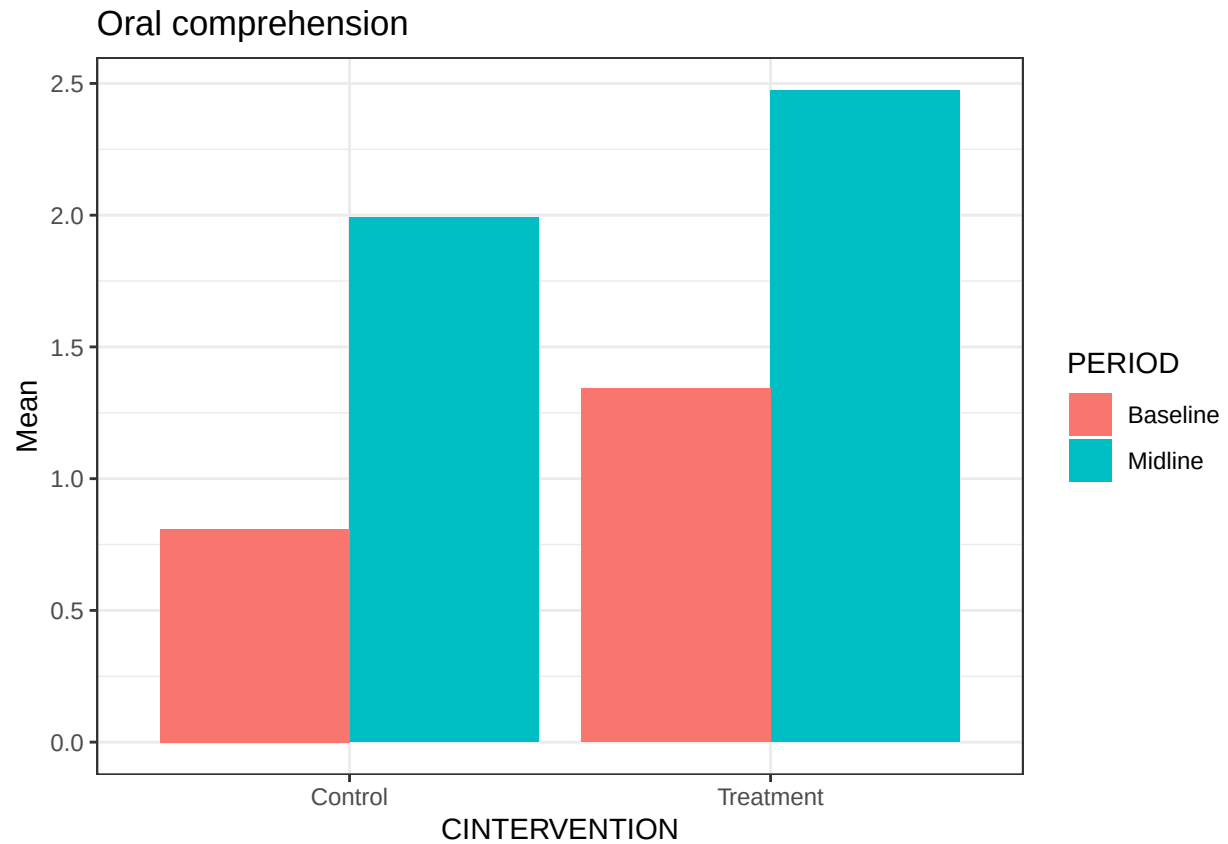


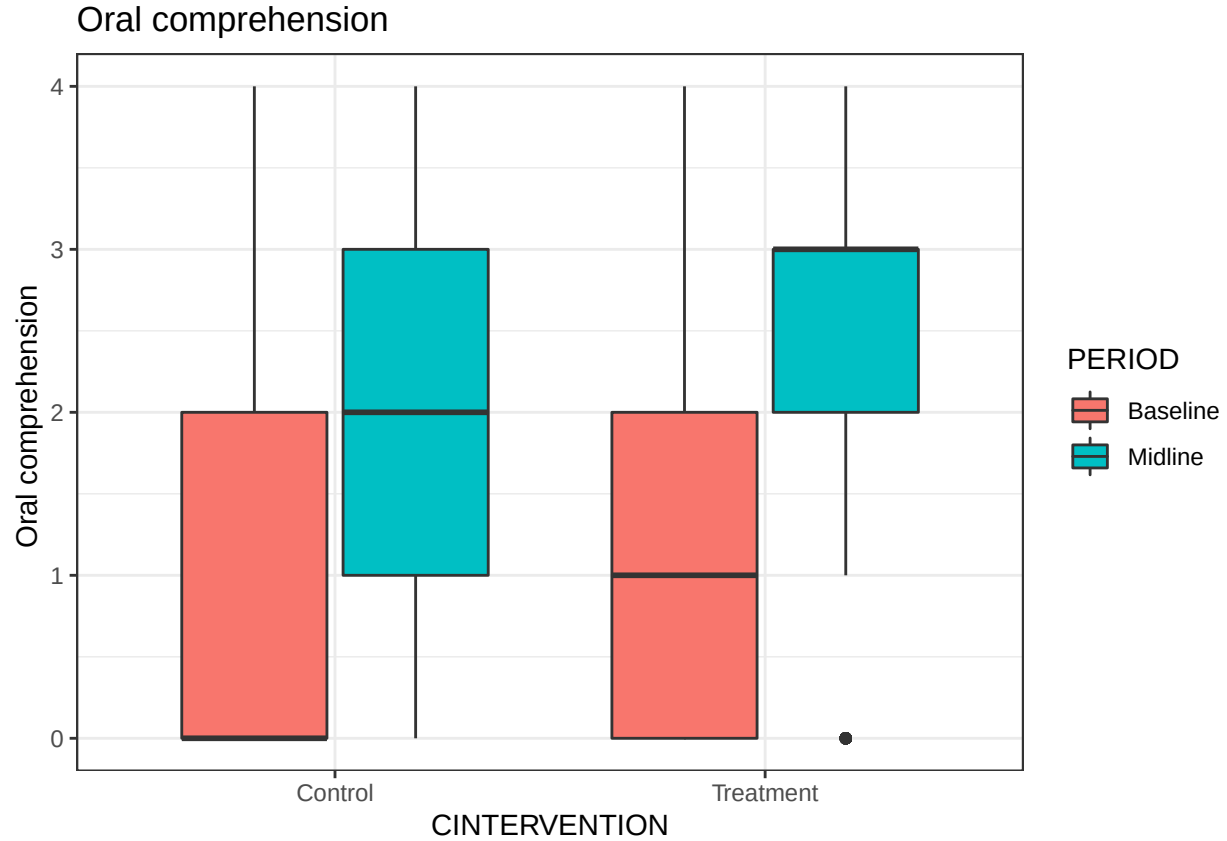
As shown in the table above, for the the Oral comprehension EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at baseline was 0.8949379 (SD = 1.164502) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 1.234615 (SD = 1.425991). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.3396775 points. The p-value for this difference was 0.009005426. The mean for the Control (Comparison (Portuguese)) condition at midline was 2.950617 (SD = 1.22101) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 2.554919 (SD = 1.309106). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.3956985 points. The p-value for this difference was 0.000293822. The change from the baseline to the midline of 2.055679 points for the Control (Comparison (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 1.320303 points. Consequently, the change for the Control (Comparison (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of -0.7353759 points. The p-value for this difference was 5.591441e-06. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Portuguese)) when compared to the Control (Comparison (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Portuguese)) (or some other unobserved process) impacted on the Oral comprehension EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed. It should be pointed out that the difference in the rate of change of the Treatment (FFE + lit (Portuguese)) relative to the Control (Comparison (Portuguese)) was negative and thus suggests that students in the Control (Comparison (Portuguese)) schools performed significantly better than those in the Treatment (FFE + lit (Portuguese)) .

1.2.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 8: Oral comprehension

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.809	1.096	89	0	4	1.991	1.118	109	0	4
Treatment	1.342	1.459	538	0	4	2.474	1.176	761	0	4



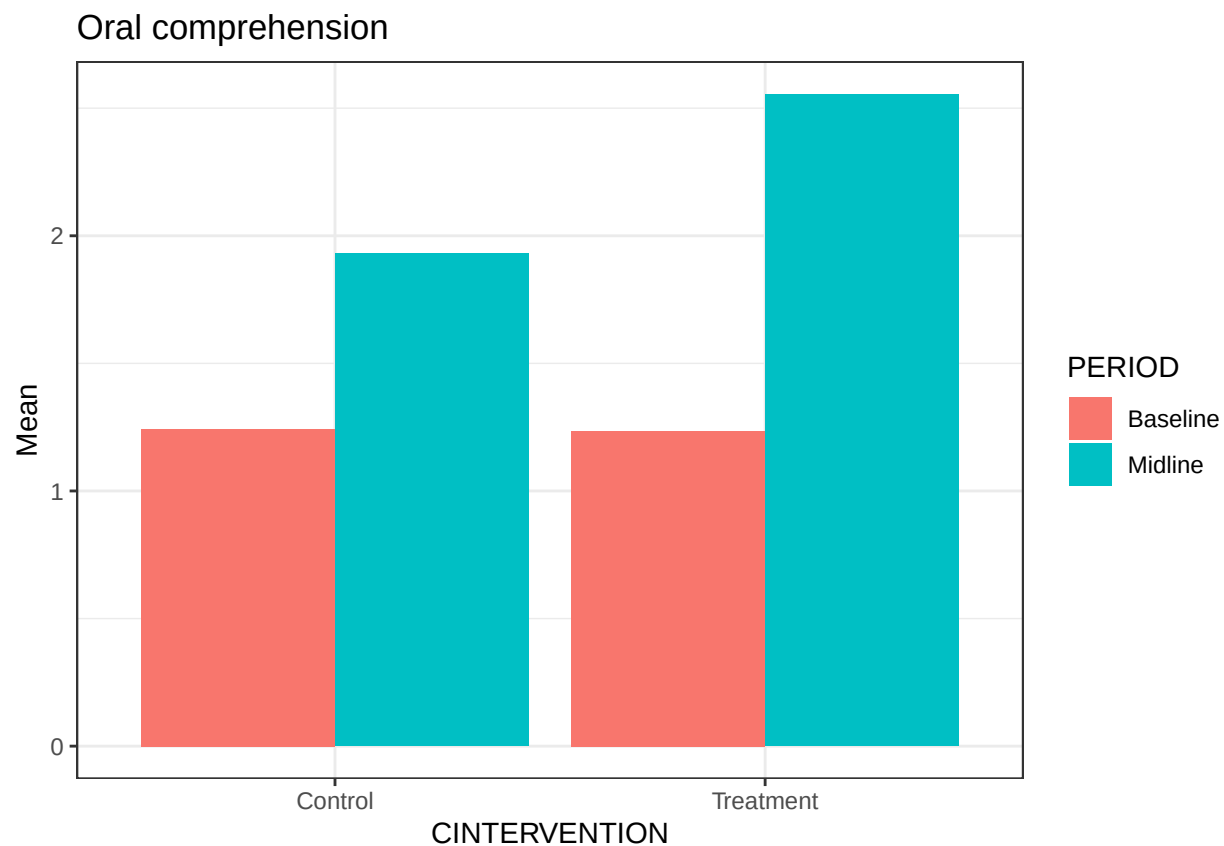


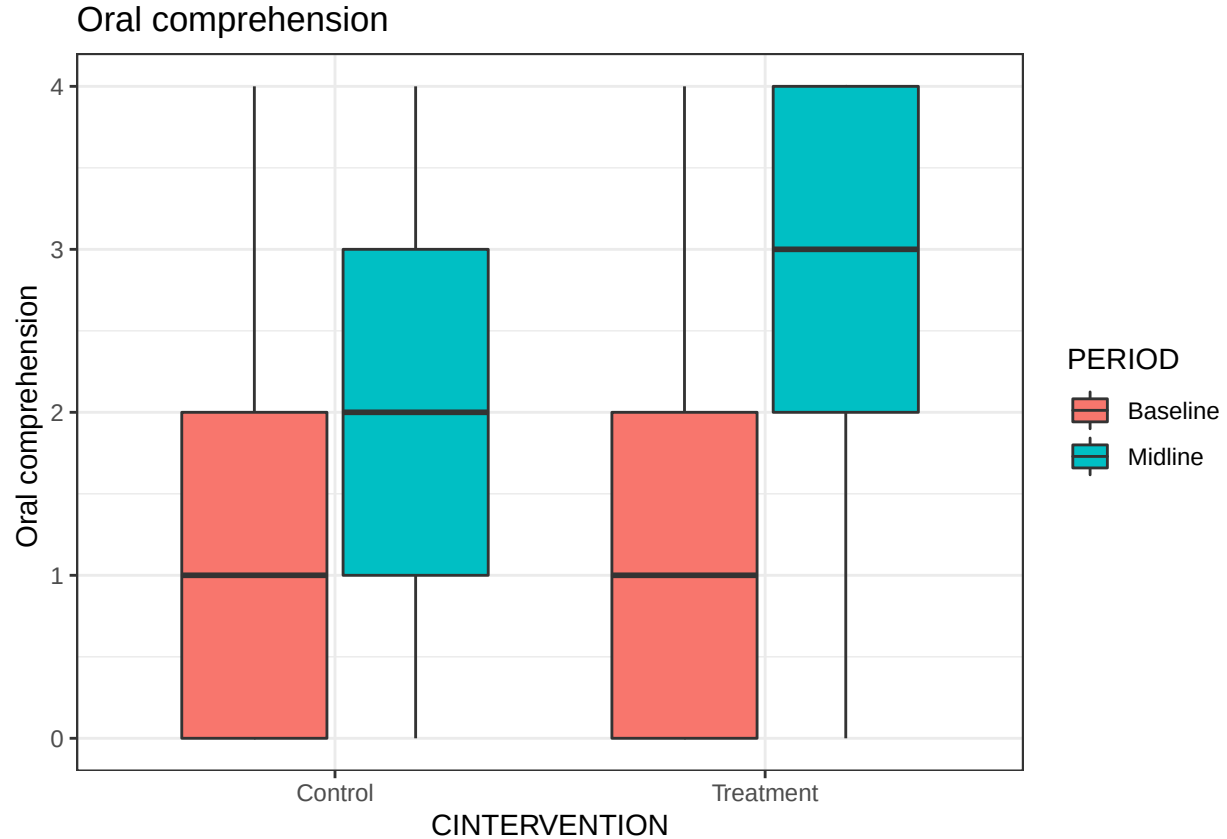
As shown in the table above, for the the Oral comprehension EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at baseline was 0.8089888 (SD = 1.096237) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 1.342007 (SD = 1.45898). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.5330187 points. The p-value for this difference was 0.03396653. The mean for the Control (Comparison (Bilingual)) condition at midline was 1.990826 (SD = 1.117996) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 2.474376 (SD = 1.176385). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.4835501 points. The p-value for this difference was 0.002500106. The change from the baseline to the midline of 1.181837 points for the Control (Comparison (Bilingual)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 1.132368 points. Consequently, the change for the Control (Comparison (Bilingual)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of -0.04946854 points. The p-value for this difference was 0.8882155. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Oral comprehension EGRA subtask across the Control (Comparison (Bilingual)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

1.2.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 9: Oral comprehension

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	1.242	1.390	1007	0	4	1.931	1.446	952	0	4
Treatment	1.235	1.426	1040	0	4	2.555	1.309	1047	0	4



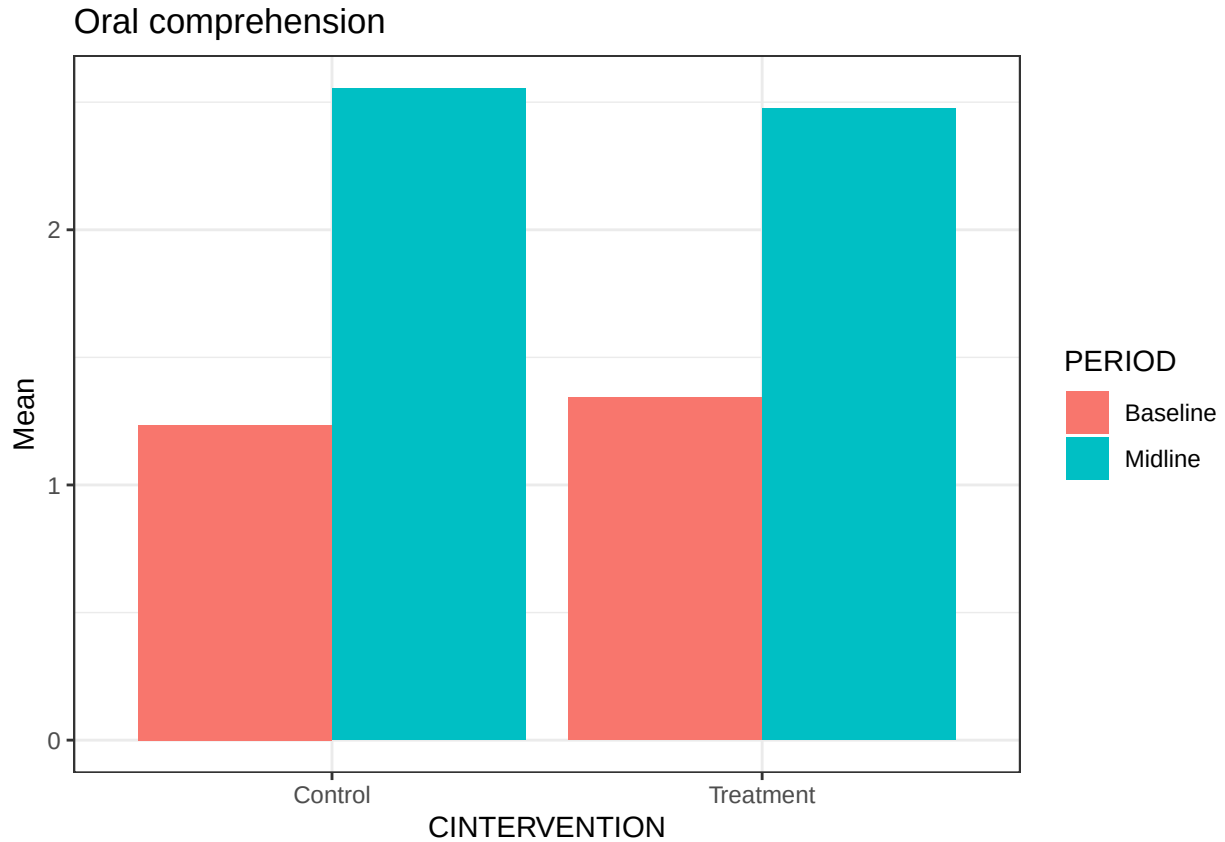


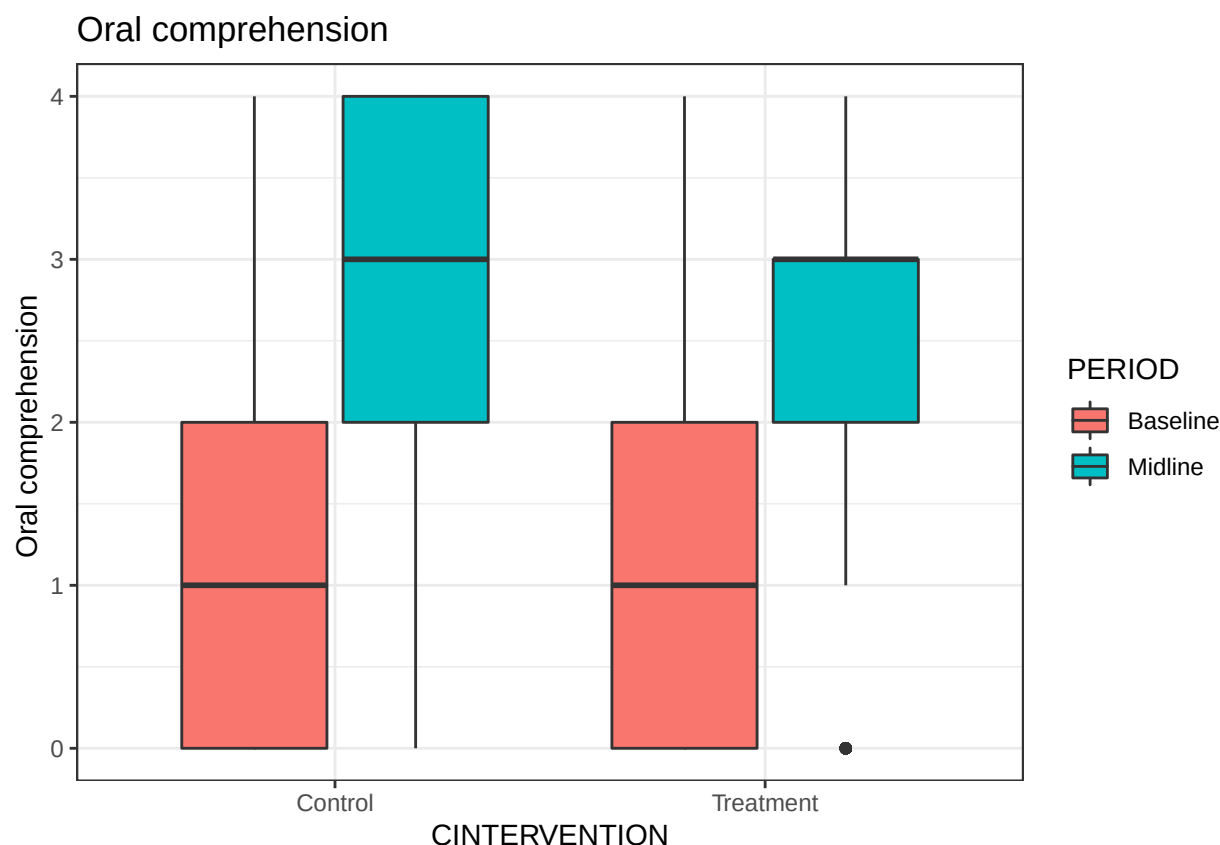
As shown in the table above, for the the Oral comprehension EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at baseline was 1.242304 (SD = 1.390424) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 1.234615 (SD = 1.425991). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.007688488 points. The p-value for this difference was 0.9560903. The mean for the Control (FFE only (Portuguese)) condition at midline was 1.930672 (SD = 1.445623) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 2.554919 (SD = 1.309106). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.6242465 points. The p-value for this difference was 2.647787e-08. The change from the baseline to the midline of 0.6883684 points for the Control (FFE only (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 1.320303 points. Consequently, the change for the Control (FFE only (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 0.631935 points. The p-value for this difference was 0.0007014284. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Portuguese)) when compared to the Control (FFE only (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Portuguese)) (or some other unobserved process) impacted on the Oral comprehension EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

1.2.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 10: Oral comprehension

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	1.235	1.426	1040	0	4	2.555	1.309	1047	0	4
Treatment	1.342	1.459	538	0	4	2.474	1.176	761	0	4





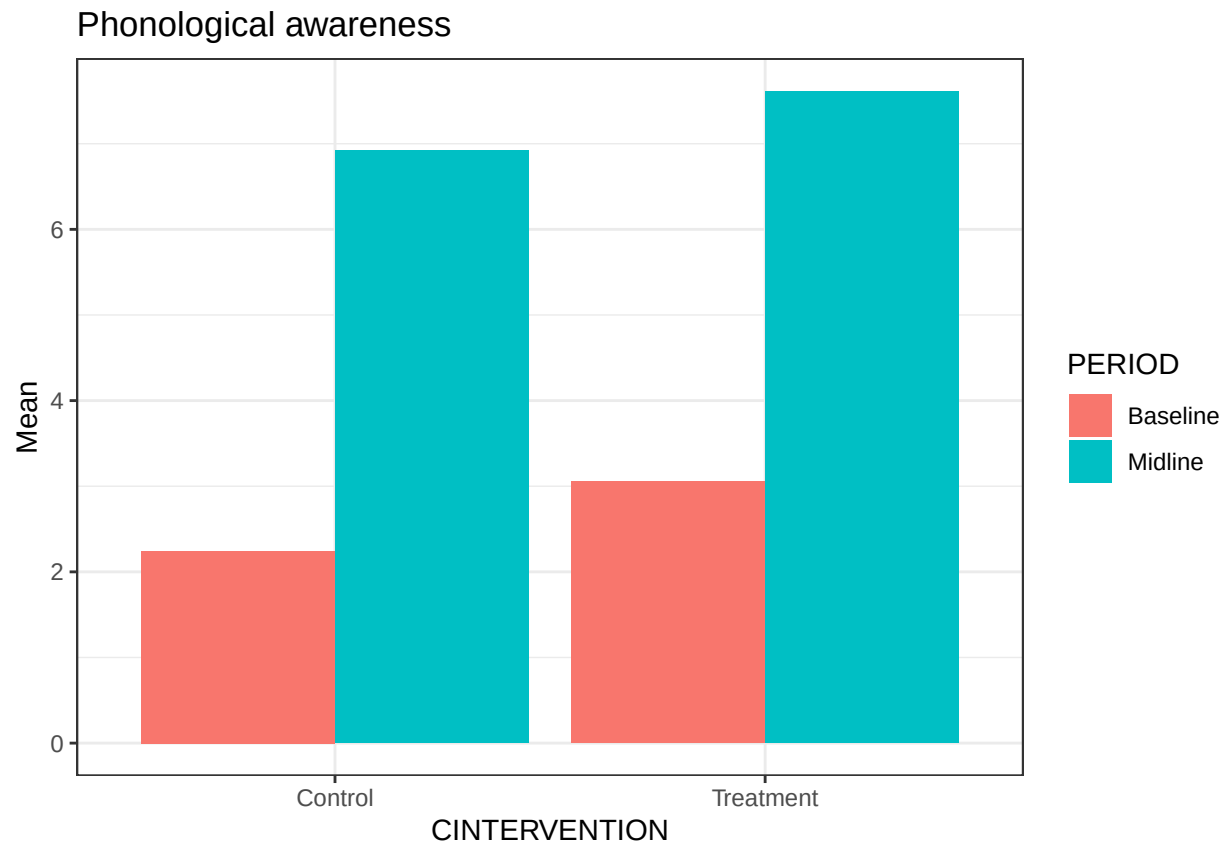
As shown in the table above, for the the Oral comprehension EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at baseline was 1.234615 (SD = 1.425991) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 1.342007 (SD = 1.45898). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.1073921 points. The p-value for this difference was 0.5485581. The mean for the Control (FFE + lit (Portuguese)) condition at midline was 2.554919 (SD = 1.309106) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 2.474376 (SD = 1.176385). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.08054299 points. The p-value for this difference was 0.4197186. The change from the baseline to the midline of 1.320303 points for the Control (FFE + lit (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 1.132368 points. Consequently, the change for the Control (FFE + lit (Portuguese)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of -0.187935 points. The p-value for this difference was 0.3774959. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Oral comprehension EGRA subtask across the Control (FFE + lit (Portuguese)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

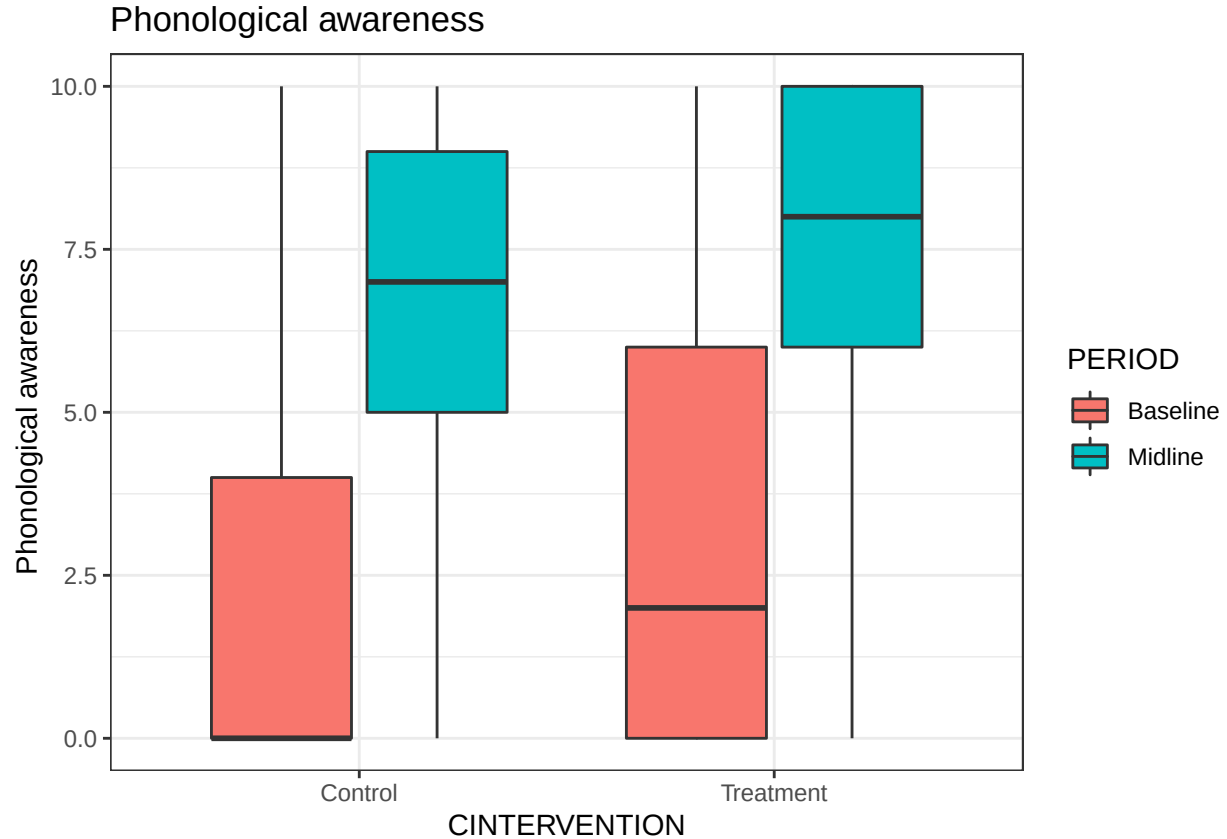
1.3 EGRA_ST3: Phonological awareness

1.3.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 11: Phonological awareness

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	2.245	3.257	1136	0	10	6.925	2.607	1081	0	10
Treatment	3.054	3.462	1578	0	10	7.614	2.399	1808	0	10



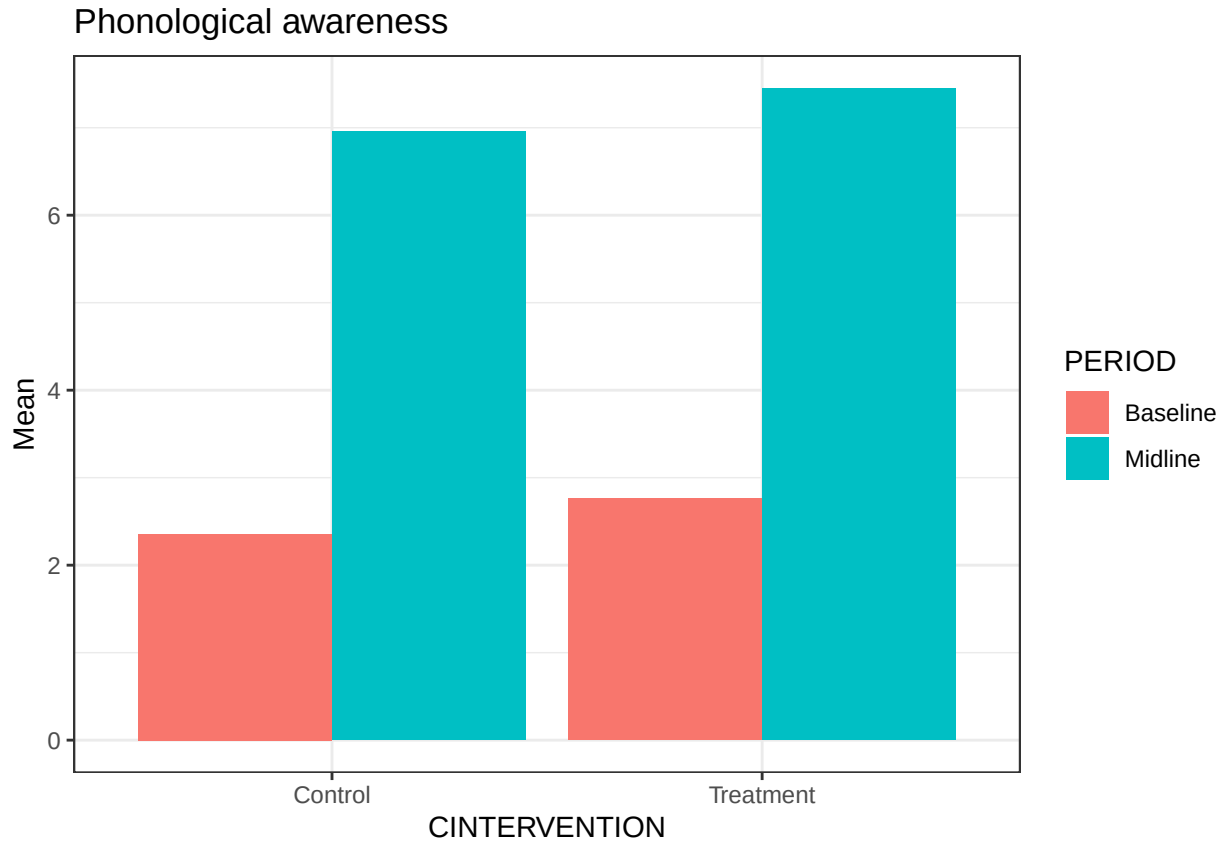


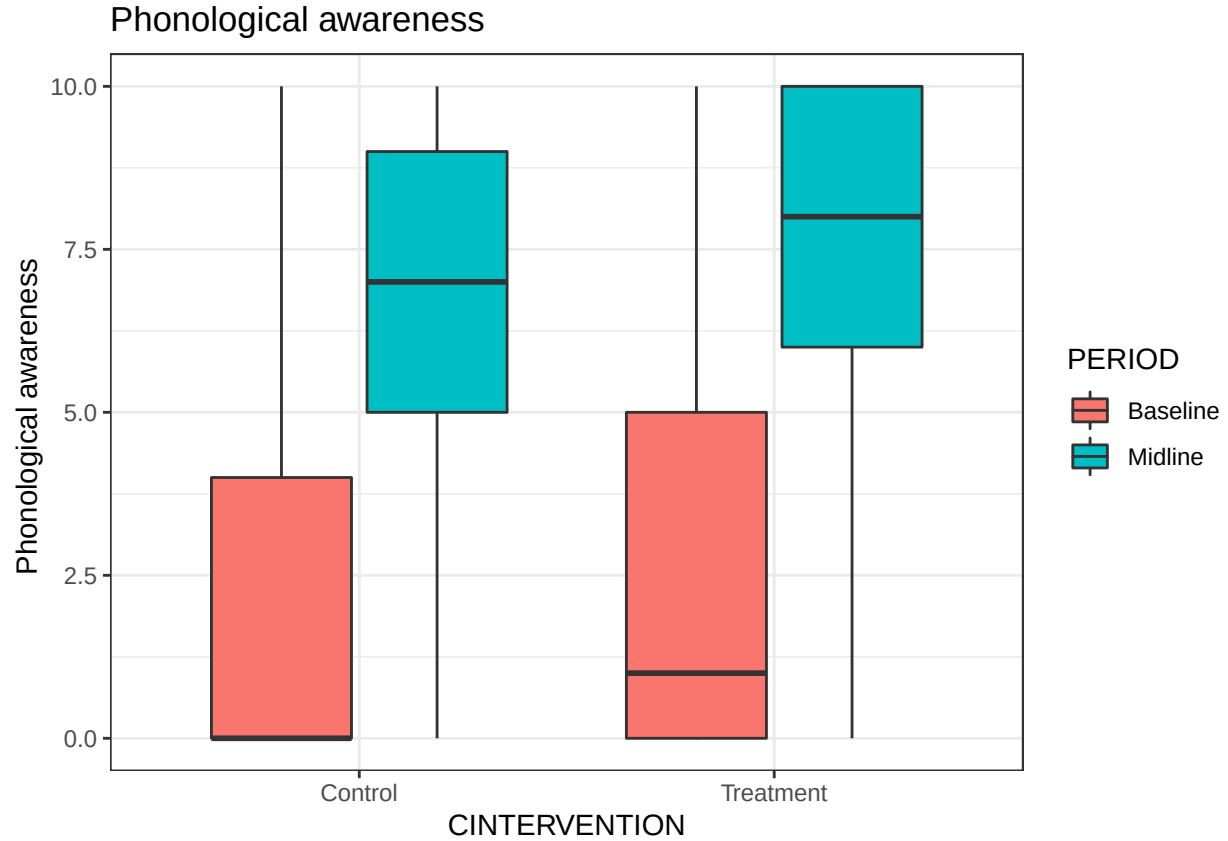
As shown in the table above, for the the Phonological awareness EGRA subtask, the mean for the Control (Comparison (all)) condition at baseline was 2.244718 (SD = 3.256707) and the mean for the Treatment (FFE + lit (all)) condition at baseline was 3.053866 (SD = 3.462126). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at baseline was thus of 0.8091473 points. The p-value for this difference was 0.02043418. The mean for the Control (Comparison (all)) condition at midline was 6.925069 (SD = 2.607136) and the mean for the Treatment (FFE + lit (all)) condition at midline was 7.613938 (SD = 2.39864). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at midline was thus of 0.6888687 points. The p-value for this difference was 2.963609e-05. The change from the baseline to the midline of 4.680351 points for the Control (Comparison (all)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (all)) can be assessed against. The change for the Treatment (FFE + lit (all)) from the baseline to the midline was of 4.560072 points. Consequently, the change for the Control (Comparison (all)) relative to the change for the Treatment (FFE + lit (all)) condition was of -0.1202787 points. The p-value for this difference was 0.7427397. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Phonological awareness EGRA subtask across the Control (Comparison (all)) and the Treatment (FFE + lit (all)) conditions beyond that expected from sampling error.

1.3.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 12: Phonological awareness

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	2.355	3.309	1047	0	10	6.955	2.617	972	0	10
Treatment	2.764	3.238	1040	0	10	7.452	2.456	1047	0	10



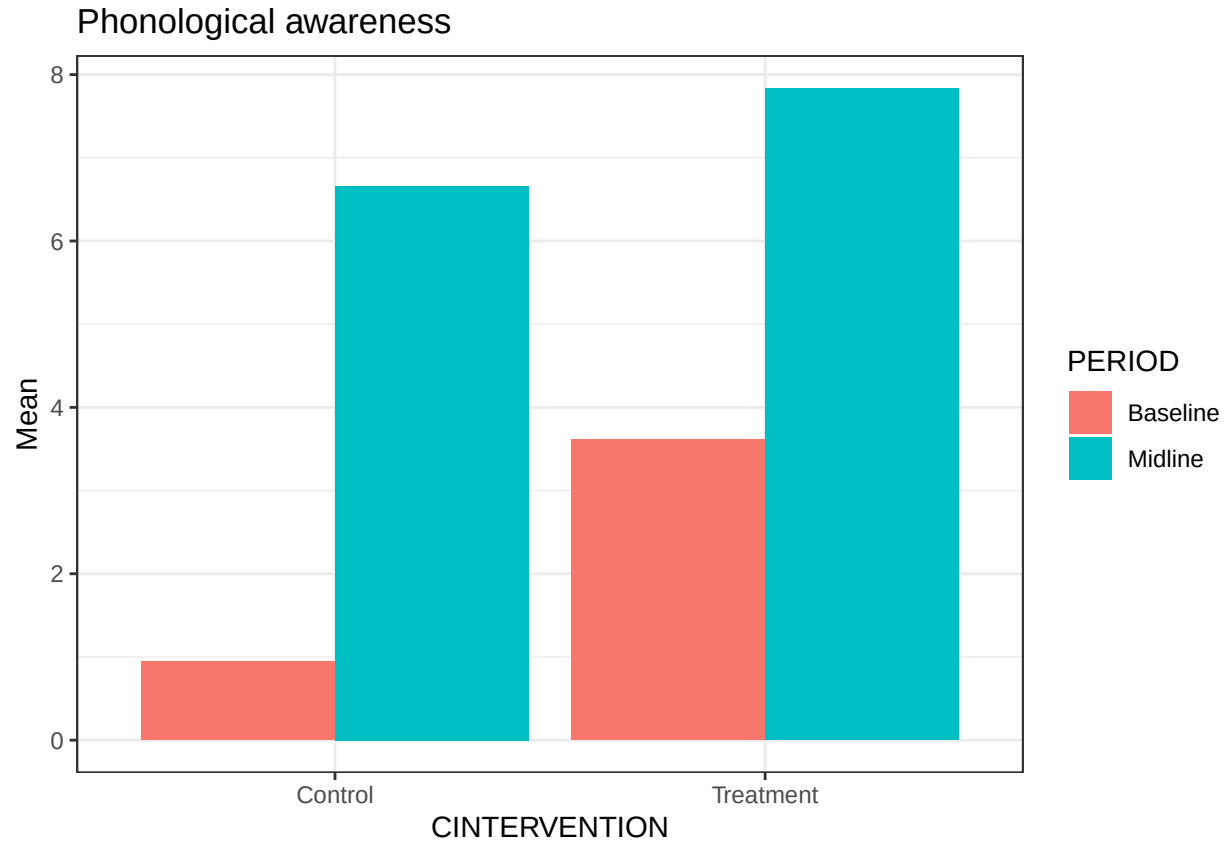


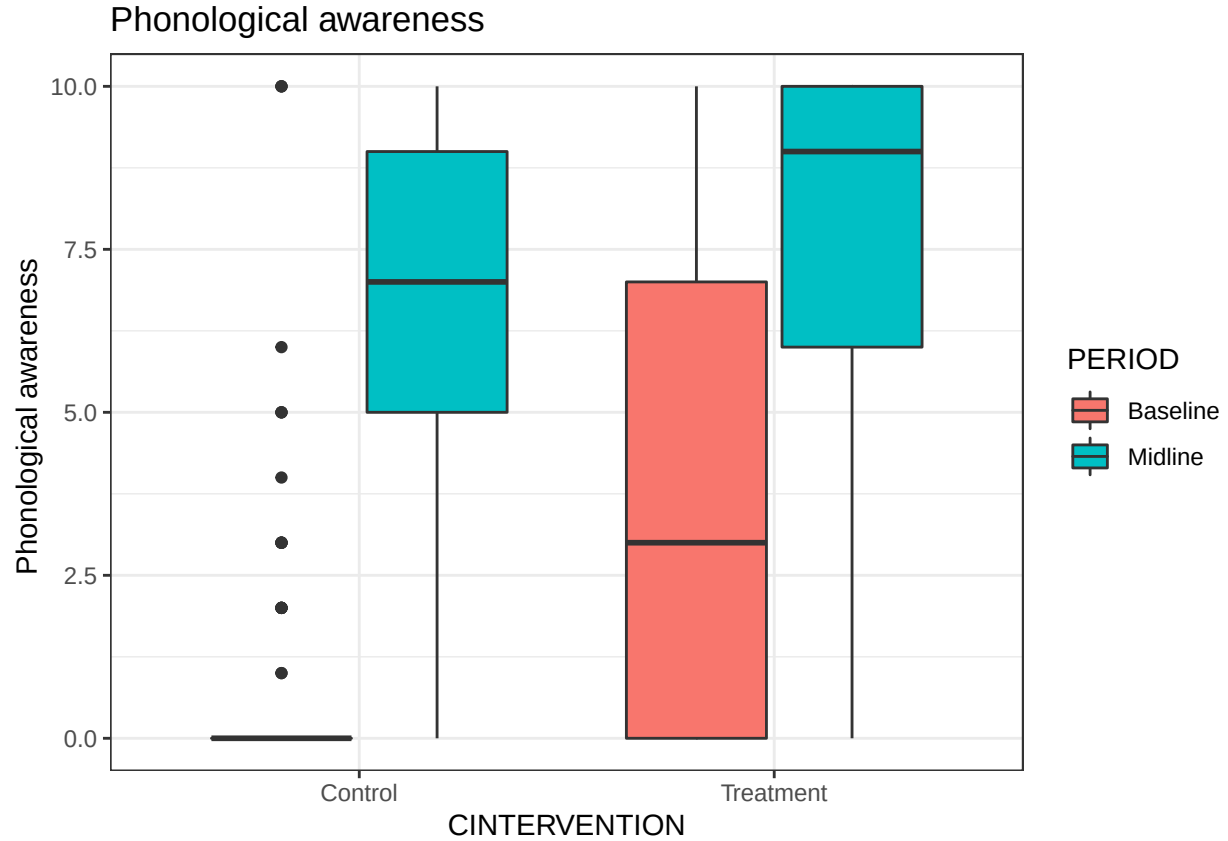
As shown in the table above, for the the Phonological awareness EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at baseline was 2.355301 (SD = 3.309386) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 2.764423 (SD = 3.238258). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.4091222 points. The p-value for this difference was 0.268457. The mean for the Control (Comparison (Portuguese)) condition at midline was 6.954733 (SD = 2.61738) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 7.451767 (SD = 2.456372). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.4970344 points. The p-value for this difference was 0.00867946. The change from the baseline to the midline of 4.599432 points for the Control (Comparison (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 4.687344 points. Consequently, the change for the Control (Comparison (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 0.08791223 points. The p-value for this difference was 0.8213137. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Phonological awareness EGRA subtask across the Control (Comparison (Portuguese)) and the Treatment (FFE + lit (Portuguese)) conditions beyond that expected from sampling error.

1.3.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 13: Phonological awareness

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.944	2.181	89	0	10	6.661	2.51	109	0	10
Treatment	3.613	3.799	538	0	10	7.837	2.30	761	0	10



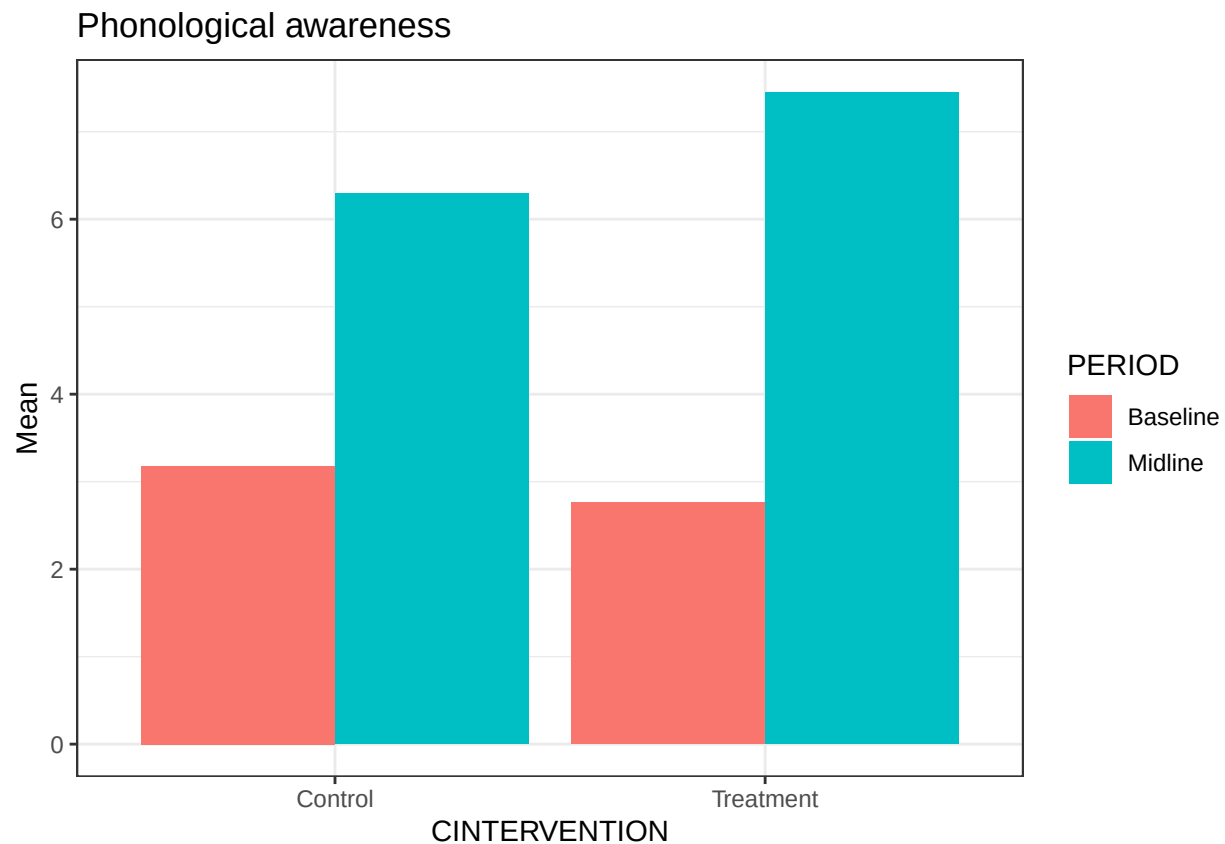


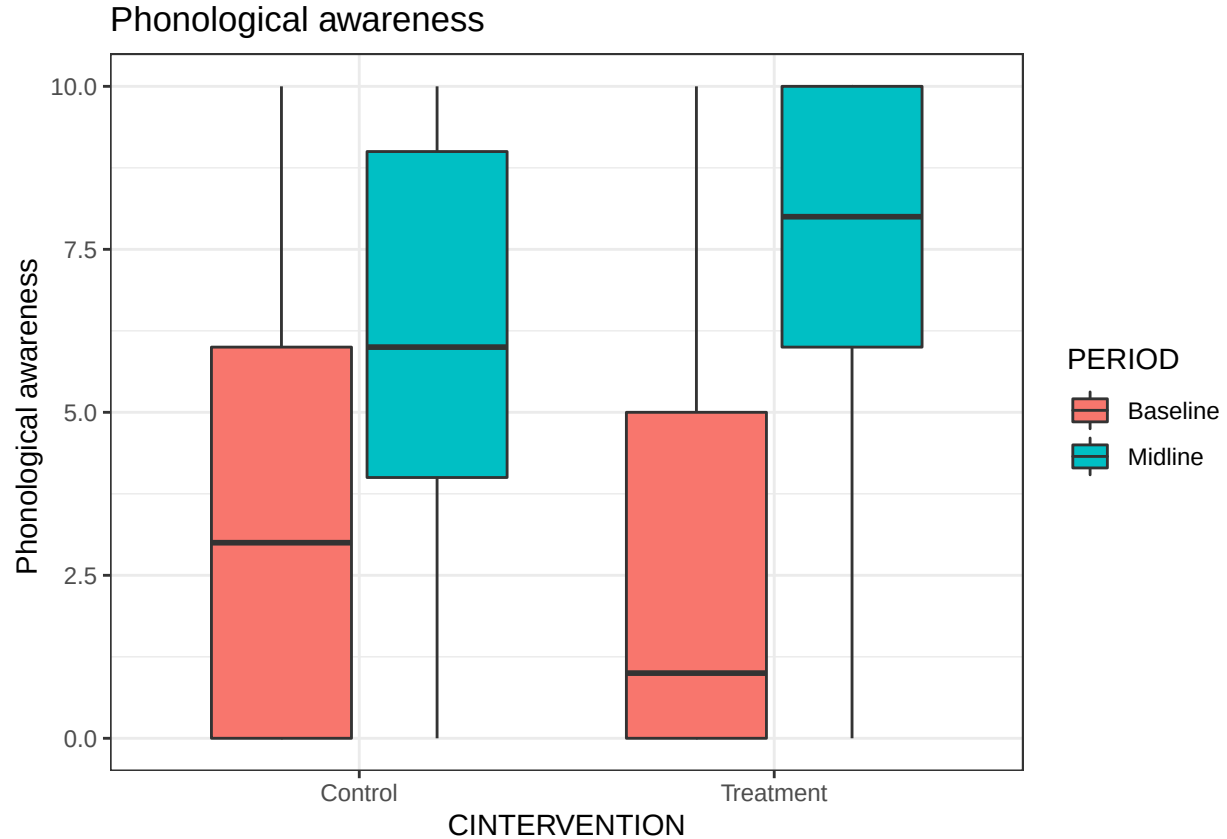
As shown in the table above, for the the Phonological awareness EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at baseline was 0.9438202 (SD = 2.181323) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 3.613383 (SD = 3.799339). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 2.669563 points. The p-value for this difference was 6.276787e-07. The mean for the Control (Comparison (Bilingual)) condition at midline was 6.66055 (SD = 2.510055) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 7.837057 (SD = 2.299827). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 1.176506 points. The p-value for this difference was 1.406602e-05. The change from the baseline to the midline of 5.71673 points for the Control (Comparison (Bilingual)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 4.223674 points. Consequently, the change for the Control (Comparison (Bilingual)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of -1.493057 points. The p-value for this difference was 0.02097117. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Phonological awareness EGRA subtask across the Control (Comparison (Bilingual)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

1.3.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 14: Phonological awareness

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	3.180	3.233	1007	0	10	6.294	2.757	952	0	10
Treatment	2.764	3.238	1040	0	10	7.452	2.456	1047	0	10



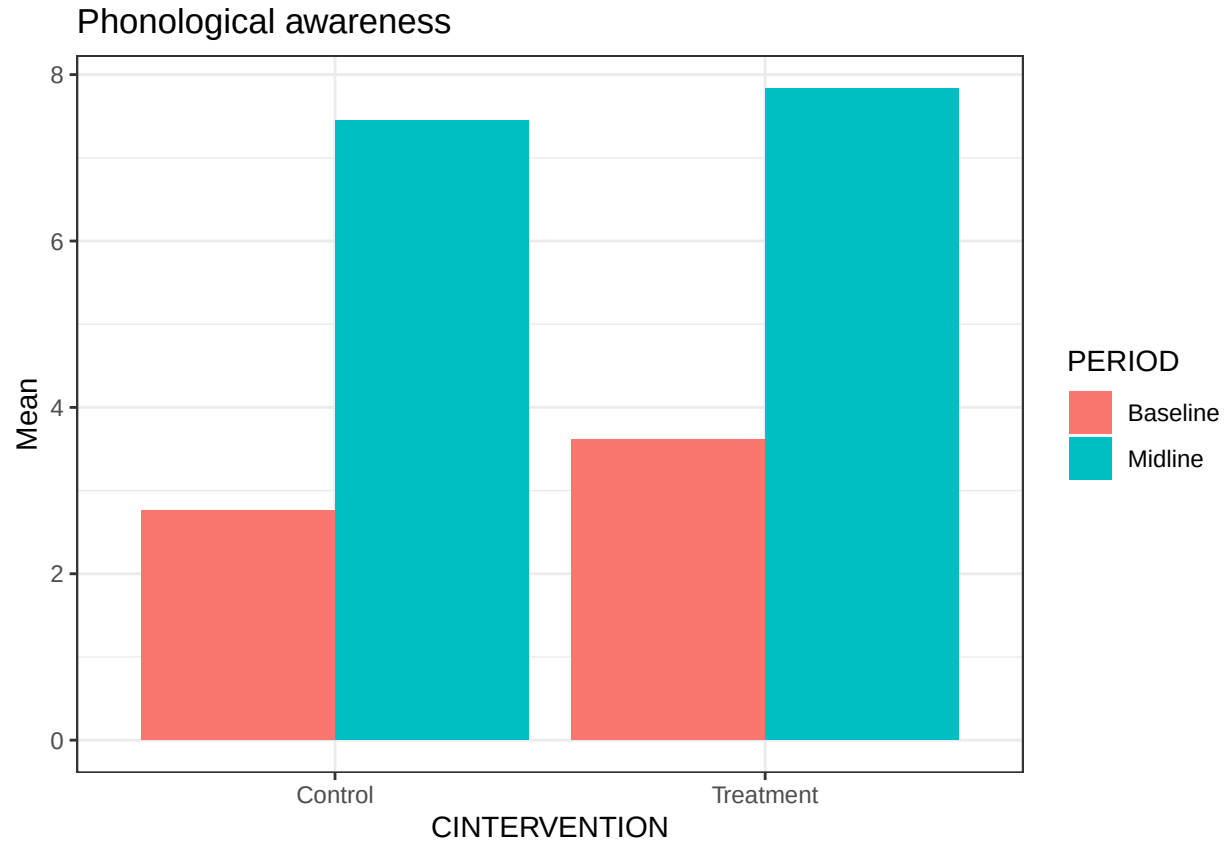


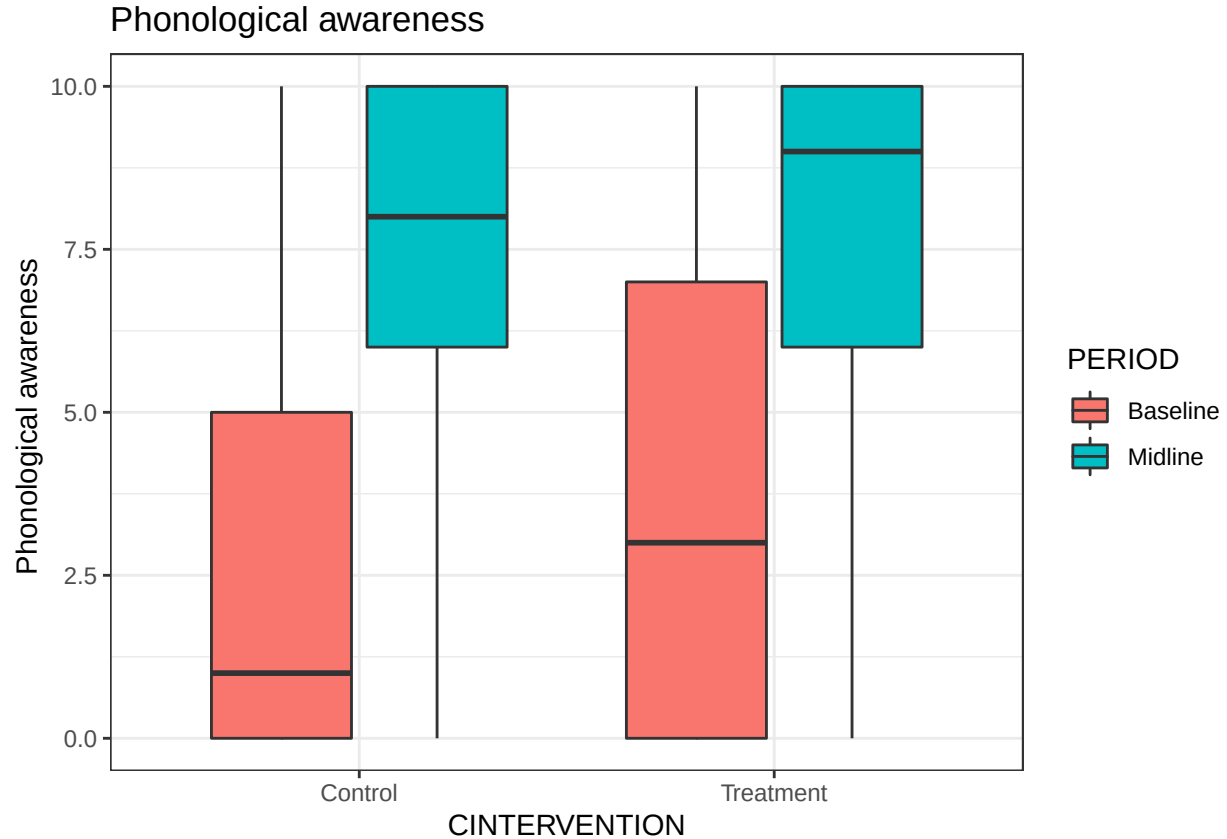
As shown in the table above, for the the Phonological awareness EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at baseline was 3.179742 (SD = 3.232918) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 2.764423 (SD = 3.238258). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.4153187 points. The p-value for this difference was 0.1928234. The mean for the Control (FFE only (Portuguese)) condition at midline was 6.294118 (SD = 2.756819) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 7.451767 (SD = 2.456372). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 1.157649 points. The p-value for this difference was 2.845414e-08. The change from the baseline to the midline of 3.114376 points for the Control (FFE only (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 4.687344 points. Consequently, the change for the Control (FFE only (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 1.572968 points. The p-value for this difference was 9.488215e-05. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Portuguese)) when compared to the Control (FFE only (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Portuguese)) (or some other unobserved process) impacted on the Phonological awareness EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

1.3.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 15: Phonological awareness

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	2.764	3.238	1040	0	10	7.452	2.456	1047	0	10
Treatment	3.613	3.799	538	0	10	7.837	2.300	761	0	10





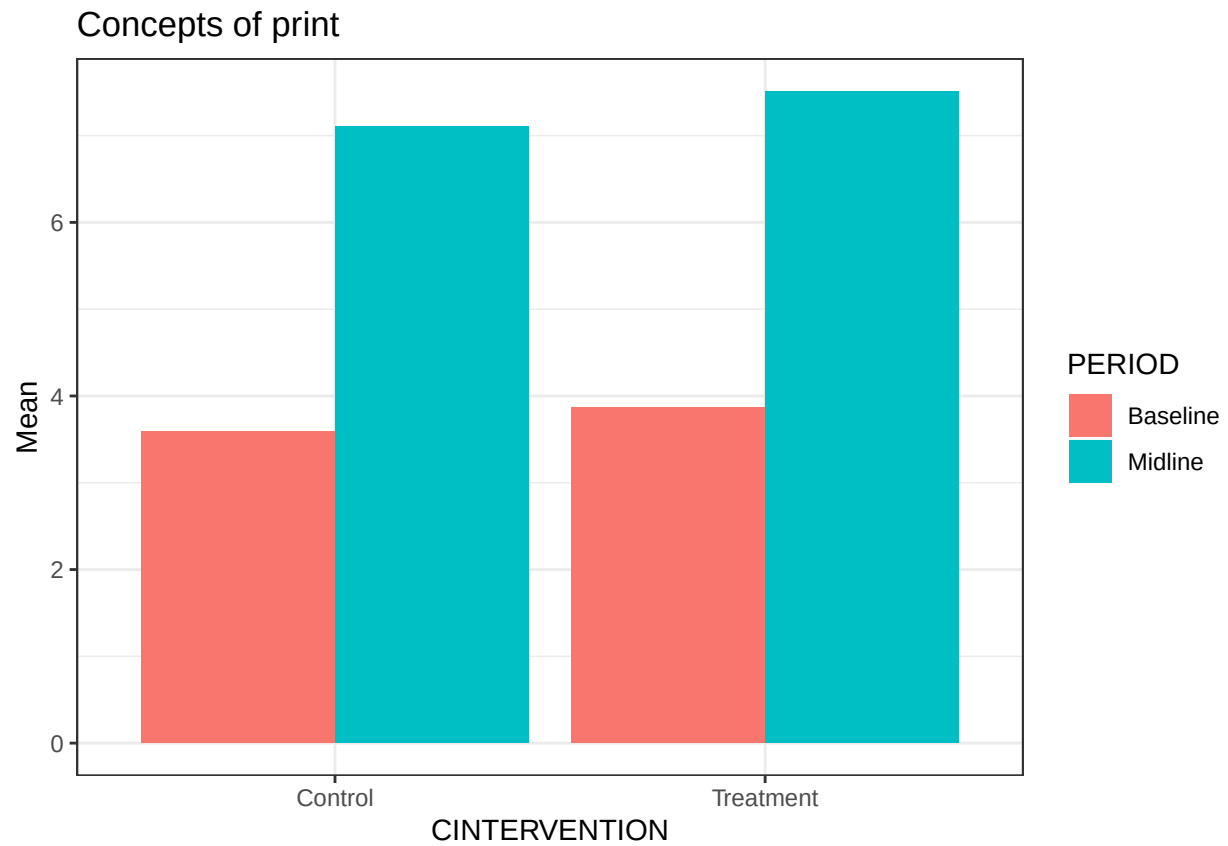
As shown in the table above, for the the Phonological awareness EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at baseline was 2.764423 (SD = 3.238258) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 3.613383 (SD = 3.799339). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.8489598 points. The p-value for this difference was 0.04410332. The mean for the Control (FFE + lit (Portuguese)) condition at midline was 7.451767 (SD = 2.456372) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 7.837057 (SD = 2.299827). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.3852896 points. The p-value for this difference was 0.02207021. The change from the baseline to the midline of 4.687344 points for the Control (FFE + lit (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 4.223674 points. Consequently, the change for the Control (FFE + lit (Portuguese)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of -0.4636703 points. The p-value for this difference was 0.3129217. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Phonological awareness EGRA subtask across the Control (FFE + lit (Portuguese)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

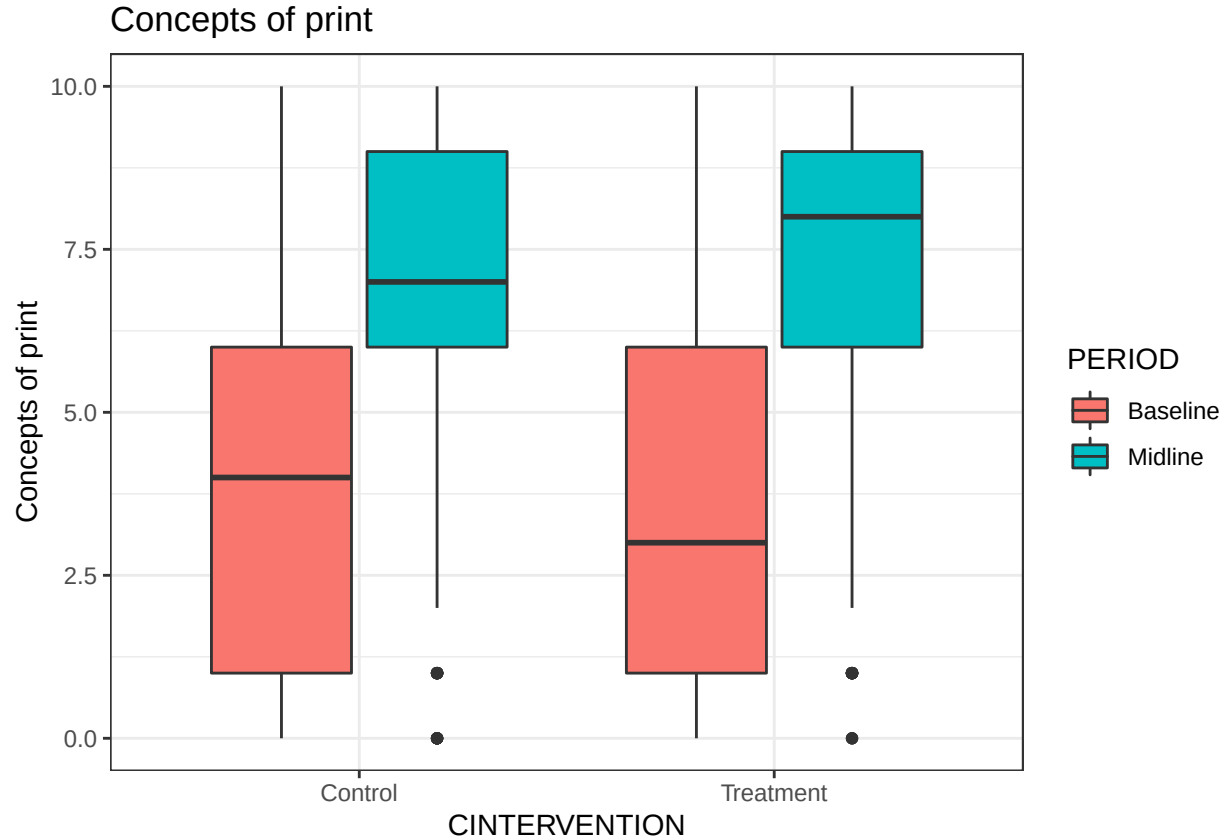
1.4 EGRA_ST4: Concepts of print

1.4.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 16: Concepts of print

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	3.589	2.675	1136	0	10	7.107	2.333	1081	0	10
Treatment	3.864	3.163	1578	0	10	7.512	2.208	1808	0	10



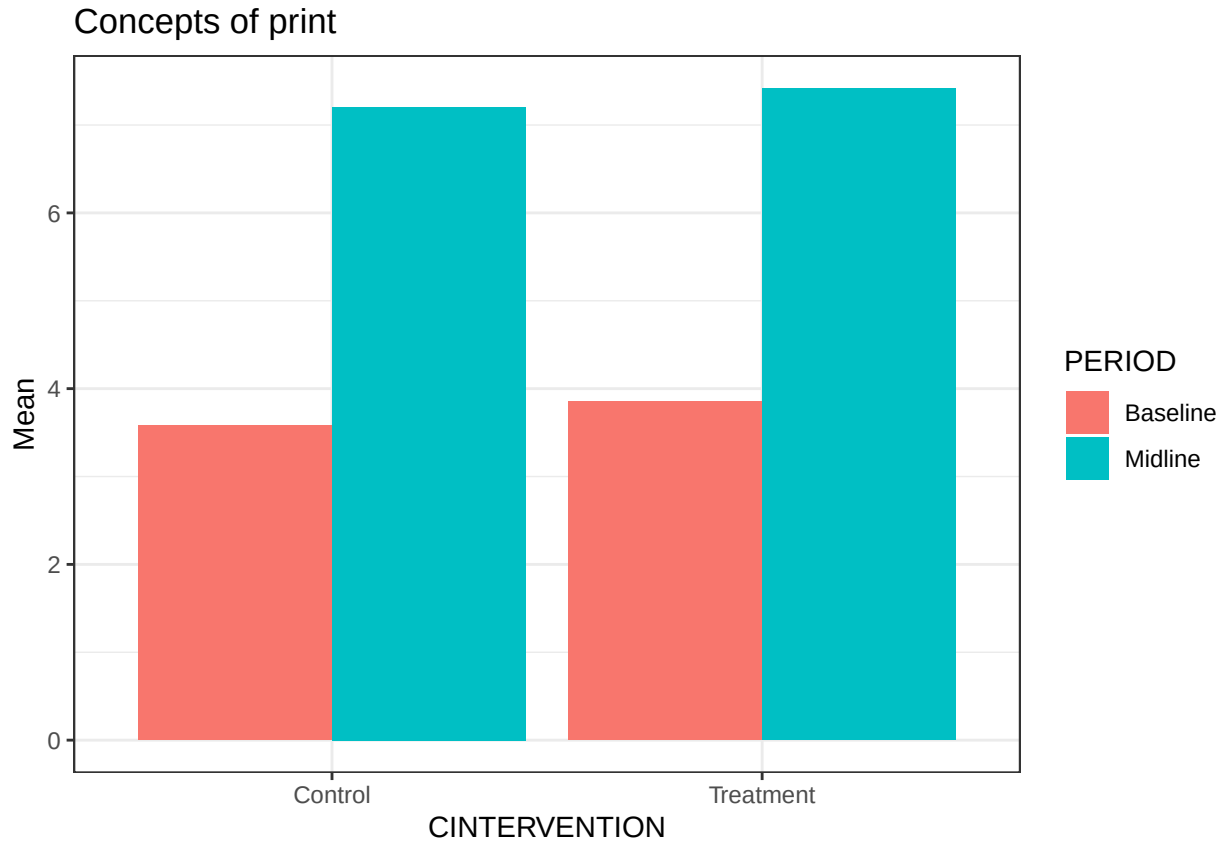


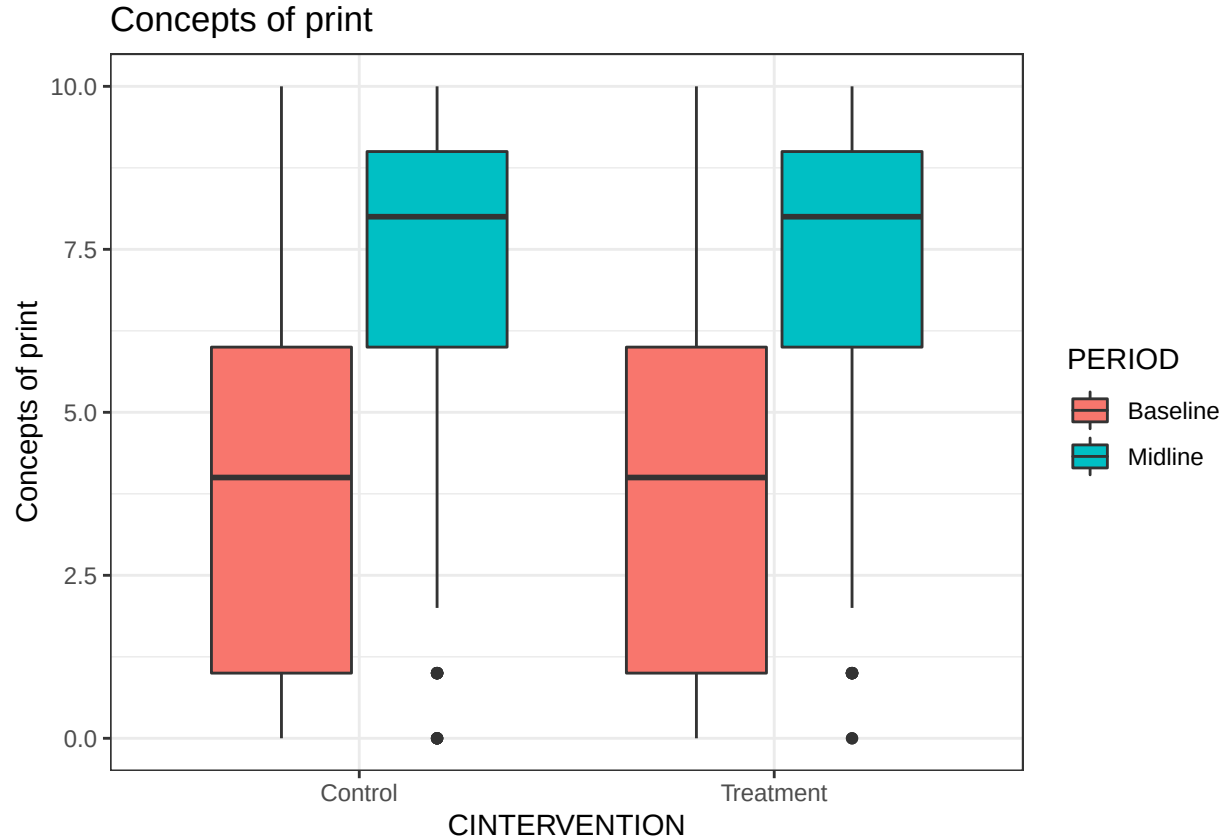
As shown in the table above, for the the Concepts of print EGRA subtask, the mean for the Control (Comparison (all)) condition at baseline was 3.588908 (SD = 2.674899) and the mean for the Treatment (FFE + lit (all)) condition at baseline was 3.864385 (SD = 3.162977). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at baseline was thus of 0.2754768 points. The p-value for this difference was 0.2499473. The mean for the Control (Comparison (all)) condition at midline was 7.107308 (SD = 2.333244) and the mean for the Treatment (FFE + lit (all)) condition at midline was 7.512168 (SD = 2.208267). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at midline was thus of 0.4048601 points. The p-value for this difference was 0.008410878. The change from the baseline to the midline of 3.5184 points for the Control (Comparison (all)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (all)) can be assessed against. The change for the Treatment (FFE + lit (all)) from the baseline to the midline was of 3.647783 points. Consequently, the change for the Control (Comparison (all)) relative to the change for the Treatment (FFE + lit (all)) condition was of 0.1293832 points. The p-value for this difference was 0.6381239. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Concepts of print EGRA subtask across the Control (Comparison (all)) and the Treatment (FFE + lit (all)) conditions beyond that expected from sampling error.

1.4.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 17: Concepts of print

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	3.585	2.676	1047	0	10	7.205	2.290	972	0	10
Treatment	3.855	3.143	1040	0	10	7.419	2.219	1047	0	10



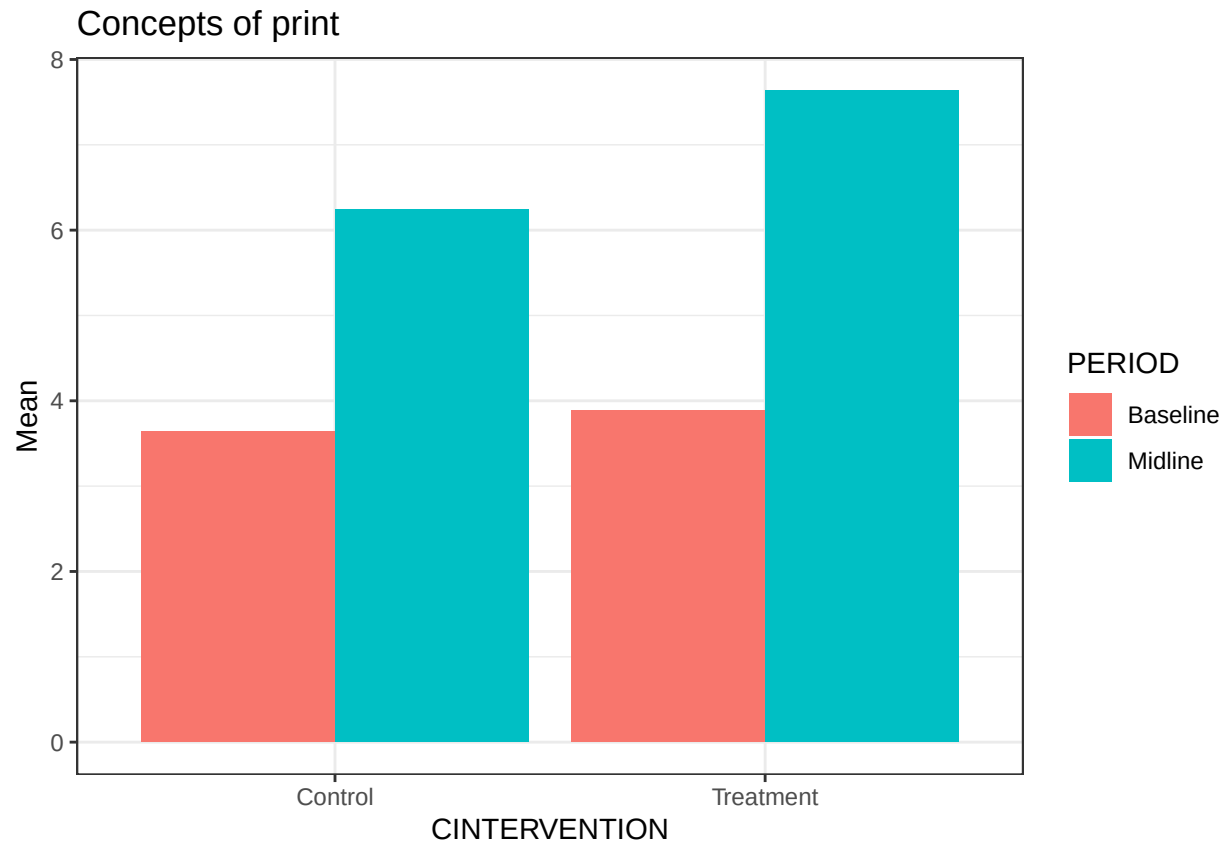


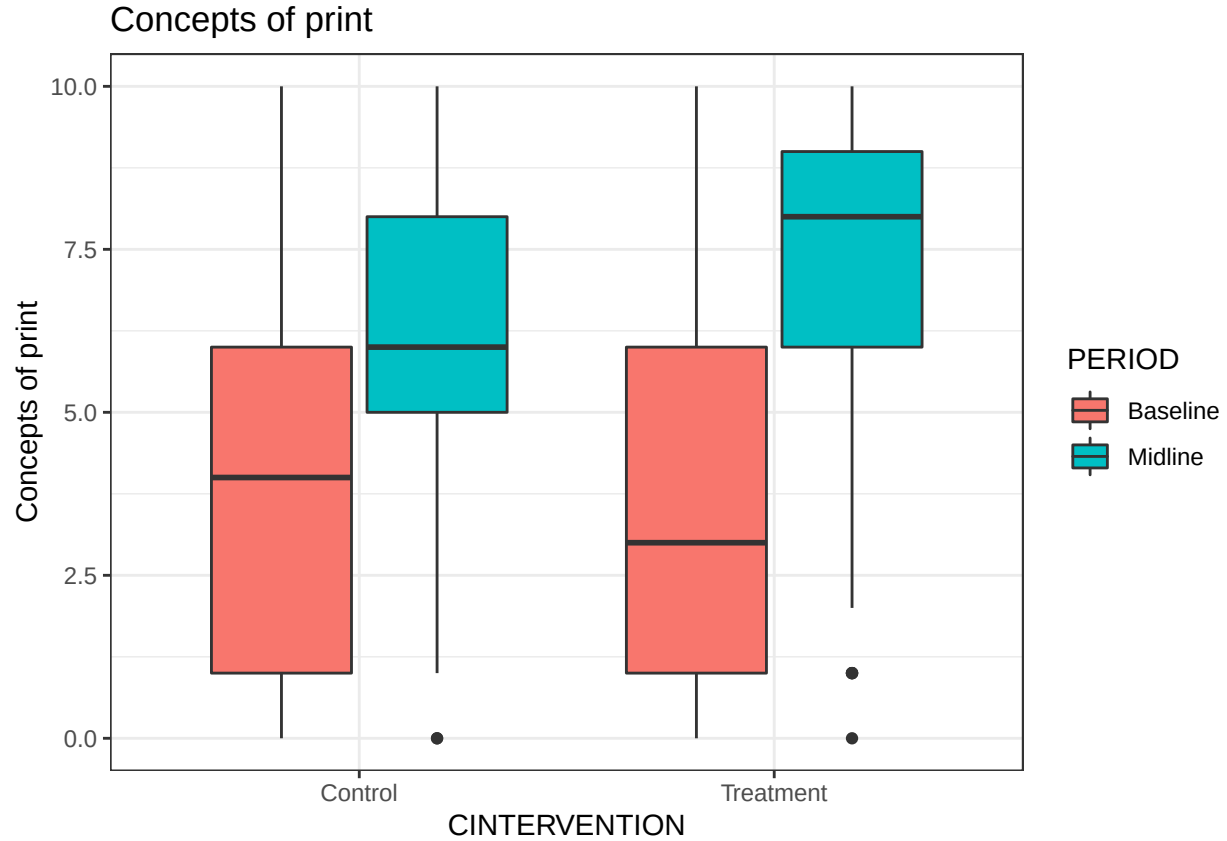
As shown in the table above, for the the Concepts of print EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at baseline was 3.584527 (SD = 2.67598) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 3.854808 (SD = 3.143209). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.2702805 points. The p-value for this difference was 0.3261805. The mean for the Control (Comparison (Portuguese)) condition at midline was 7.204733 (SD = 2.290053) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 7.419293 (SD = 2.21867). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.2145607 points. The p-value for this difference was 0.2067414. The change from the baseline to the midline of 3.620205 points for the Control (Comparison (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 3.564486 points. Consequently, the change for the Control (Comparison (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of -0.05571976 points. The p-value for this difference was 0.8602412. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Concepts of print EGRA subtask across the Control (Comparison (Portuguese)) and the Treatment (FFE + lit (Portuguese)) conditions beyond that expected from sampling error.

1.4.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 18: Concepts of print

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	3.640	2.677	89	0	10	6.239	2.538	109	0	10
Treatment	3.883	3.204	538	0	10	7.640	2.189	761	0	10



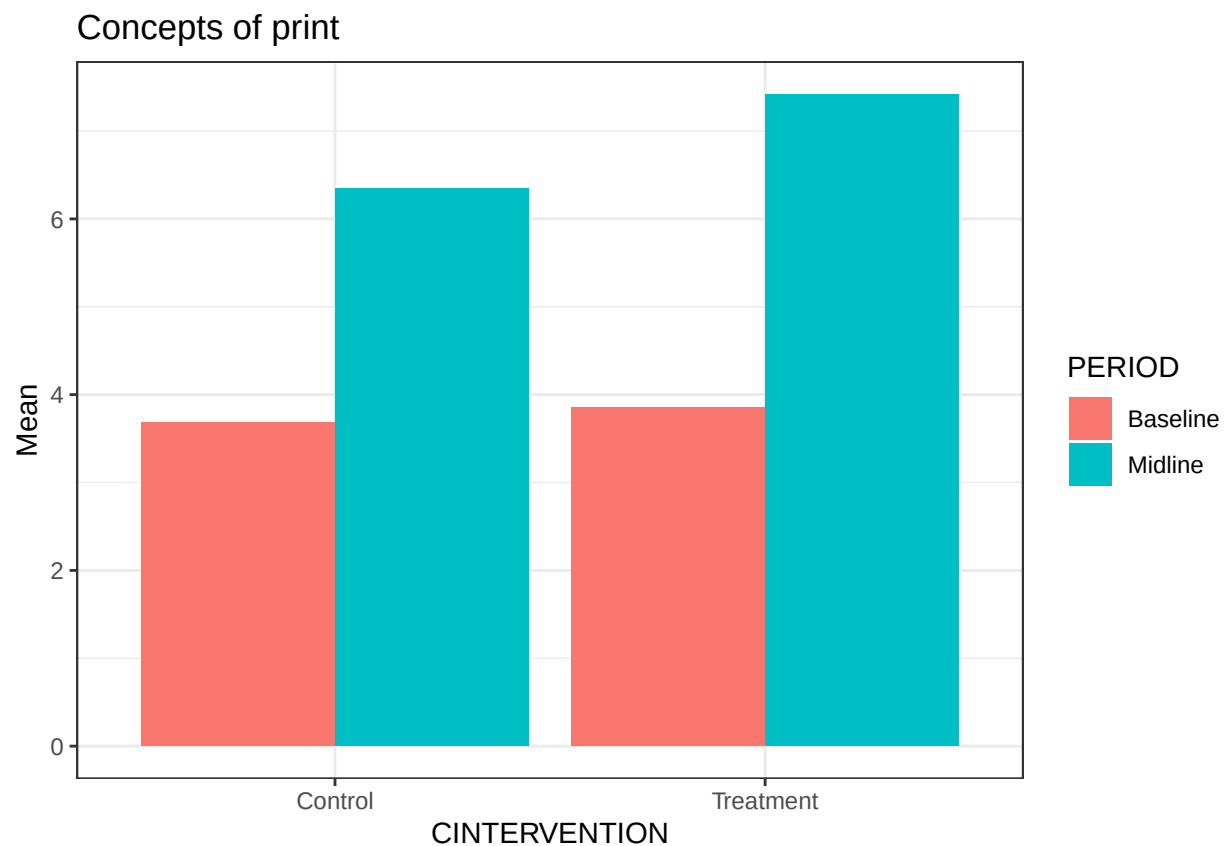


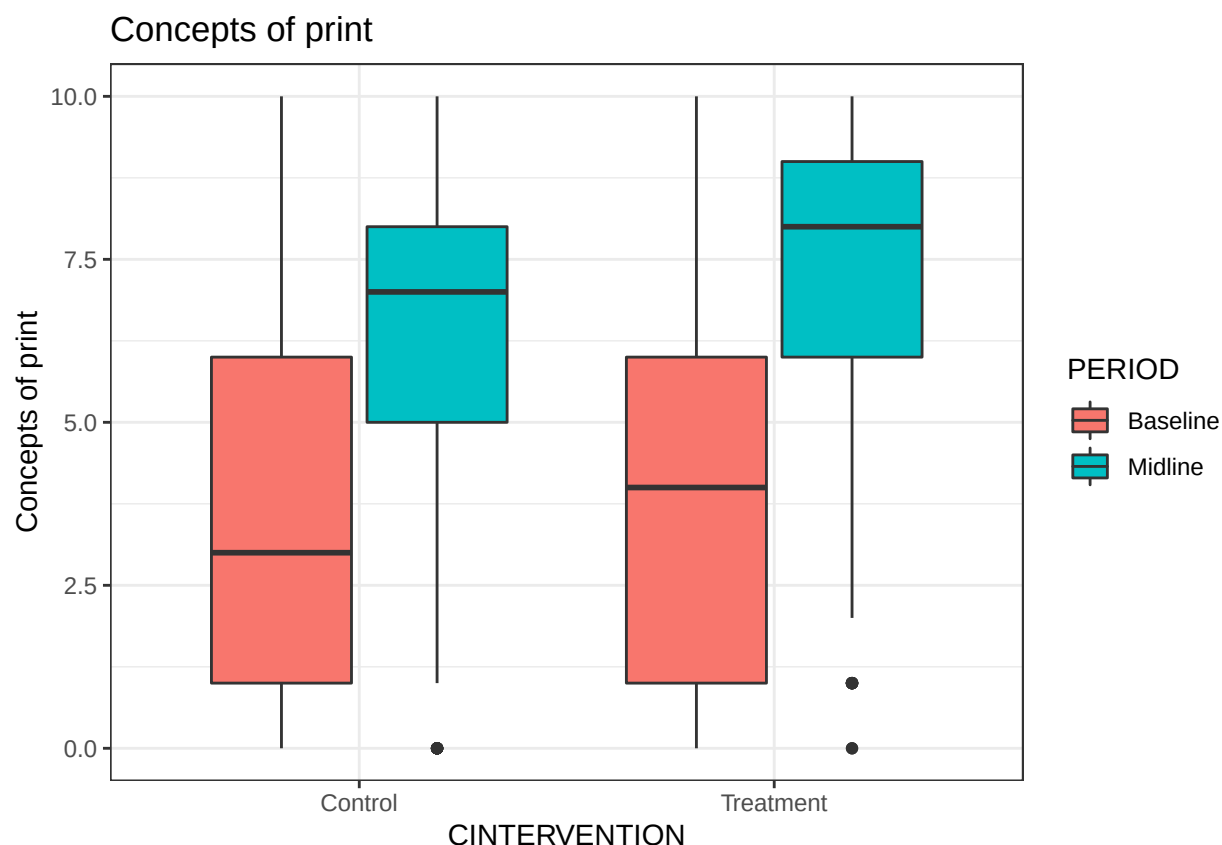
As shown in the table above, for the the Concepts of print EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at baseline was 3.640449 (SD = 2.676697) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 3.8829 (SD = 3.203706). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.2424502 points. The p-value for this difference was 0.7076607. The mean for the Control (Comparison (Bilingual)) condition at midline was 6.238532 (SD = 2.538223) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 7.639947 (SD = 2.188891). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 1.401415 points. The p-value for this difference was 1.156373e-06. The change from the baseline to the midline of 2.598083 points for the Control (Comparison (Bilingual)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 3.757048 points. Consequently, the change for the Control (Comparison (Bilingual)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 1.158965 points. The p-value for this difference was 0.06620163. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Concepts of print EGRA subtask across the Control (Comparison (Bilingual)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

1.4.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 19: Concepts of print

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	3.680	3.092	1007	0	10	6.348	2.453	952	0	10
Treatment	3.855	3.143	1040	0	10	7.419	2.219	1047	0	10



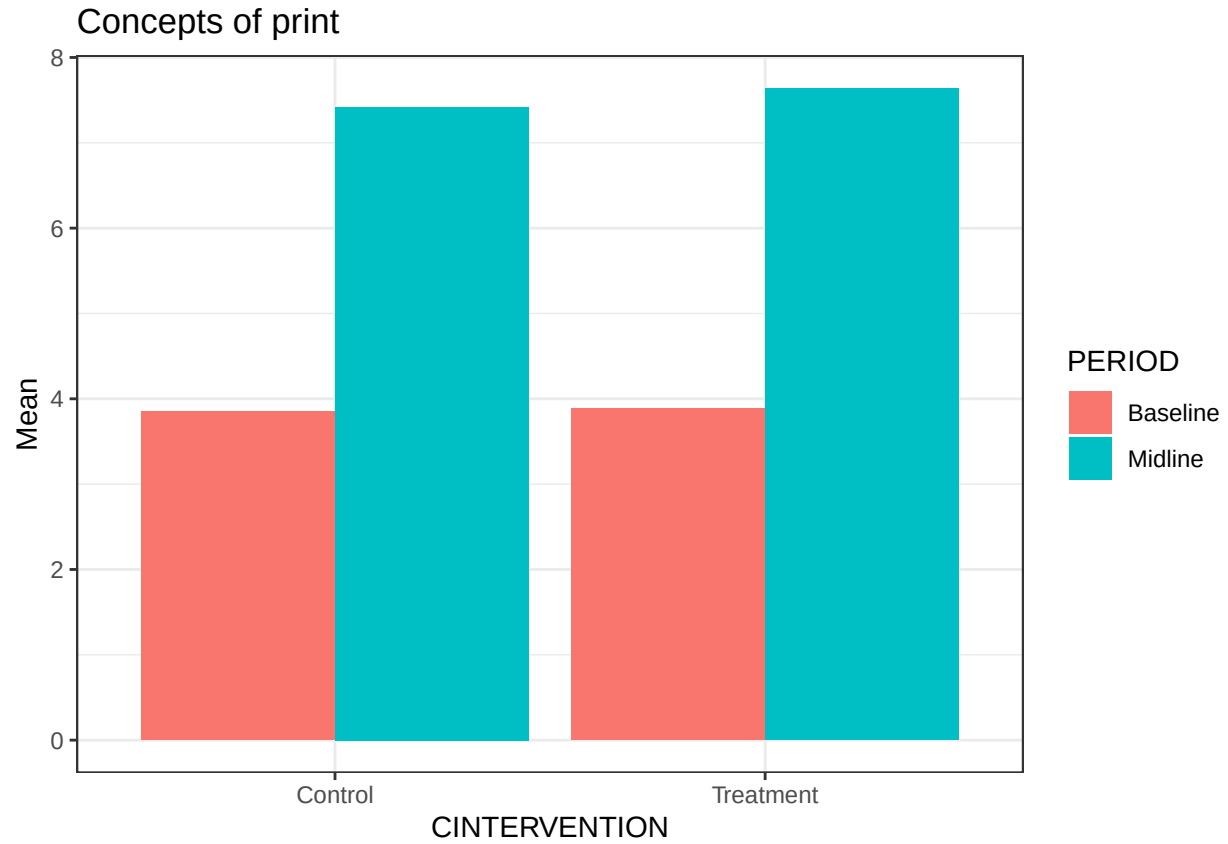


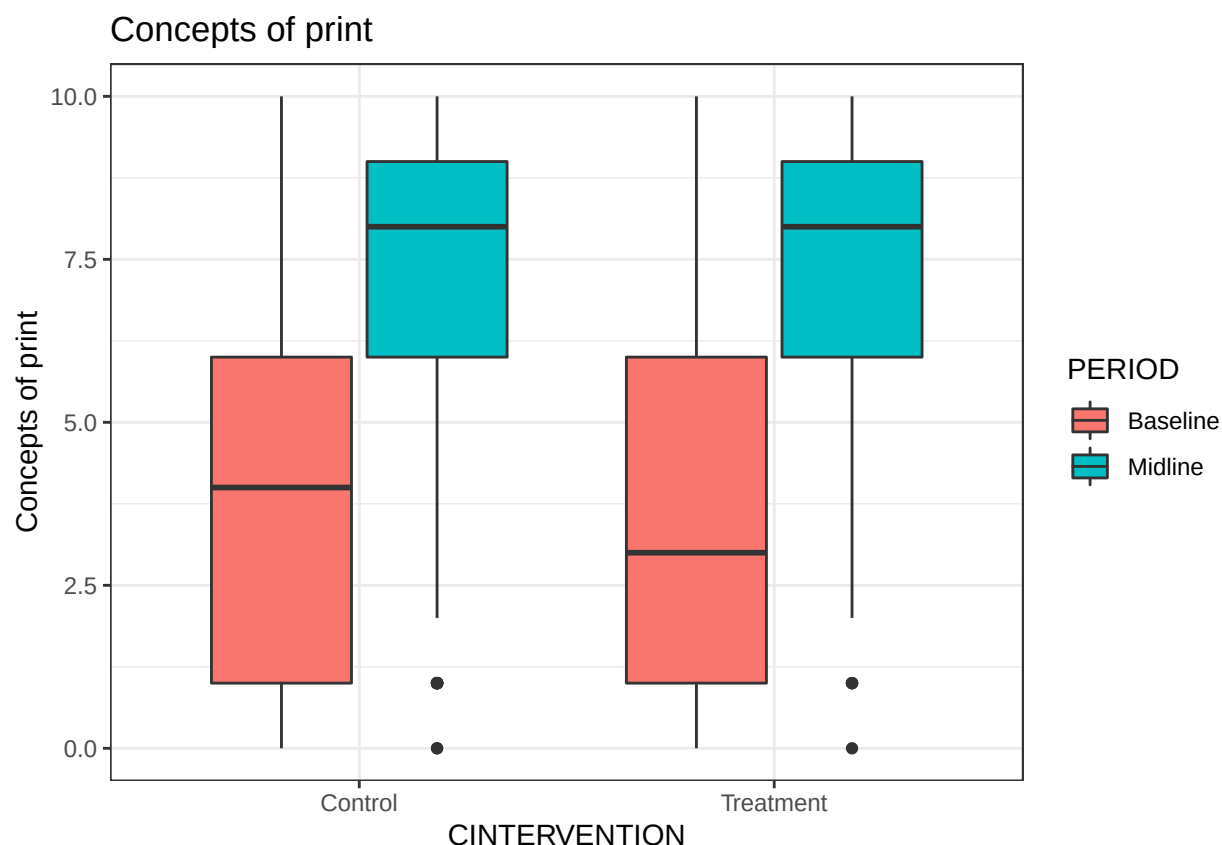
As shown in the table above, for the the Concepts of print EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at baseline was 3.680238 (SD = 3.091551) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 3.854808 (SD = 3.143209). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.1745694 points. The p-value for this difference was 0.5574752. The mean for the Control (FFE only (Portuguese)) condition at midline was 6.347689 (SD = 2.452903) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 7.419293 (SD = 2.21867). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 1.071604 points. The p-value for this difference was 4.766178e-10. The change from the baseline to the midline of 2.667451 points for the Control (FFE only (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 3.564486 points. Consequently, the change for the Control (FFE only (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 0.8970348 points. The p-value for this difference was 0.01683061. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Concepts of print EGRA subtask across the Control (FFE only (Portuguese)) and the Treatment (FFE + lit (Portuguese)) conditions beyond that expected from sampling error.

1.4.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 20: Concepts of print

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	3.855	3.143	1040	0	10	7.419	2.219	1047	0	10
Treatment	3.883	3.204	538	0	10	7.640	2.189	761	0	10





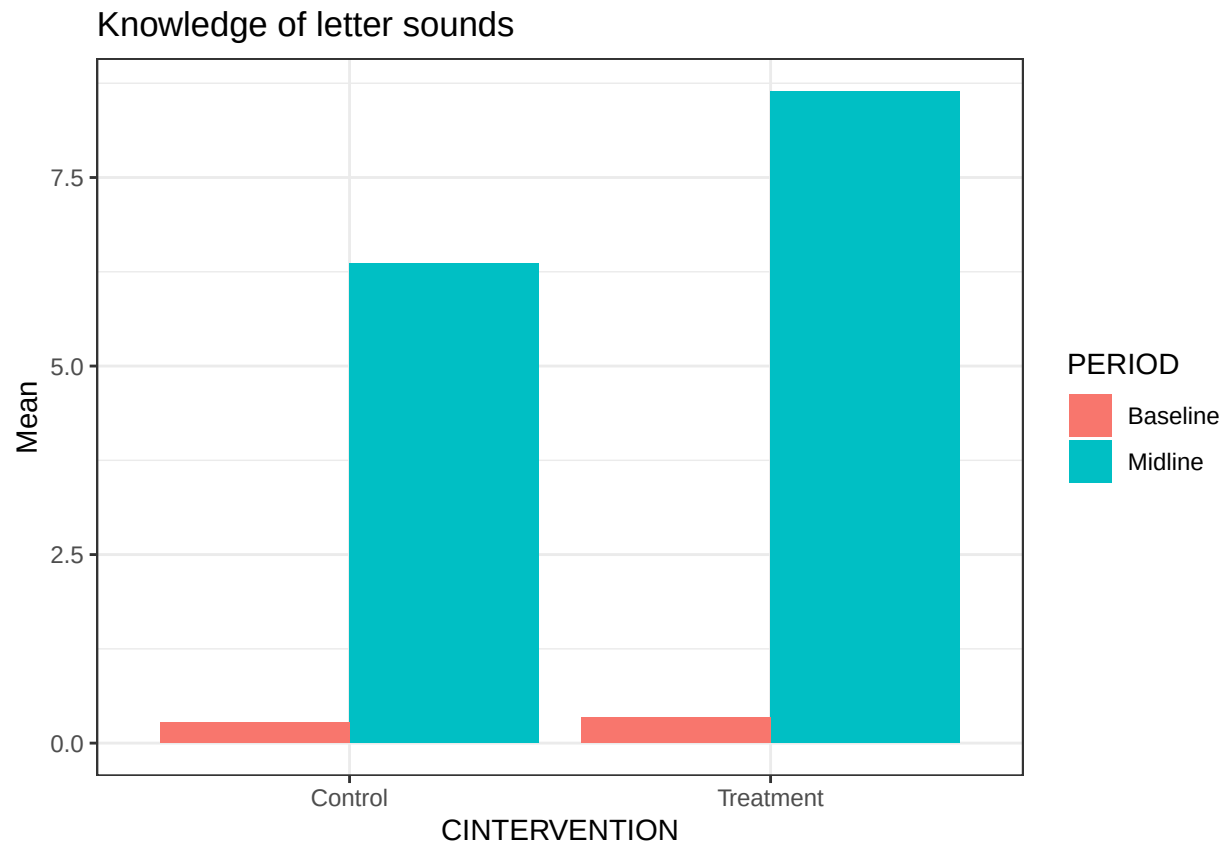
As shown in the table above, for the the Concepts of print EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at baseline was 3.854808 (SD = 3.143209) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 3.8829 (SD = 3.203706). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.02809194 points. The p-value for this difference was 0.9391503. The mean for the Control (FFE + lit (Portuguese)) condition at midline was 7.419293 (SD = 2.21867) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 7.639947 (SD = 2.188891). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.2206542 points. The p-value for this difference was 0.2185148. The change from the baseline to the midline of 3.564486 points for the Control (FFE + lit (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 3.757048 points. Consequently, the change for the Control (FFE + lit (Portuguese)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 0.1925623 points. The p-value for this difference was 0.646684. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Concepts of print EGRA subtask across the Control (FFE + lit (Portuguese)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

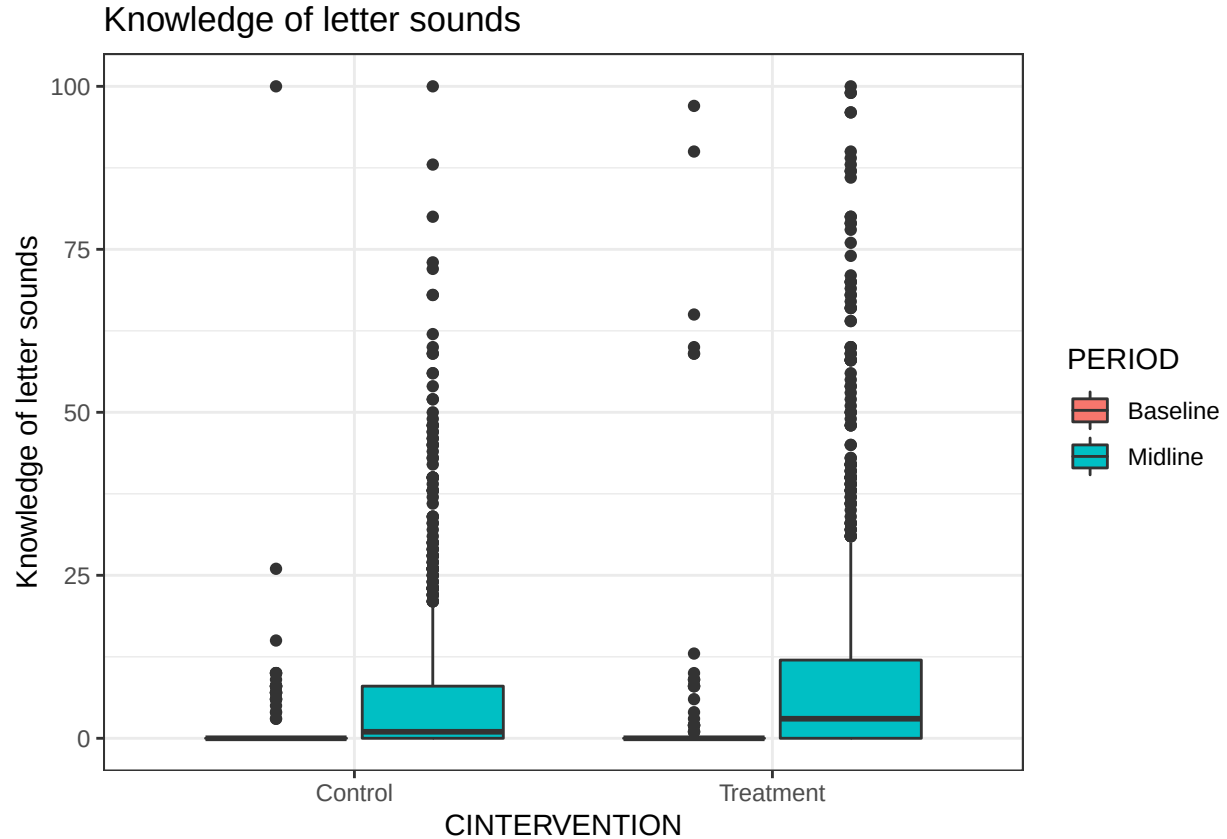
1.5 EGRA_ST5: Knowledge of letter sounds

1.5.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 21: Knowledge of letter sounds

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.278	3.271	1136	0	100	6.355	11.876	1081	0	100
Treatment	0.341	4.572	1578	0	97	8.645	14.183	1808	0	100



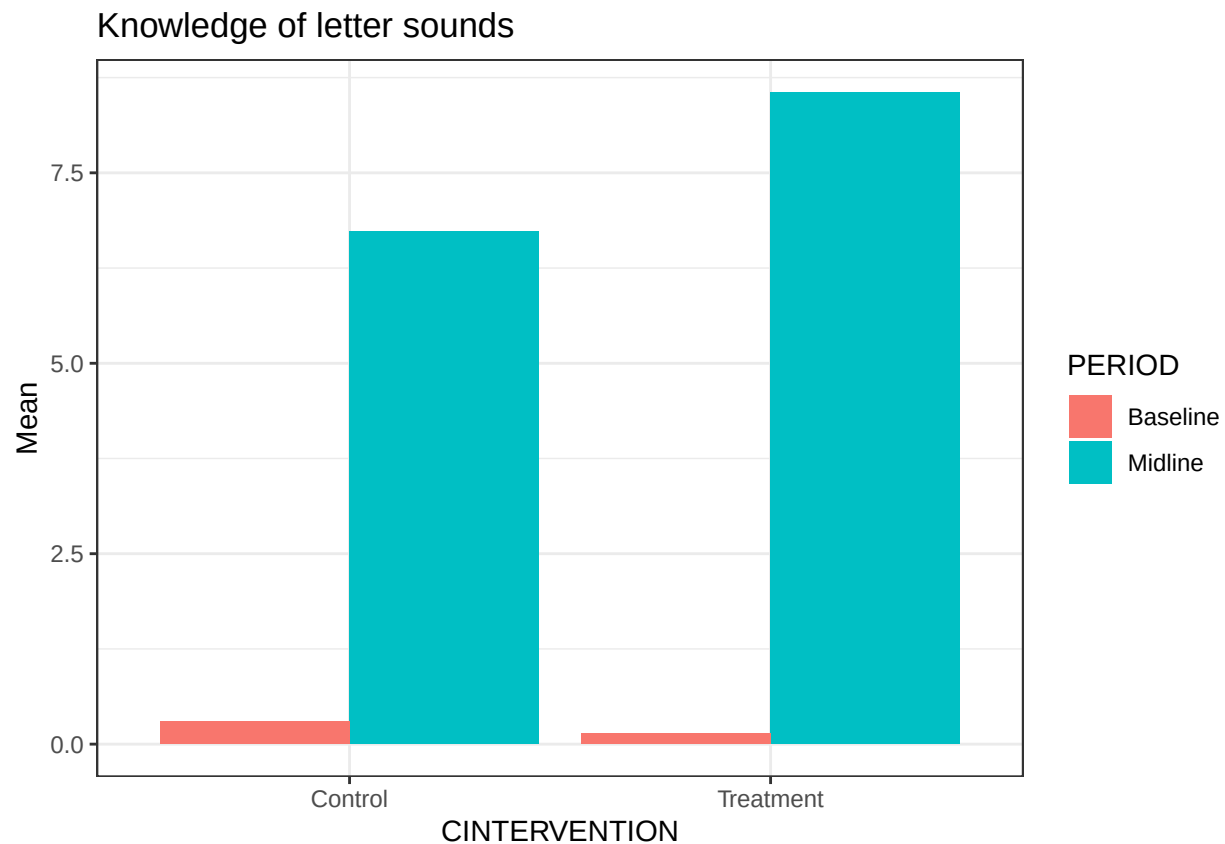


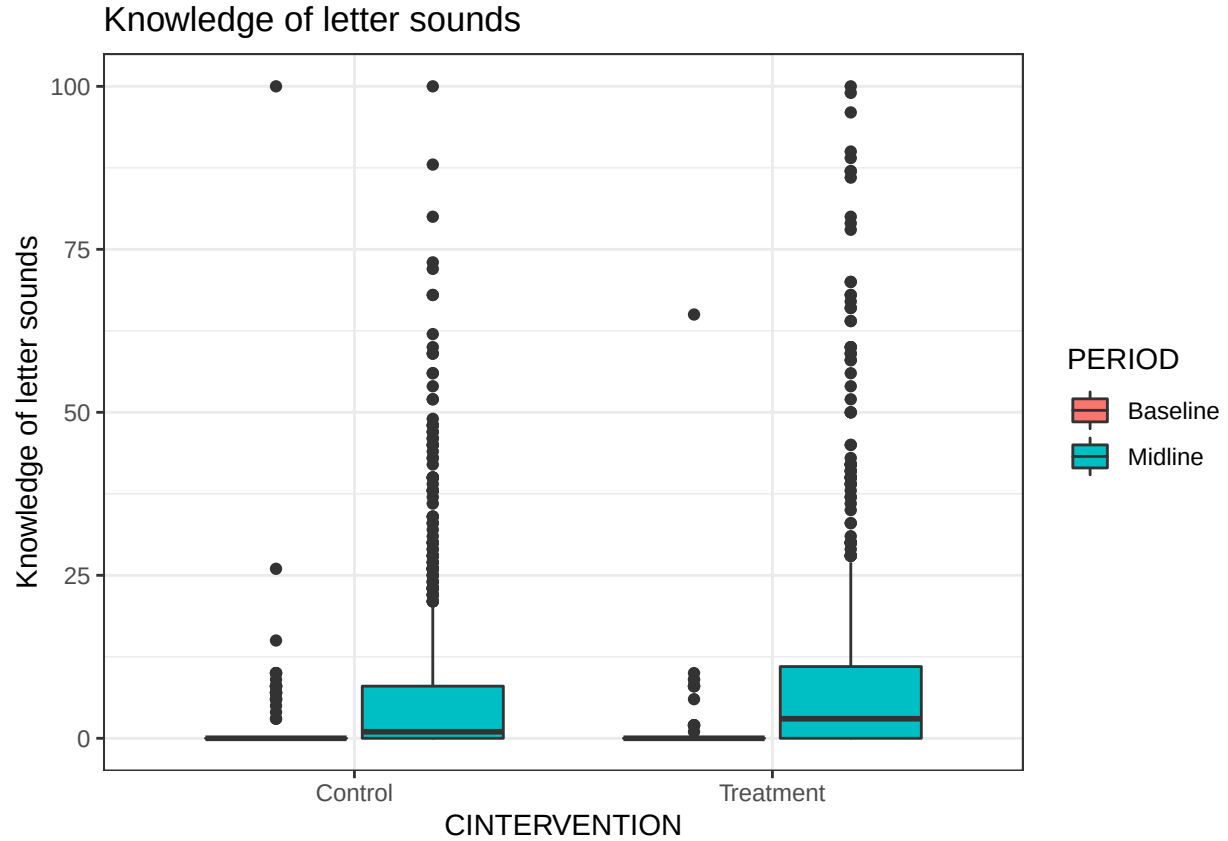
As shown in the table above, for the the Knowledge of letter sounds EGRA subtask, the mean for the Control (Comparison (all)) condition at baseline was 0.278169 (SD = 3.270762) and the mean for the Treatment (FFE + lit (all)) condition at baseline was 0.3409379 (SD = 4.572018). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at baseline was thus of 0.06276888 points. The p-value for this difference was 0.7037663. The mean for the Control (Comparison (all)) condition at midline was 6.355227 (SD = 11.87612) and the mean for the Treatment (FFE + lit (all)) condition at midline was 8.644912 (SD = 14.18262). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at midline was thus of 2.289685 points. The p-value for this difference was 0.04166406. The change from the baseline to the midline of 6.077058 points for the Control (Comparison (all)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (all)) can be assessed against. The change for the Treatment (FFE + lit (all)) from the baseline to the midline was of 8.303974 points. Consequently, the change for the Control (Comparison (all)) relative to the change for the Treatment (FFE + lit (all)) condition was of 2.226916 points. The p-value for this difference was 0.04633249. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Knowledge of letter sounds EGRA subtask across the Control (Comparison (all)) and the Treatment (FFE + lit (all)) conditions beyond that expected from sampling error.

1.5.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 22: Knowledge of letter sounds

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.298	3.404	1047	0	100	6.726	12.243	972	0	100
Treatment	0.137	2.145	1040	0	65	8.559	14.534	1047	0	100



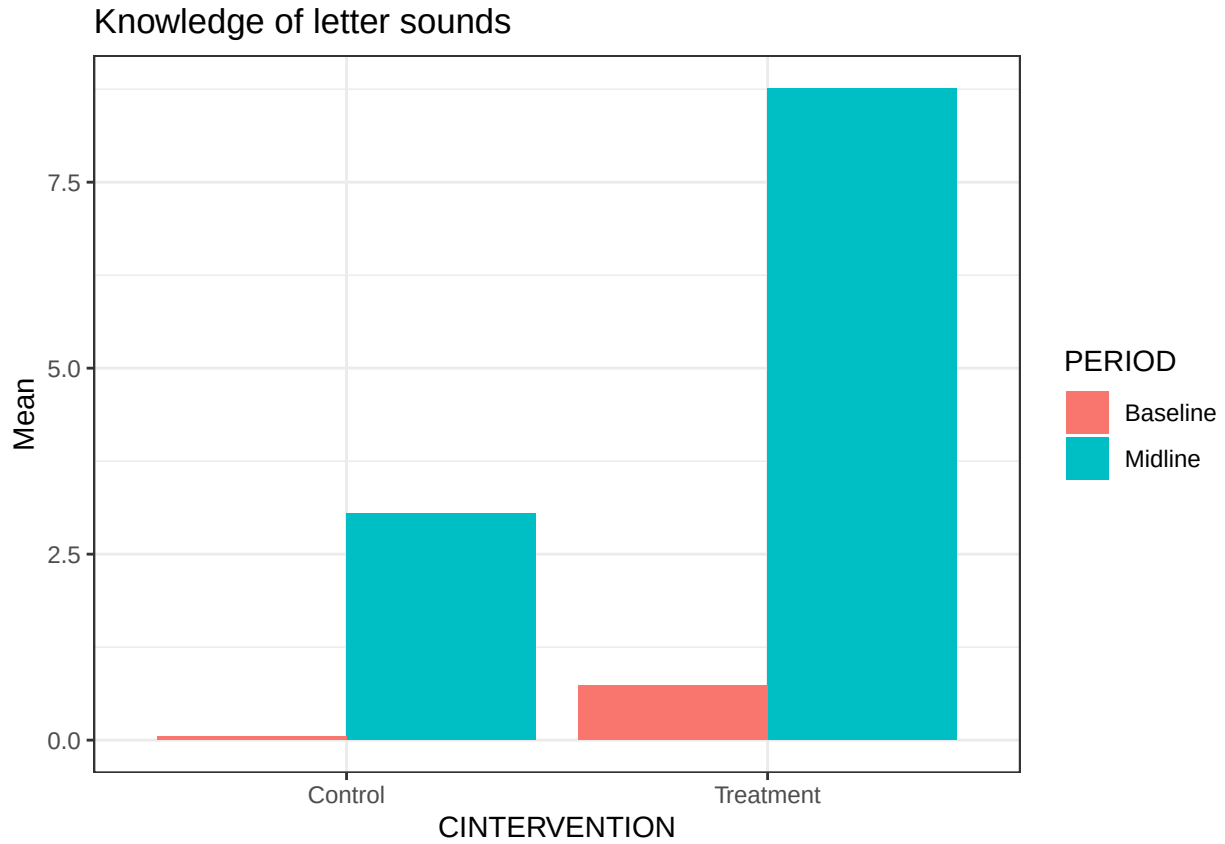


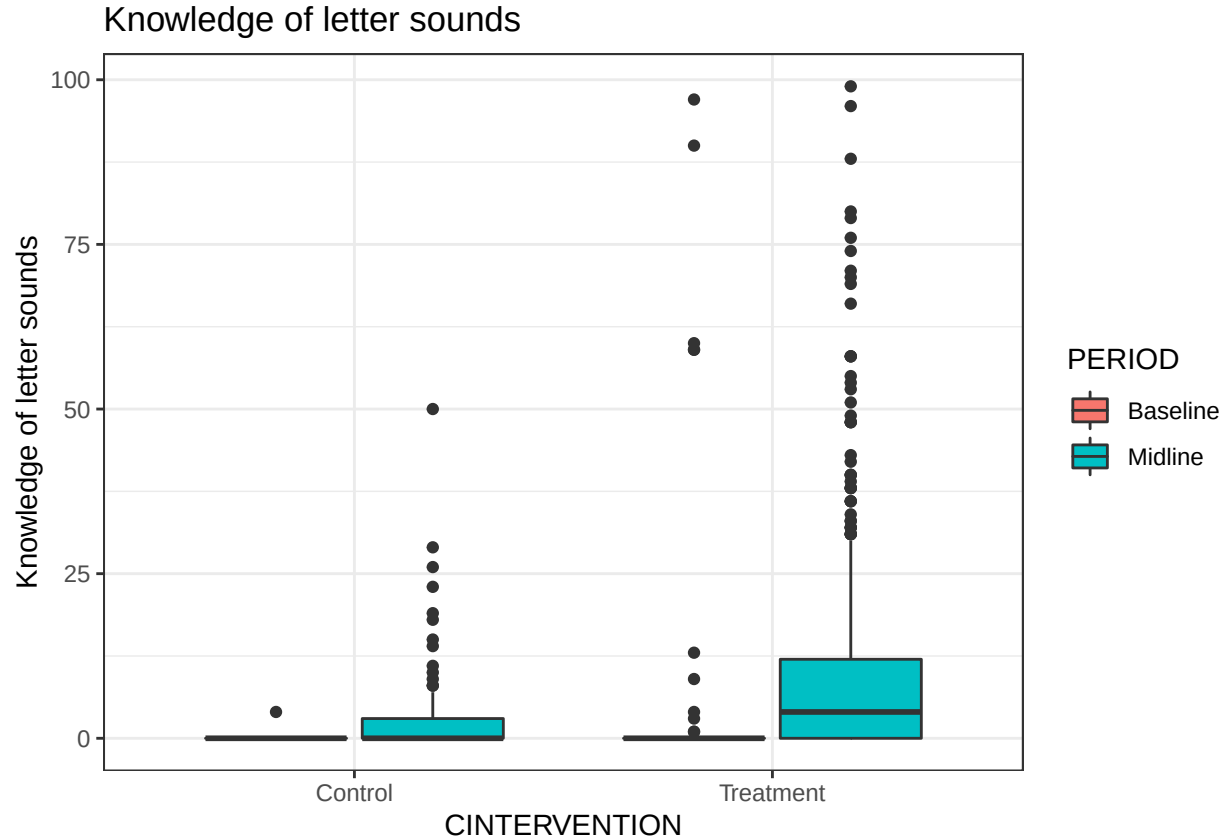
As shown in the table above, for the the Knowledge of letter sounds EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at baseline was 0.2979943 (SD = 3.404112) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.1365385 (SD = 2.145033). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.1614558 points. The p-value for this difference was 0.2308036. The mean for the Control (Comparison (Portuguese)) condition at midline was 6.726337 (SD = 12.24287) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 8.558739 (SD = 14.53446). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 1.832402 points. The p-value for this difference was 0.1684851. The change from the baseline to the midline of 6.428343 points for the Control (Comparison (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 8.422201 points. Consequently, the change for the Control (Comparison (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 1.993858 points. The p-value for this difference was 0.1320772. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Knowledge of letter sounds EGRA subtask across the Control (Comparison (Portuguese)) and the Treatment (FFE + lit (Portuguese)) conditions beyond that expected from sampling error.

1.5.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 23: Knowledge of letter sounds

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.045	0.424	89	0	4	3.046	7.108	109	0	50
Treatment	0.736	7.228	538	0	97	8.763	13.692	761	0	99



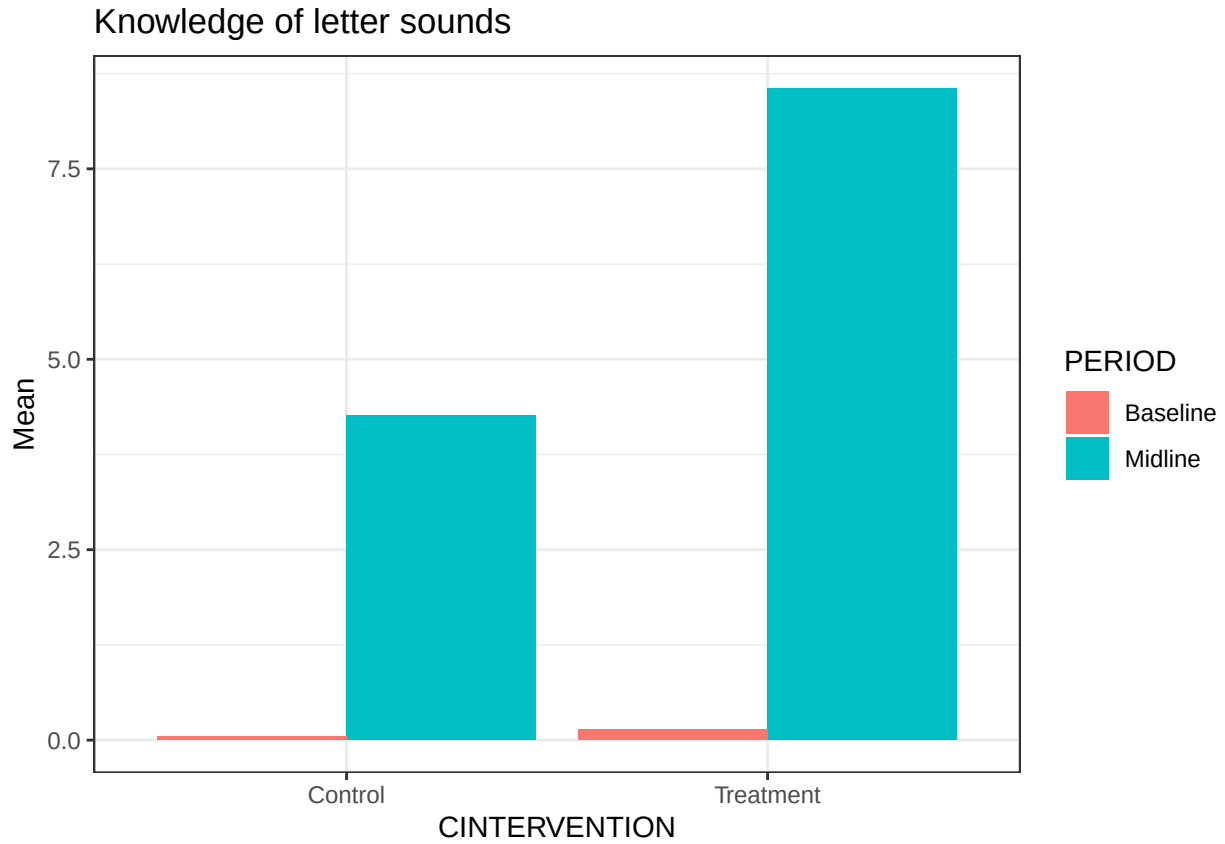


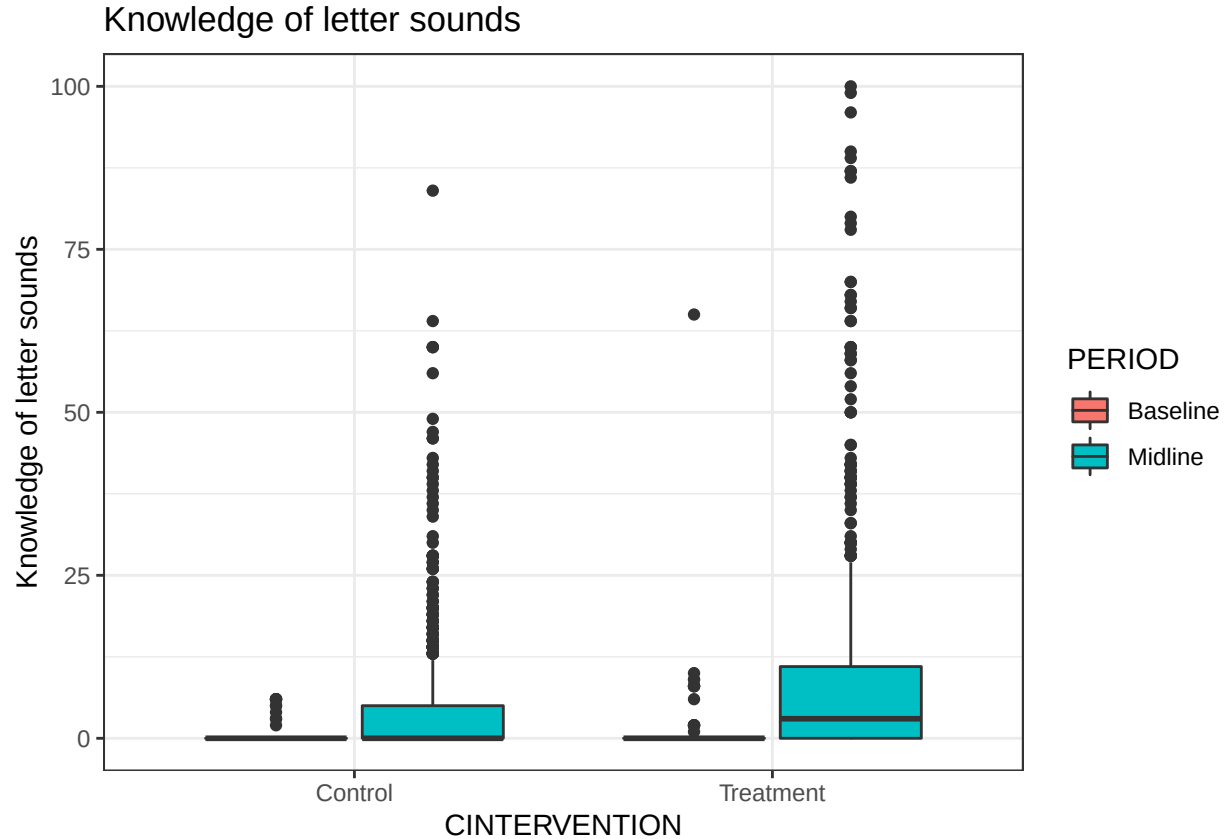
As shown in the table above, for the the Knowledge of letter sounds EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at baseline was 0.04494382 (SD = 0.4239992) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.7360595 (SD = 7.228193). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.6911157 points. The p-value for this difference was 0.04152365. The mean for the Control (Comparison (Bilingual)) condition at midline was 3.045872 (SD = 7.10814) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 8.763469 (SD = 13.69236). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 5.717598 points. The p-value for this difference was 4.525924e-05. The change from the baseline to the midline of 3.000928 points for the Control (Comparison (Bilingual)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 8.02741 points. Consequently, the change for the Control (Comparison (Bilingual)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 5.026482 points. The p-value for this difference was 0.0004872545. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Bilingual)) when compared to the Control (Comparison (Bilingual)) condition. This provides evidence that the Treatment (FFE + lit (Bilingual)) (or some other unobserved process) impacted on the Knowledge of letter sounds EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

1.5.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 24: Knowledge of letter sounds

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.052	0.514	1007	0	6	4.263	8.859	952	0	84
Treatment	0.137	2.145	1040	0	65	8.559	14.534	1047	0	100



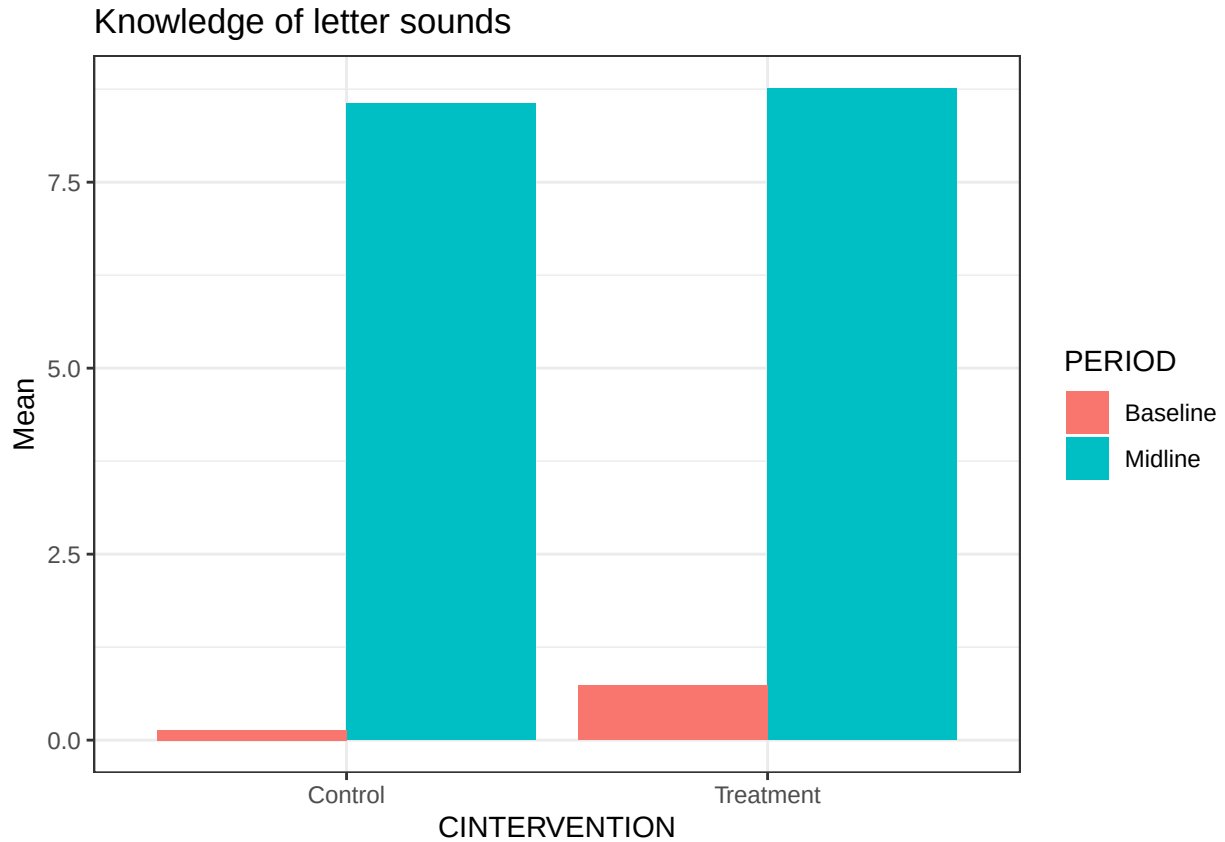


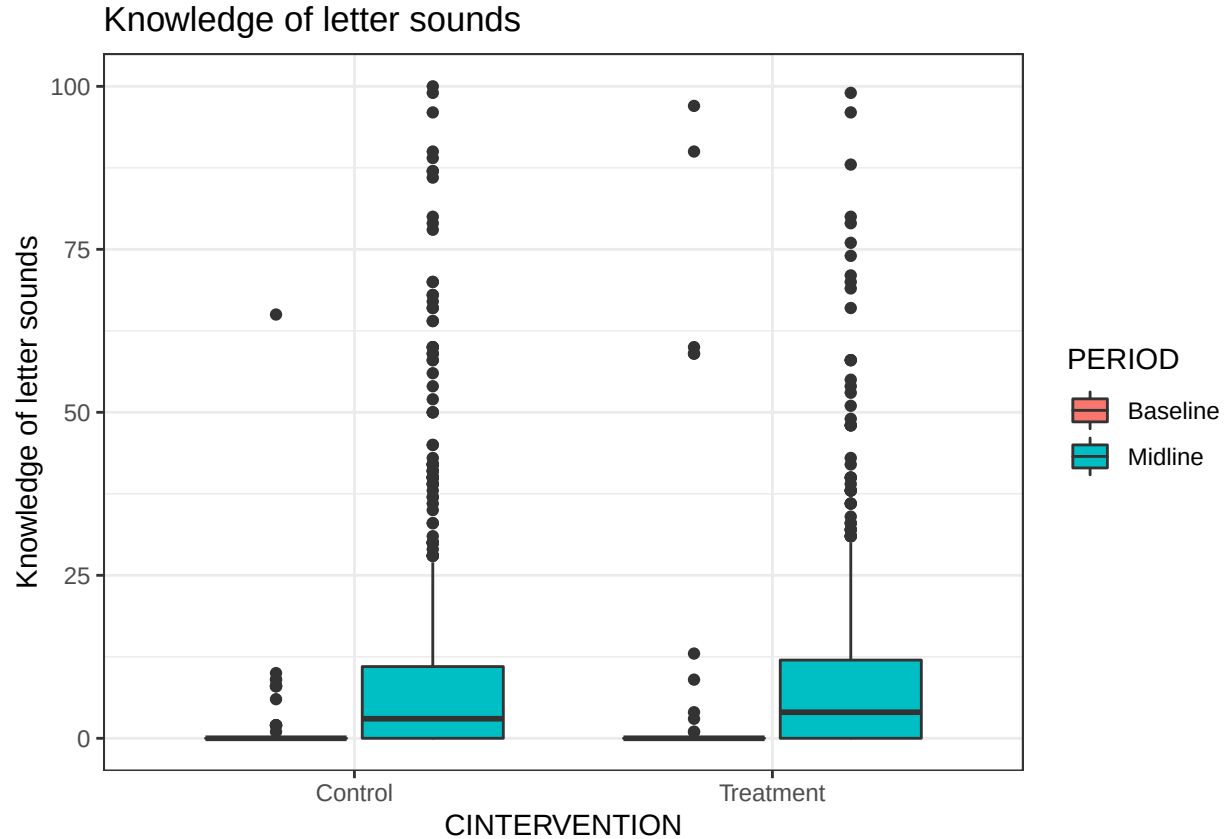
As shown in the table above, for the the Knowledge of letter sounds EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at baseline was 0.05163853 (SD = 0.5135488) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.1365385 (SD = 2.145033). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.08489993 points. The p-value for this difference was 0.2218159. The mean for the Control (FFE only (Portuguese)) condition at midline was 4.262605 (SD = 8.85853) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 8.558739 (SD = 14.53446). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 4.296134 points. The p-value for this difference was 6.7168e-05. The change from the baseline to the midline of 4.210967 points for the Control (FFE only (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 8.422201 points. Consequently, the change for the Control (FFE only (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 4.211234 points. The p-value for this difference was 9.450926e-05. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Portuguese)) when compared to the Control (FFE only (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Portuguese)) (or some other unobserved process) impacted on the Knowledge of letter sounds EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

1.5.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 25: Knowledge of letter sounds

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.137	2.145	1040	0	65	8.559	14.534	1047	0	100
Treatment	0.736	7.228	538	0	97	8.763	13.692	761	0	99





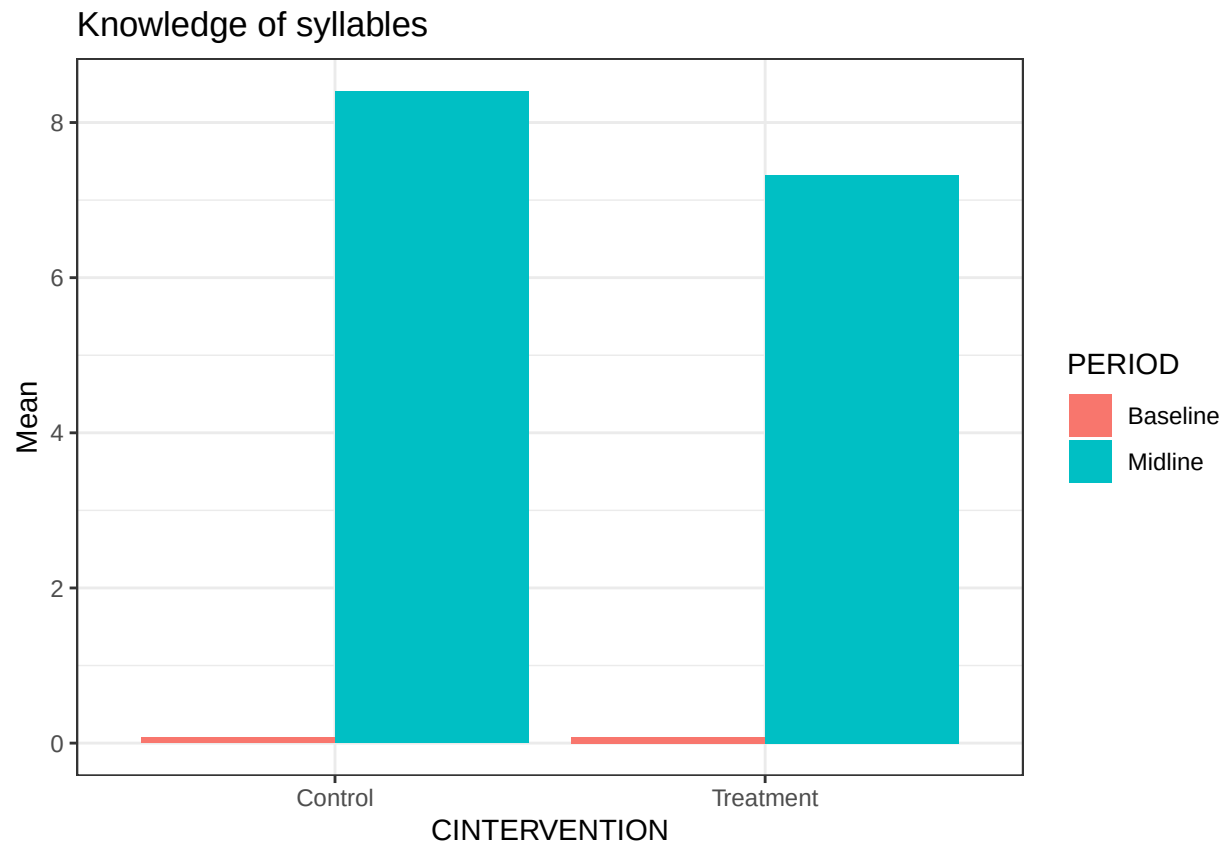
As shown in the table above, for the the Knowledge of letter sounds EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at baseline was 0.1365385 (SD = 2.145033) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.7360595 (SD = 7.228193). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.599521 points. The p-value for this difference was 0.08007998. The mean for the Control (FFE + lit (Portuguese)) condition at midline was 8.558739 (SD = 14.53446) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 8.763469 (SD = 13.69236). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.2047299 points. The p-value for this difference was 0.8862875. The change from the baseline to the midline of 8.422201 points for the Control (FFE + lit (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 8.02741 points. Consequently, the change for the Control (FFE + lit (Portuguese)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of -0.3947912 points. The p-value for this difference was 0.7887049. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Knowledge of letter sounds EGRA subtask across the Control (FFE + lit (Portuguese)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

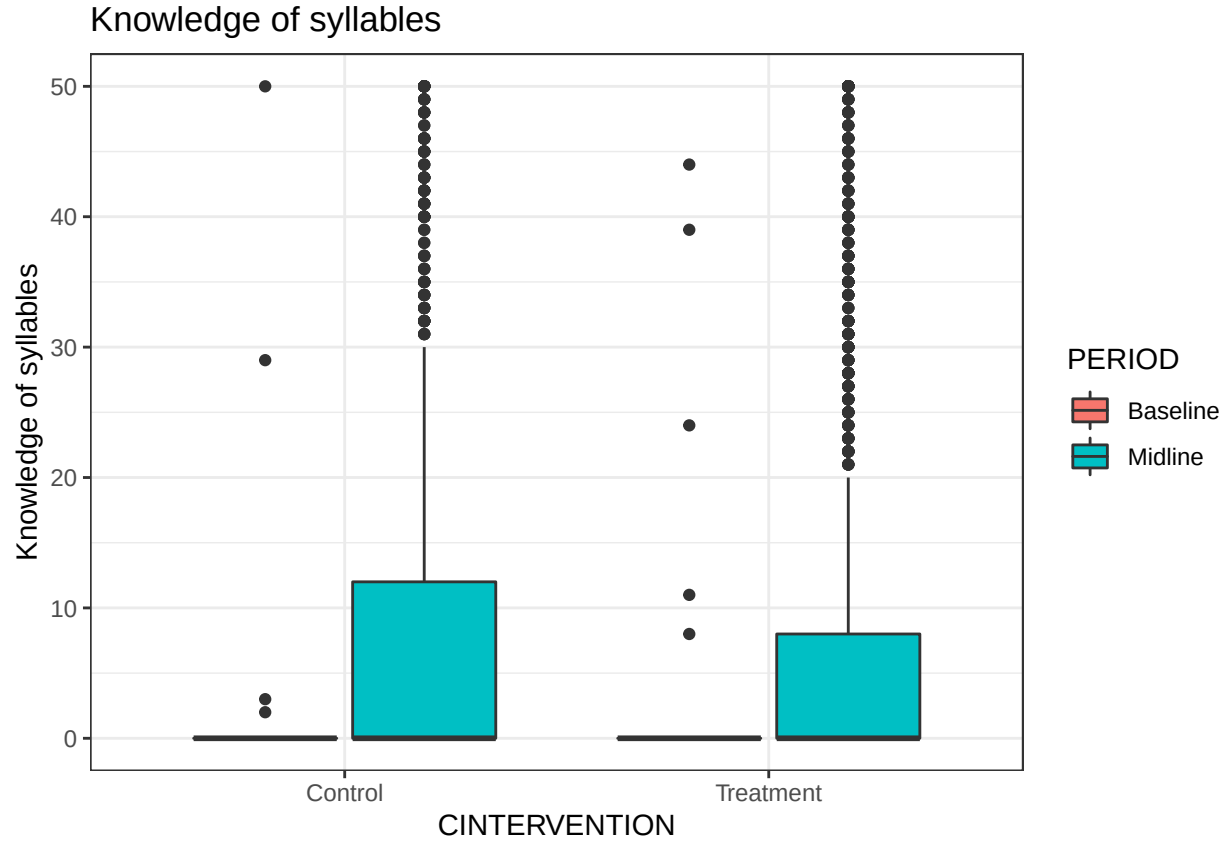
1.6 EGRA_ST6: Knowledge of syllables

1.6.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 26: Knowledge of syllables

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.074	1.717	1136	0	50	8.404	14.072	1081	0	50
Treatment	0.080	1.633	1578	0	44	7.323	12.905	1808	0	50



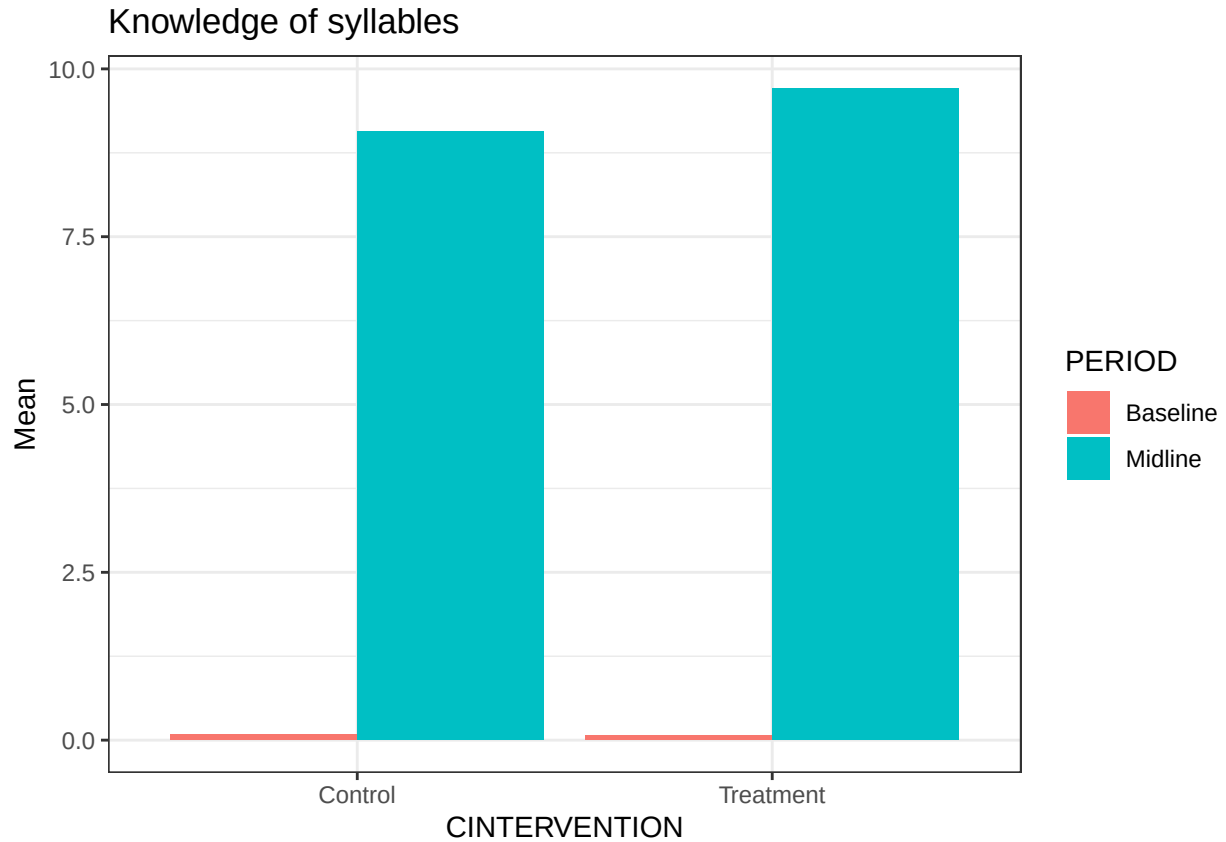


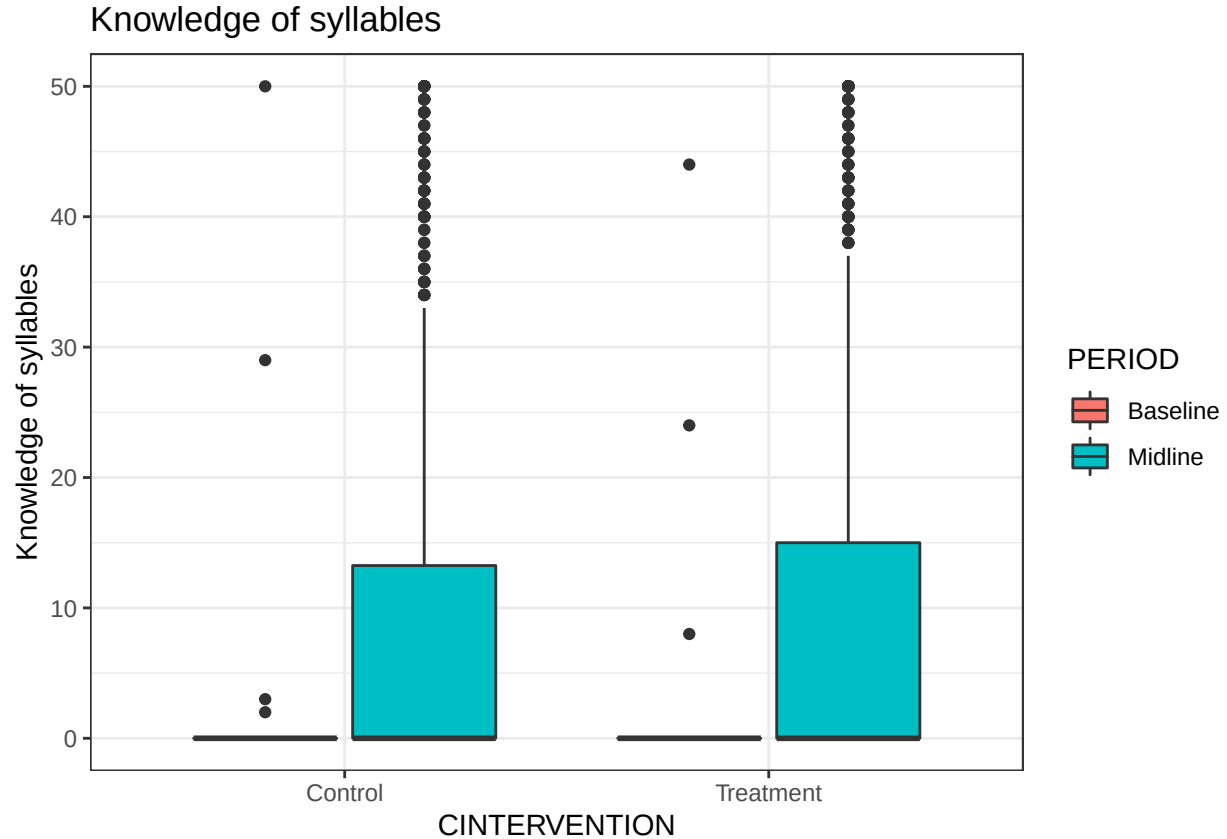
As shown in the table above, for the the Knowledge of syllables EGRA subtask, the mean for the Control (Comparison (all)) condition at baseline was 0.07394366 (SD = 1.717438) and the mean for the Treatment (FFE + lit (all)) condition at baseline was 0.07984791 (SD = 1.633499). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at baseline was thus of 0.005904247 points. The p-value for this difference was 0.9274075. The mean for the Control (Comparison (all)) condition at midline was 8.404255 (SD = 14.07192) and the mean for the Treatment (FFE + lit (all)) condition at midline was 7.323009 (SD = 12.90529). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at midline was thus of 1.081246 points. The p-value for this difference was 0.3311634. The change from the baseline to the midline of 8.330312 points for the Control (Comparison (all)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (all)) can be assessed against. The change for the Treatment (FFE + lit (all)) from the baseline to the midline was of 7.243161 points. Consequently, the change for the Control (Comparison (all)) relative to the change for the Treatment (FFE + lit (all)) condition was of -1.087151 points. The p-value for this difference was 0.3283146. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Knowledge of syllables EGRA subtask across the Control (Comparison (all)) and the Treatment (FFE + lit (all)) conditions beyond that expected from sampling error.

1.6.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 27: Knowledge of syllables

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.080	1.789	1047	0	50	9.062	14.499	972	0	50
Treatment	0.073	1.573	1040	0	44	9.713	14.938	1047	0	50



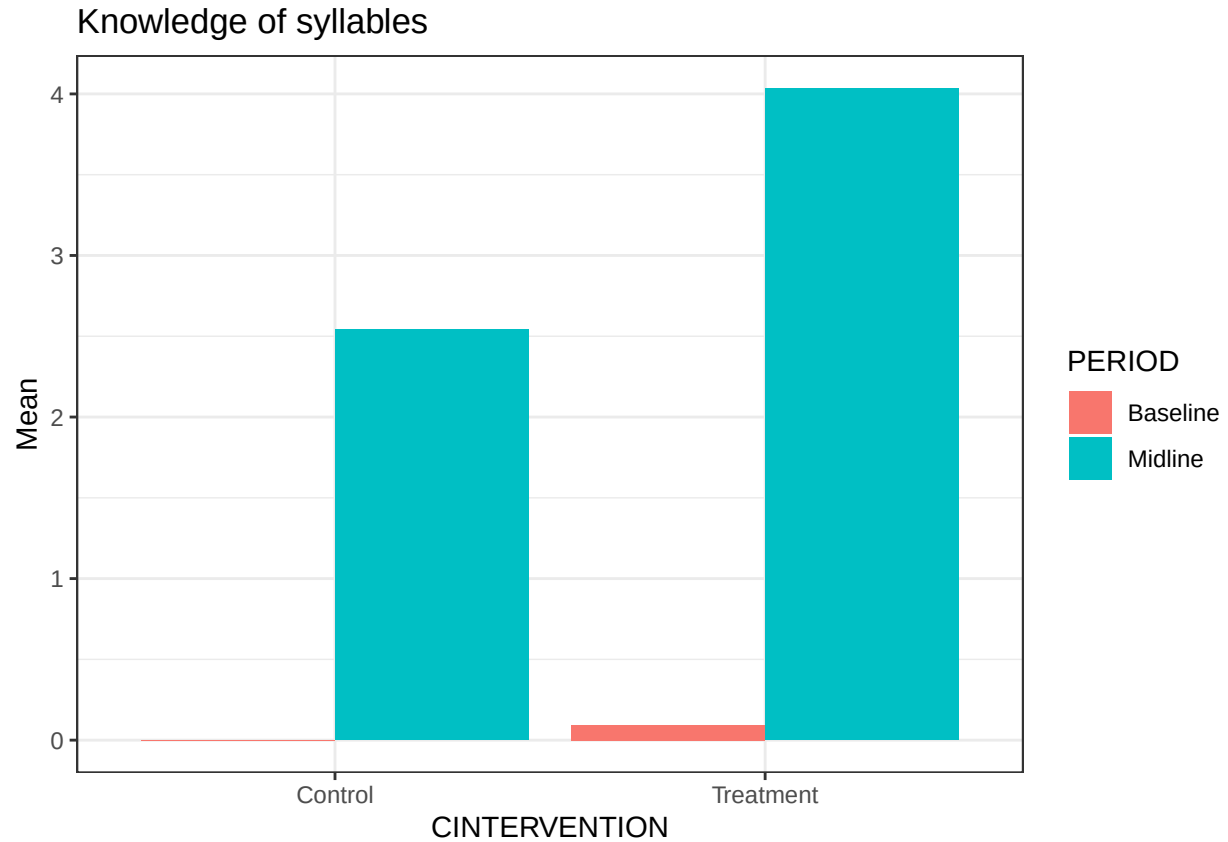


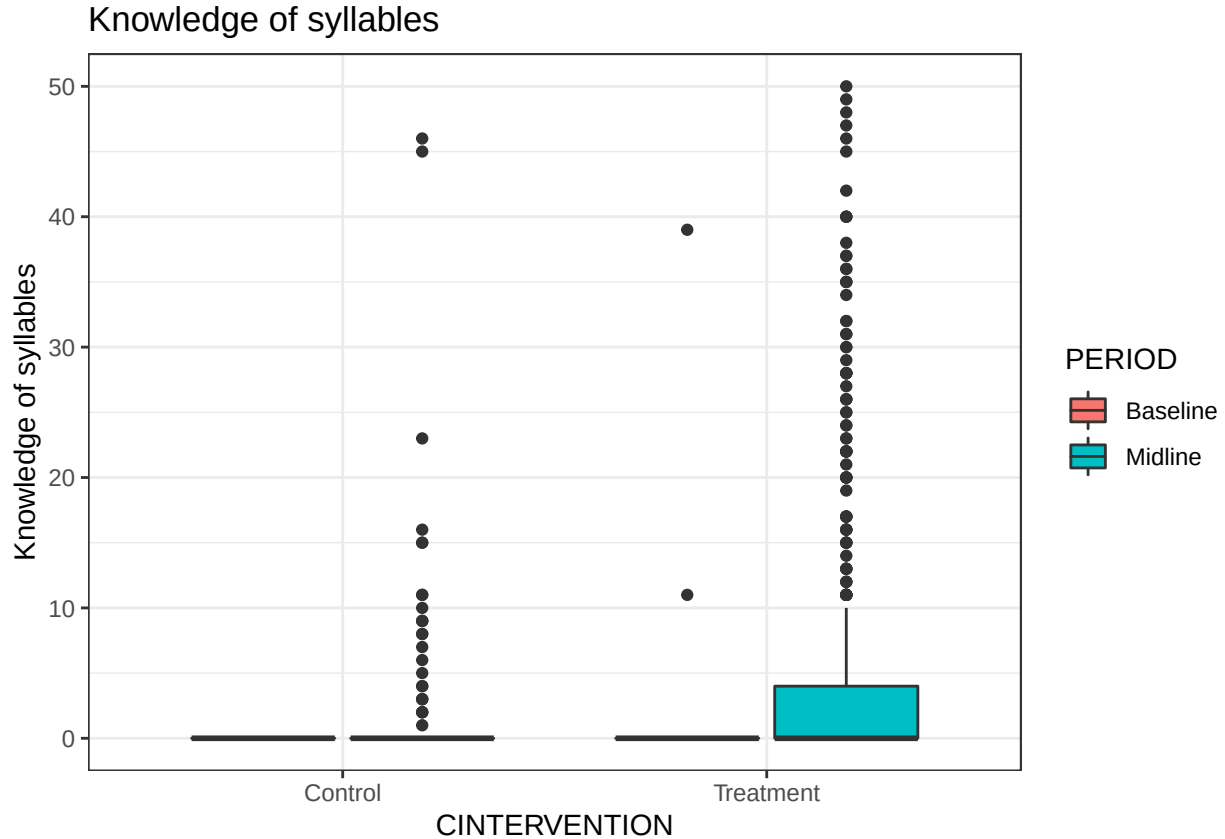
As shown in the table above, for the the Knowledge of syllables EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at baseline was 0.08022923 (SD = 1.788871) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.07307692 (SD = 1.572883). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.007152303 points. The p-value for this difference was 0.9222329. The mean for the Control (Comparison (Portuguese)) condition at midline was 9.061728 (SD = 14.49906) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 9.712512 (SD = 14.93781). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.6507835 points. The p-value for this difference was 0.6286879. The change from the baseline to the midline of 8.981499 points for the Control (Comparison (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 9.639435 points. Consequently, the change for the Control (Comparison (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 0.6579358 points. The p-value for this difference was 0.6247483. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Knowledge of syllables EGRA subtask across the Control (Comparison (Portuguese)) and the Treatment (FFE + lit (Portuguese)) conditions beyond that expected from sampling error.

1.6.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 28: Knowledge of syllables

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.000	0.000	89	0	0	2.541	7.180	109	0	46
Treatment	0.093	1.746	538	0	39	4.035	8.378	761	0	50



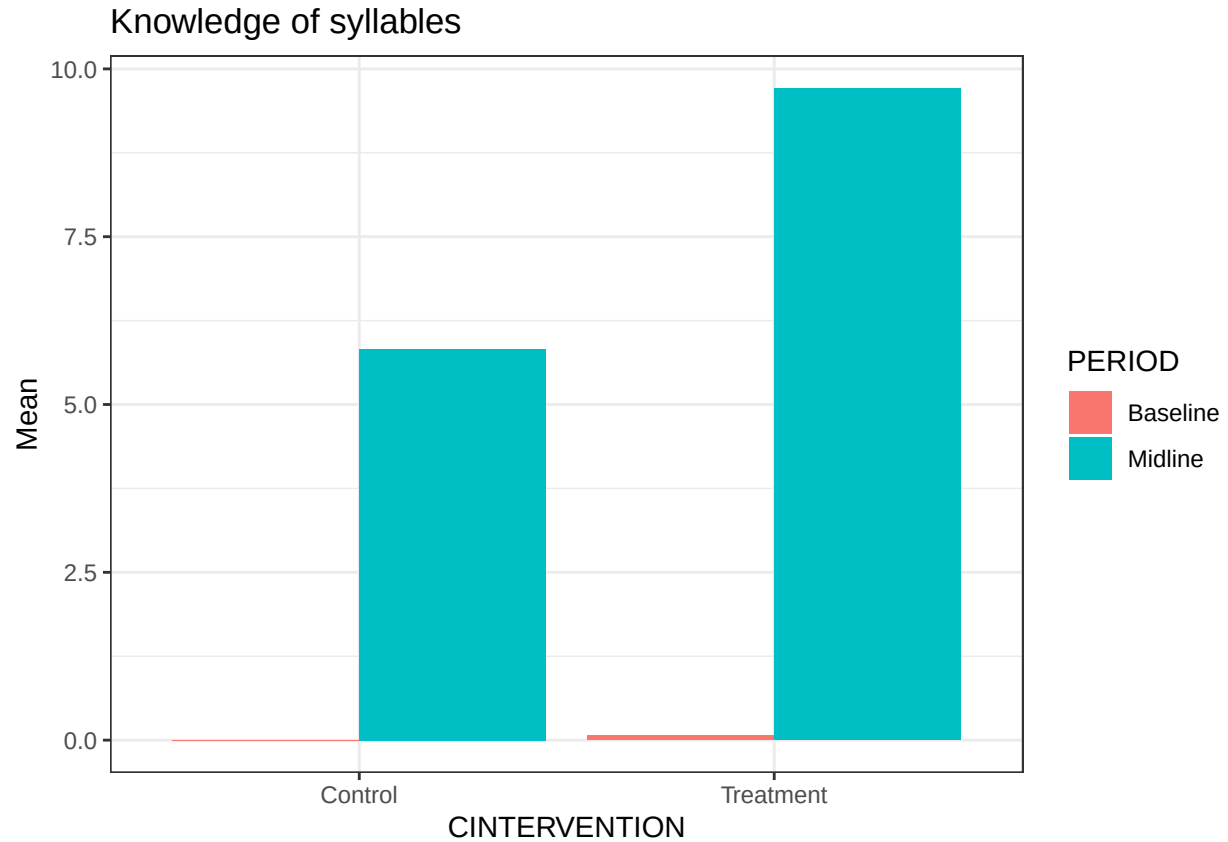


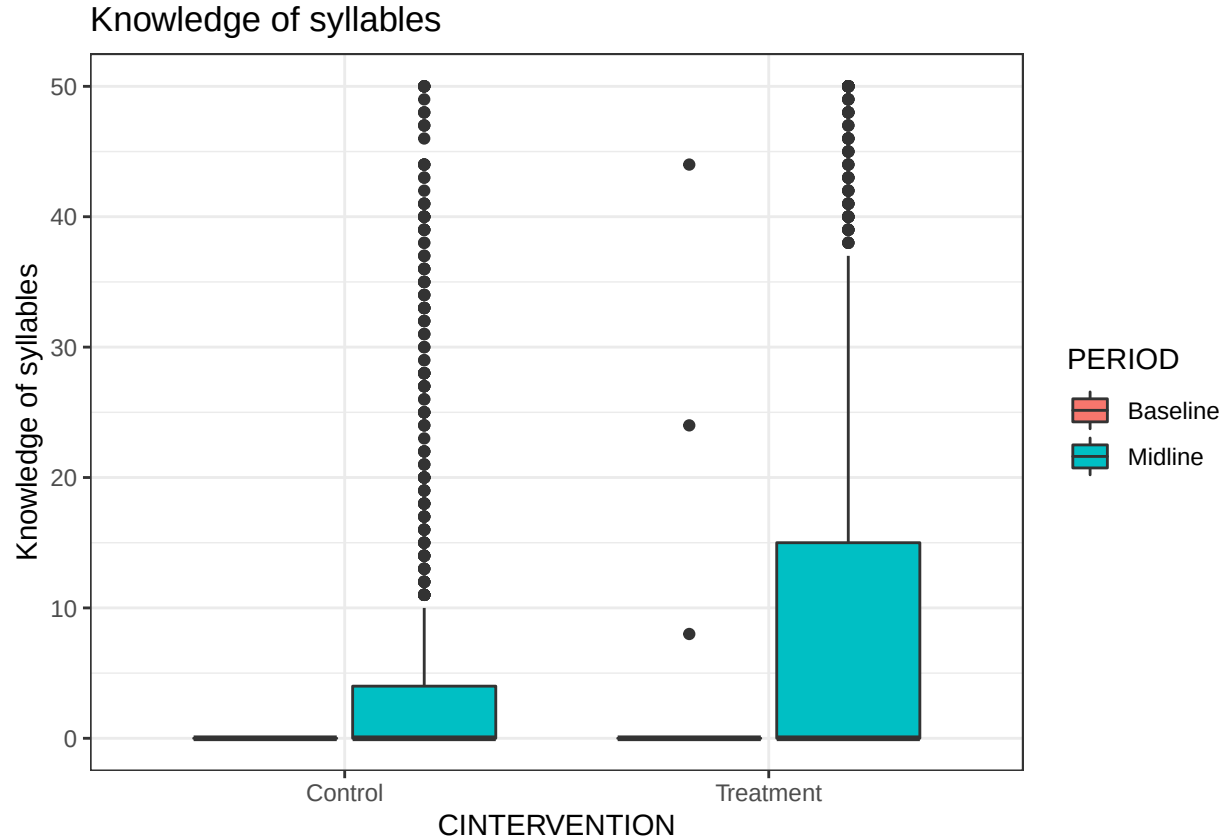
As shown in the table above, for the the Knowledge of syllables EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at baseline was 0 (SD = 0) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.0929368 (SD = 1.74616). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.0929368 points. The p-value for this difference was 0.2070935. The mean for the Control (Comparison (Bilingual)) condition at midline was 2.541284 (SD = 7.179616) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 4.03548 (SD = 8.377919). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 1.494195 points. The p-value for this difference was 0.1210532. The change from the baseline to the midline of 2.541284 points for the Control (Comparison (Bilingual)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 3.942543 points. Consequently, the change for the Control (Comparison (Bilingual)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 1.401258 points. The p-value for this difference was 0.1464555. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Knowledge of syllables EGRA subtask across the Control (Comparison (Bilingual)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

1.6.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 29: Knowledge of syllables

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.000	0.000	1007	0	0	5.828	11.903	952	0	50
Treatment	0.073	1.573	1040	0	44	9.713	14.938	1047	0	50



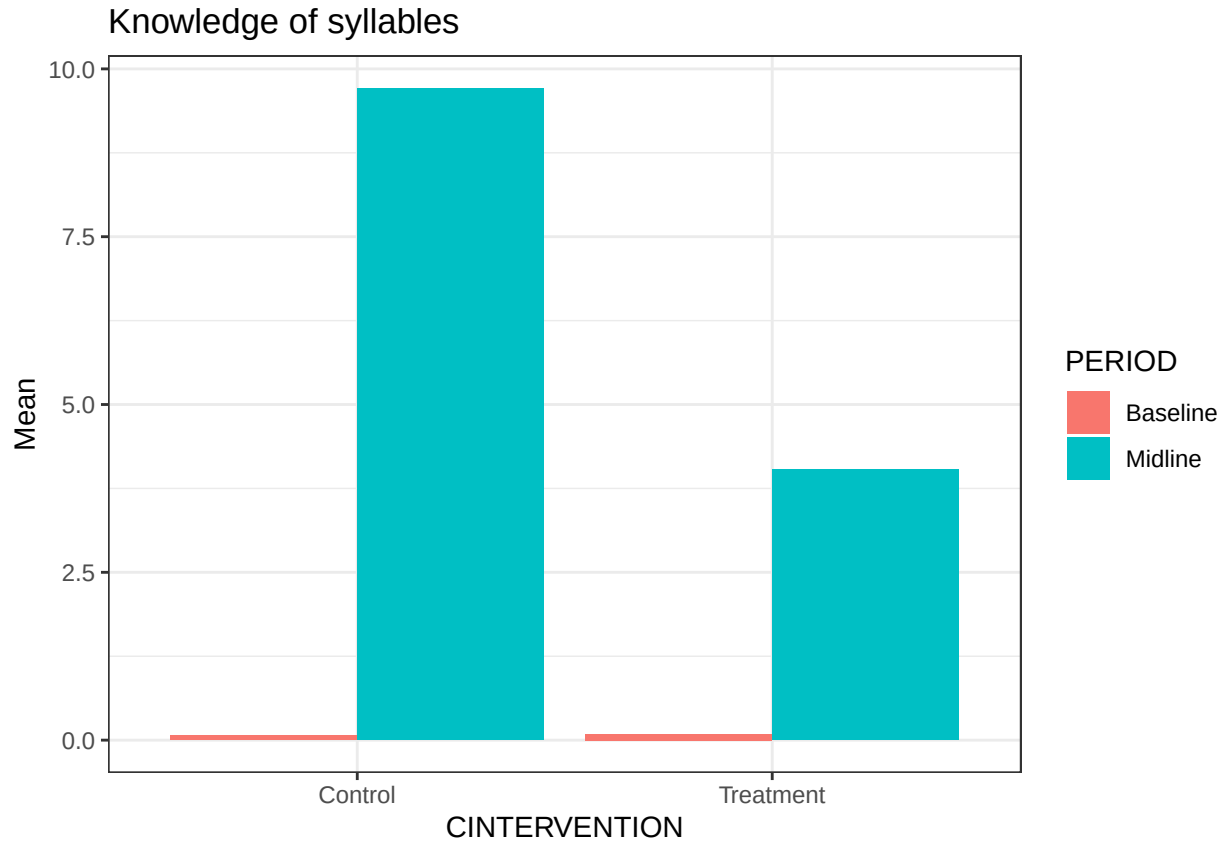


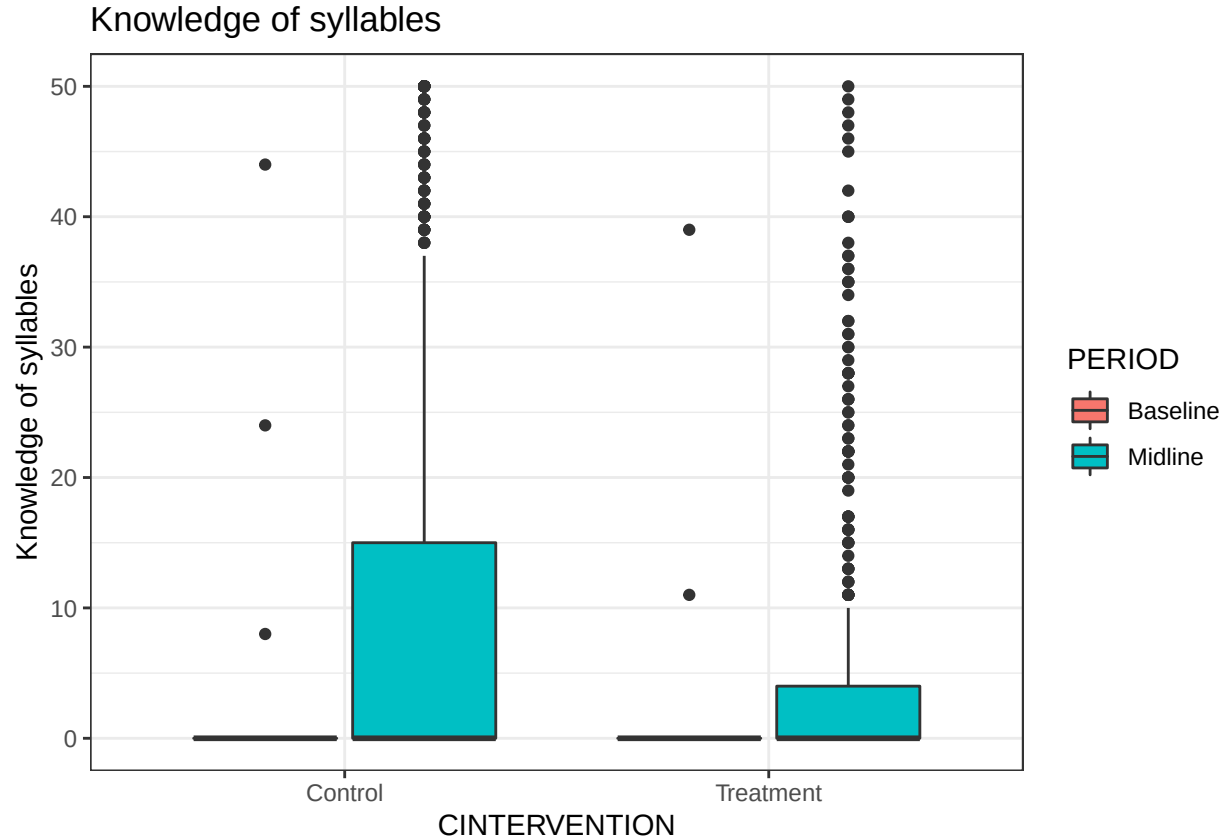
As shown in the table above, for the the Knowledge of syllables EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at baseline was 0 (SD = 0) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.07307692 (SD = 1.572883). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.07307692 points. The p-value for this difference was 0.1348755. The mean for the Control (FFE only (Portuguese)) condition at midline was 5.827731 (SD = 11.90347) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 9.712512 (SD = 14.93781). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 3.884781 points. The p-value for this difference was 0.001668309. The change from the baseline to the midline of 5.827731 points for the Control (FFE only (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 9.639435 points. Consequently, the change for the Control (FFE only (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 3.811704 points. The p-value for this difference was 0.002023932. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Portuguese)) when compared to the Control (FFE only (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Portuguese)) (or some other unobserved process) impacted on the Knowledge of syllables EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

1.6.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 30: Knowledge of syllables

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.073	1.573	1040	0	44	9.713	14.938	1047	0	50
Treatment	0.093	1.746	538	0	39	4.035	8.378	761	0	50





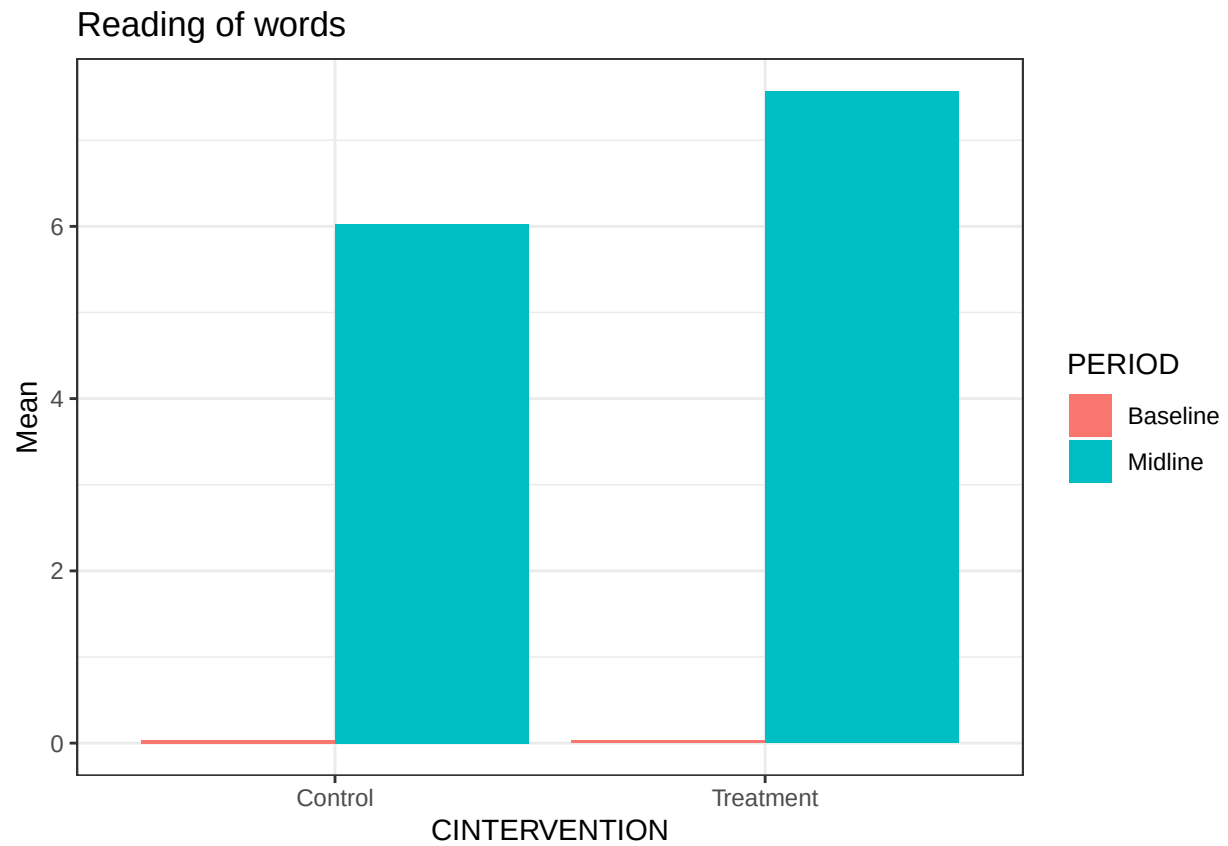
As shown in the table above, for the the Knowledge of syllables EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at baseline was 0.07307692 (SD = 1.572883) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.0929368 (SD = 1.74616). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.01985988 points. The p-value for this difference was 0.8216617. The mean for the Control (FFE + lit (Portuguese)) condition at midline was 9.712512 (SD = 14.93781) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 4.03548 (SD = 8.377919). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 5.677032 points. The p-value for this difference was 7.226164e-08. The change from the baseline to the midline of 9.639435 points for the Control (FFE + lit (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 3.942543 points. Consequently, the change for the Control (FFE + lit (Portuguese)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of -5.696892 points. The p-value for this difference was 6.542178e-08. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Bilingual)) when compared to the Control (FFE + lit (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Bilingual)) (or some other unobserved process) impacted on the Knowledge of syllables EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed. It should be pointed out that the difference in the rate of change of the Treatment (FFE + lit (Bilingual)) relative to the Control (FFE + lit (Portuguese)) was negative and thus suggests that students in the Control (FFE + lit (Portuguese)) schools performed significantly better than those in the Treatment (FFE + lit (Bilingual)) .

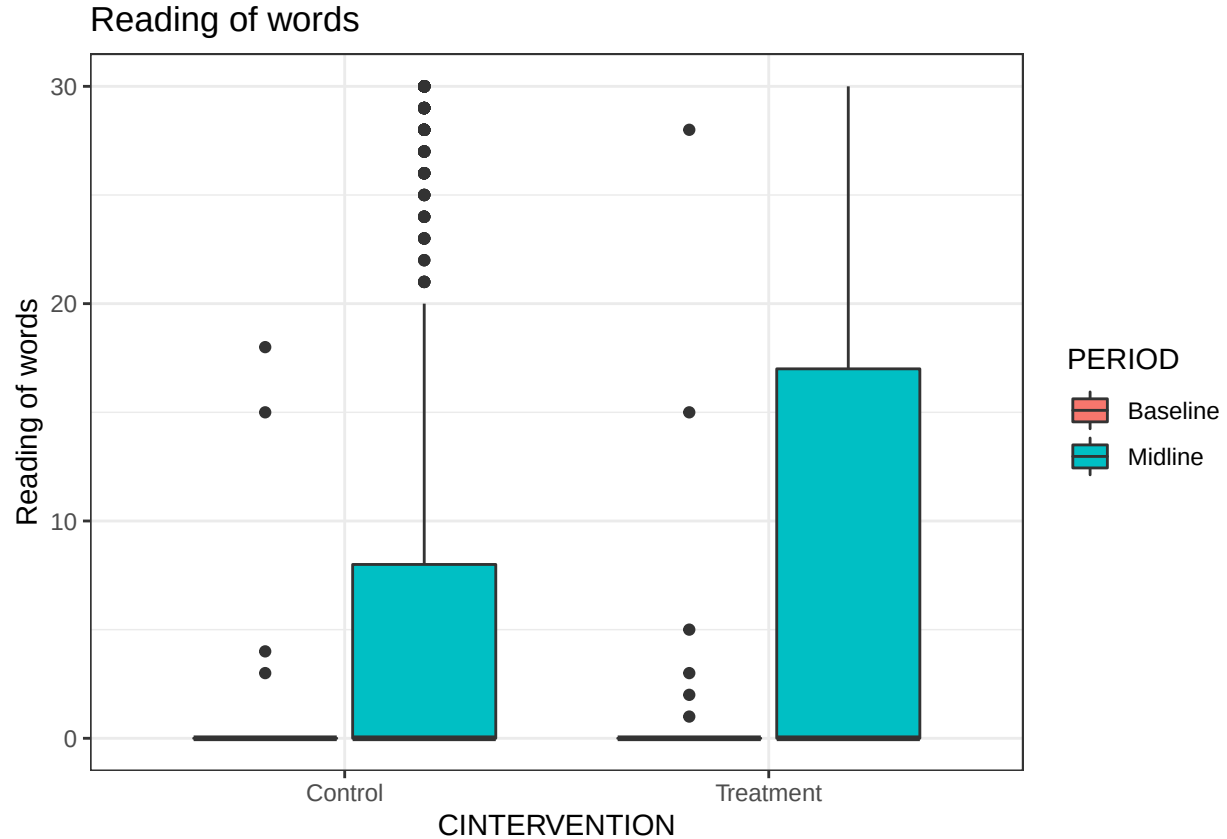
1.7 EGRA_ST7: Reading of words

1.7.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 31: Reading of words

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.035	0.710	1136	0	18	6.027	10.347	1081	0	30
Treatment	0.034	0.814	1578	0	28	7.570	11.225	1808	0	30



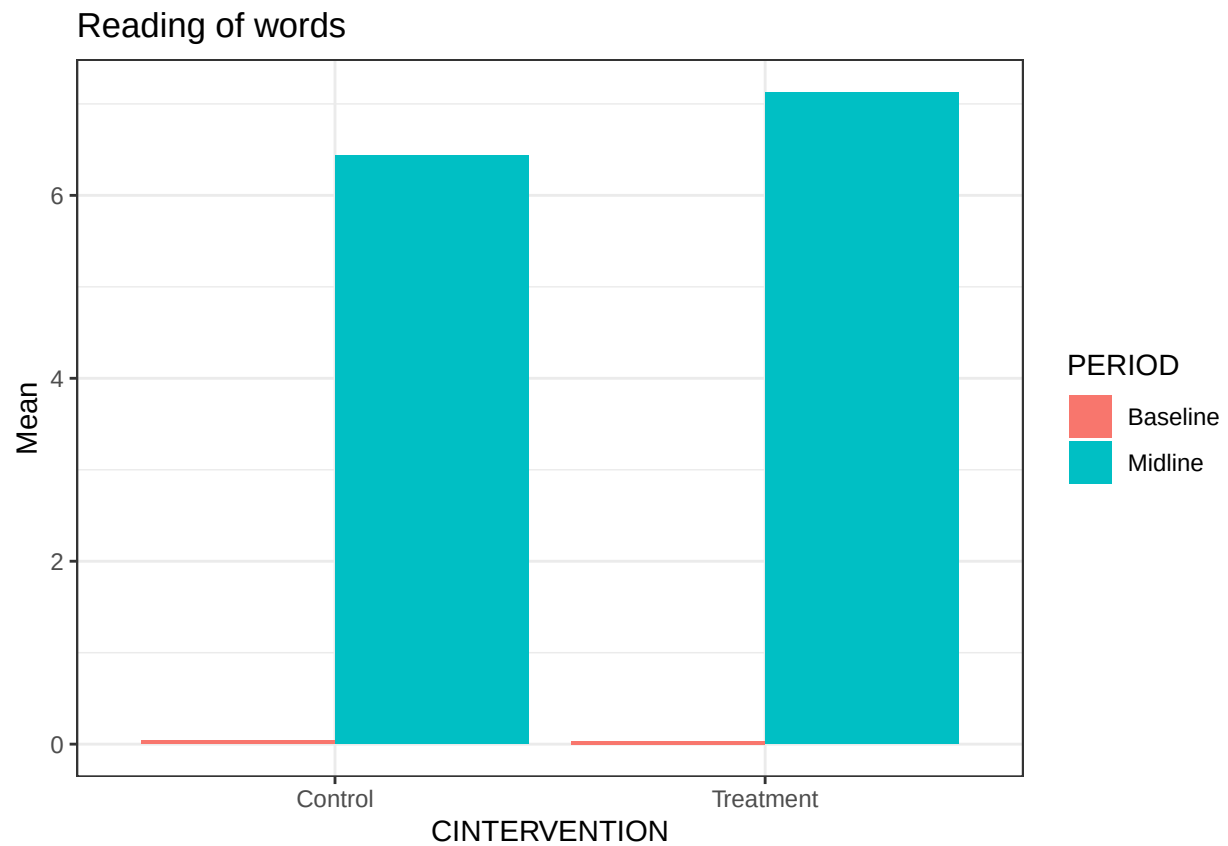


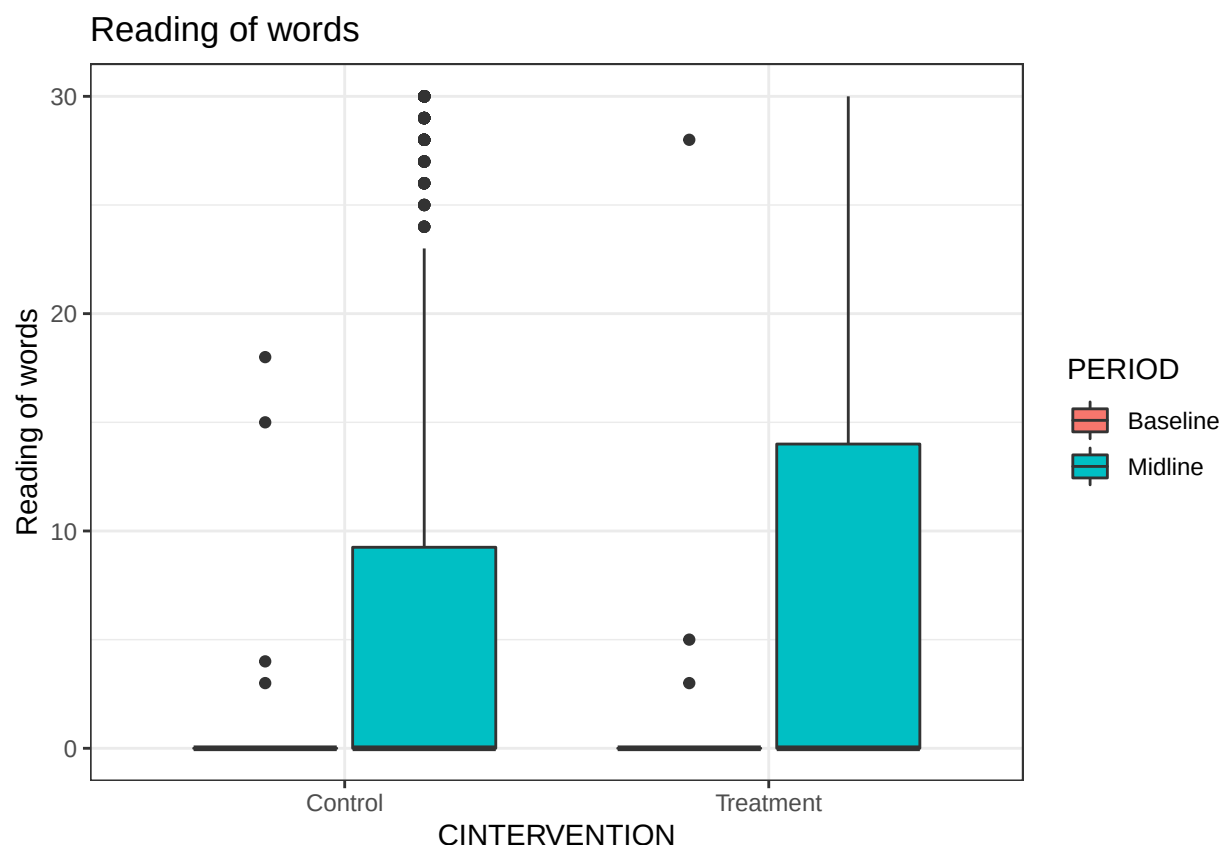
As shown in the table above, for the the Reading of words EGRA subtask, the mean for the Control (Comparison (all)) condition at baseline was 0.03521127 (SD = 0.7102717) and the mean for the Treatment (FFE + lit (all)) condition at baseline was 0.03422053 (SD = 0.8144821). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at baseline was thus of 0.0009907353 points. The p-value for this difference was 0.9728826. The mean for the Control (Comparison (all)) condition at midline was 6.026827 (SD = 10.3466) and the mean for the Treatment (FFE + lit (all)) condition at midline was 7.56969 (SD = 11.22461). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at midline was thus of 1.542863 points. The p-value for this difference was 0.06638155. The change from the baseline to the midline of 5.991616 points for the Control (Comparison (all)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (all)) can be assessed against. The change for the Treatment (FFE + lit (all)) from the baseline to the midline was of 7.53547 points. Consequently, the change for the Control (Comparison (all)) relative to the change for the Treatment (FFE + lit (all)) condition was of 1.543854 points. The p-value for this difference was 0.06591212. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Reading of words EGRA subtask across the Control (Comparison (all)) and the Treatment (FFE + lit (all)) conditions beyond that expected from sampling error.

1.7.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 32: Reading of words

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.038	0.740	1047	0	18	6.438	10.635	972	0	30
Treatment	0.035	0.887	1040	0	28	7.128	11.039	1047	0	30



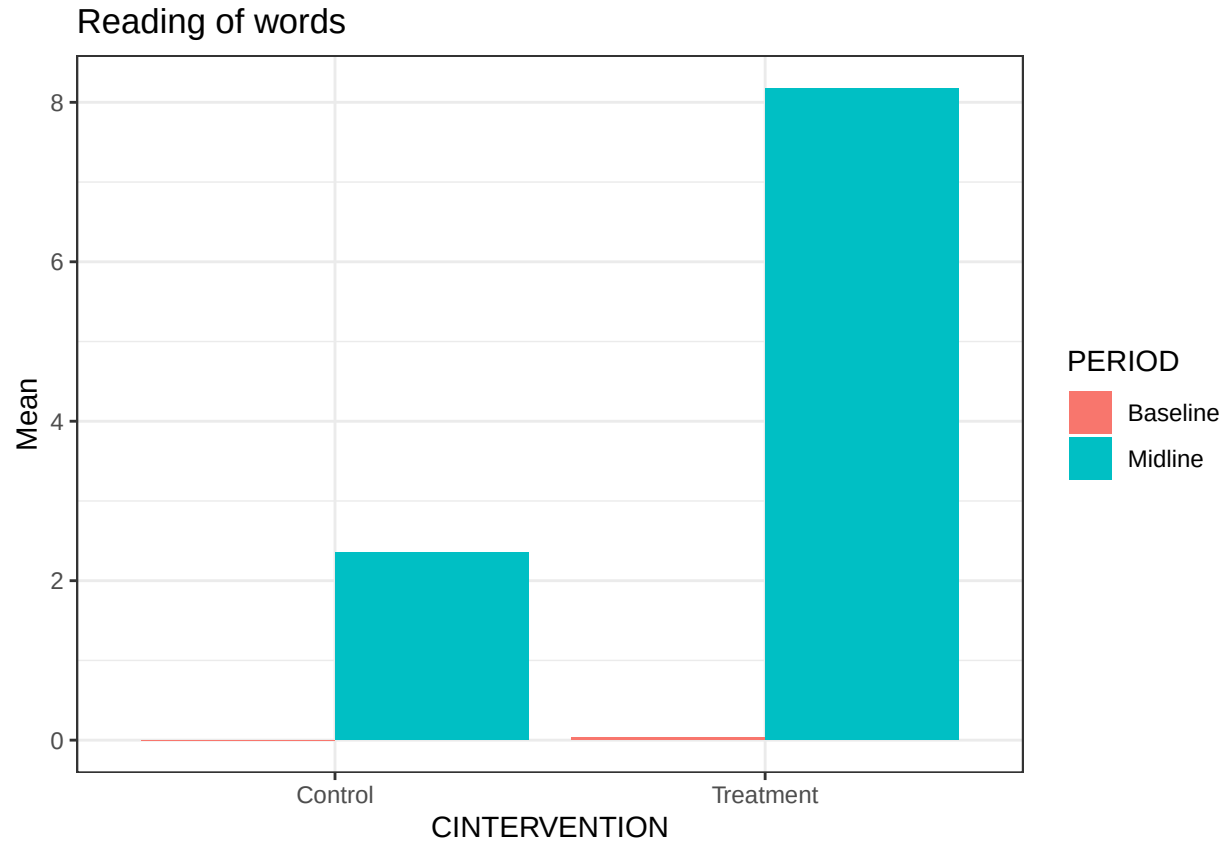


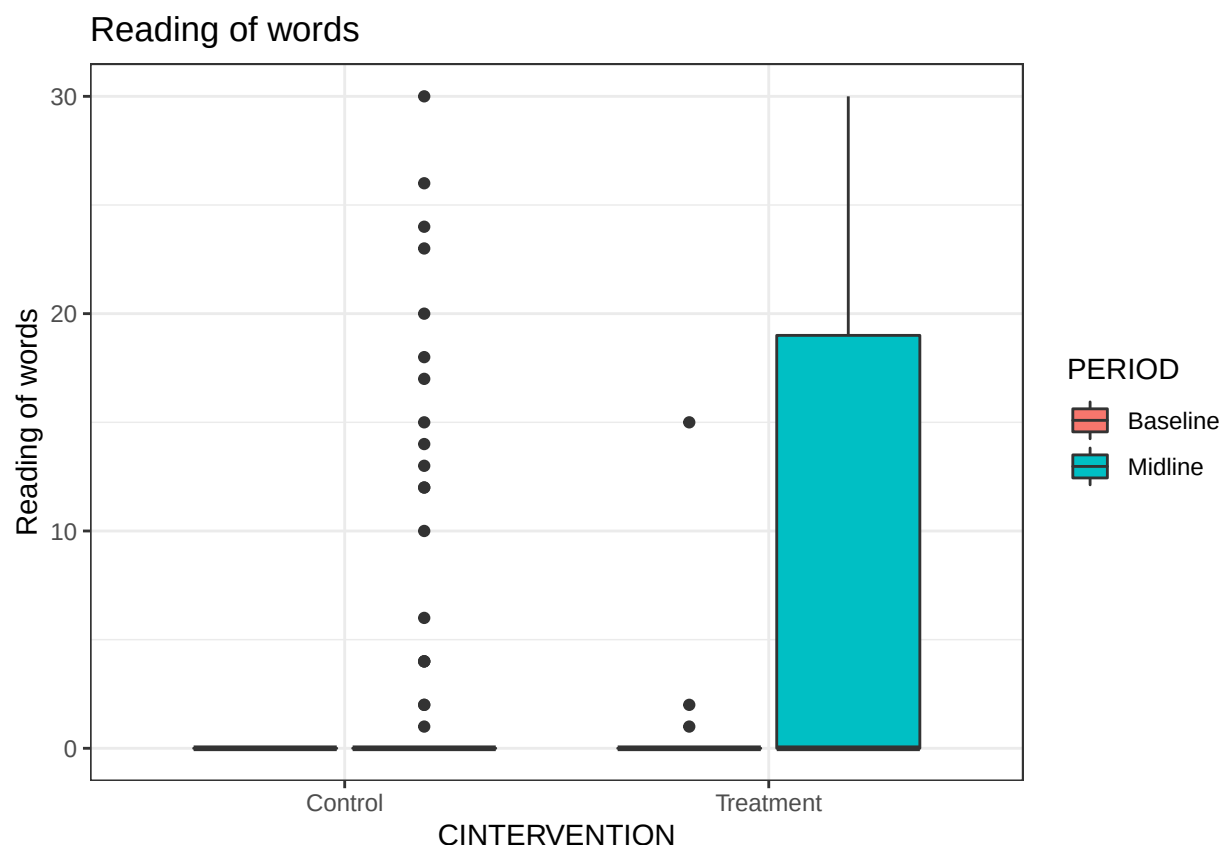
As shown in the table above, for the the Reading of words EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at baseline was 0.03820439 (SD = 0.7397947) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.03461538 (SD = 0.8866206). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.003589009 points. The p-value for this difference was 0.9193661. The mean for the Control (Comparison (Portuguese)) condition at midline was 6.438272 (SD = 10.6351) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 7.127985 (SD = 11.0389). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.6897131 points. The p-value for this difference was 0.4770892. The change from the baseline to the midline of 6.400067 points for the Control (Comparison (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 7.093369 points. Consequently, the change for the Control (Comparison (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 0.6933021 points. The p-value for this difference was 0.4745508. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Reading of words EGRA subtask across the Control (Comparison (Portuguese)) and the Treatment (FFE + lit (Portuguese)) conditions beyond that expected from sampling error.

1.7.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 33: Reading of words

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.000	0.000	89	0	0	2.358	6.206	109	0	30
Treatment	0.033	0.654	538	0	15	8.177	11.455	761	0	30



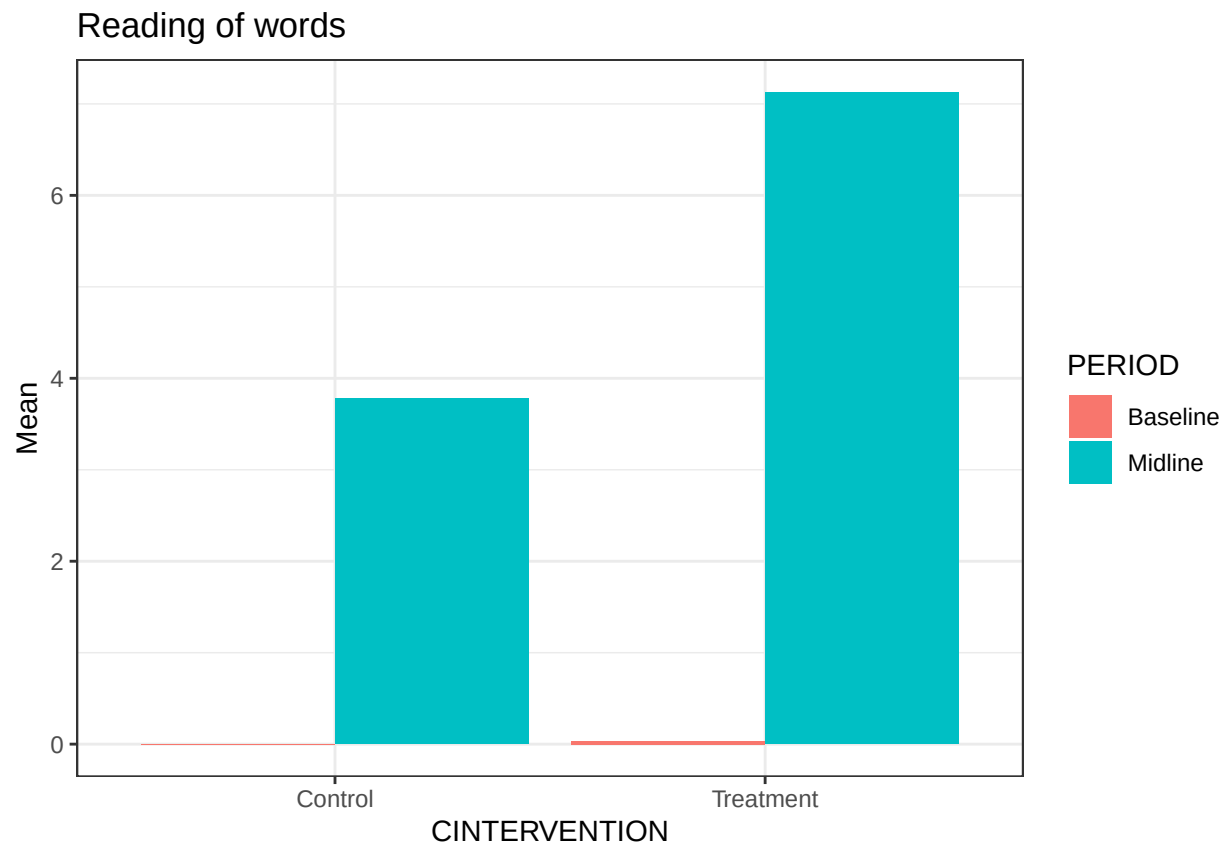


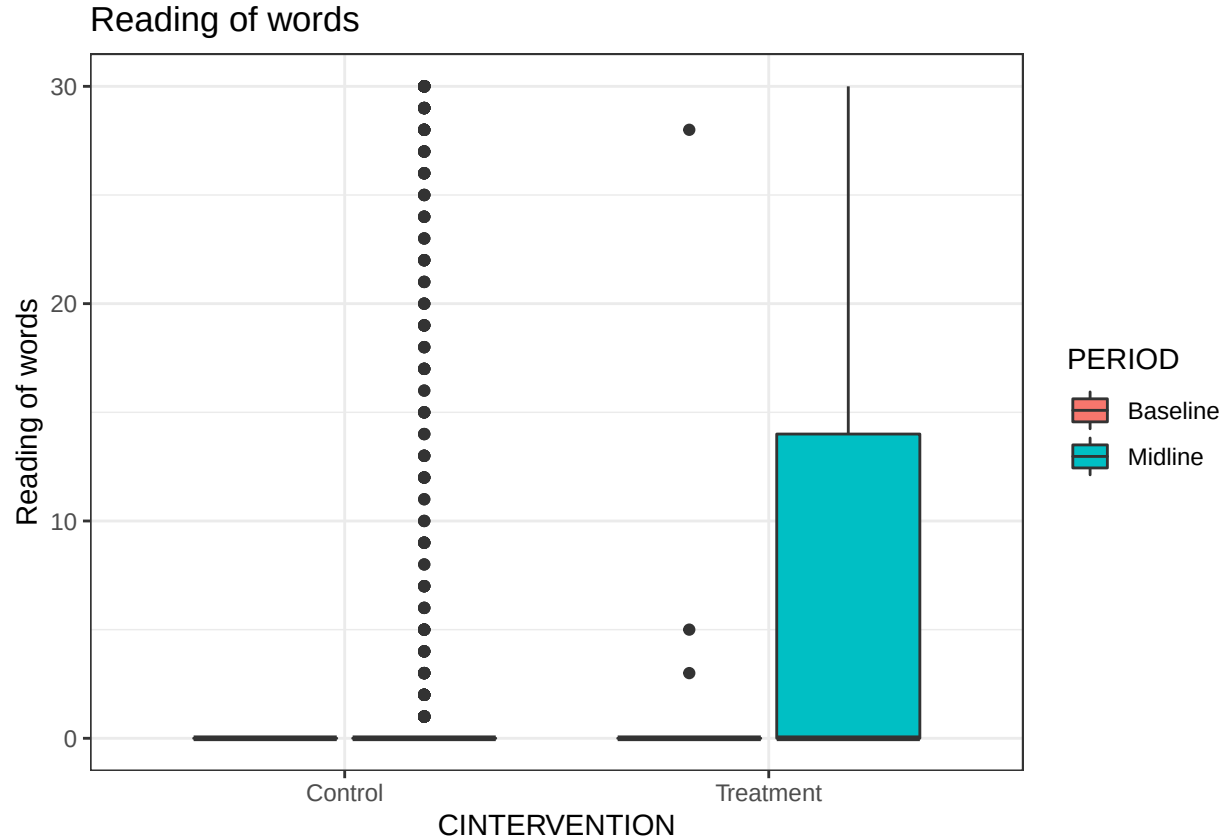
As shown in the table above, for the the Reading of words EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at baseline was 0 (SD = 0) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.03345725 (SD = 0.6535931). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.03345725 points. The p-value for this difference was 0.2300851. The mean for the Control (Comparison (Bilingual)) condition at midline was 2.357798 (SD = 6.205617) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 8.177398 (SD = 11.45466). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 5.8196 points. The p-value for this difference was 7.35086e-06. The change from the baseline to the midline of 2.357798 points for the Control (Comparison (Bilingual)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 8.143941 points. Consequently, the change for the Control (Comparison (Bilingual)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 5.786143 points. The p-value for this difference was 8.351949e-06. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Bilingual)) when compared to the Control (Comparison (Bilingual)) condition. This provides evidence that the Treatment (FFE + lit (Bilingual)) (or some other unobserved process) impacted on the Reading of words EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

1.7.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 34: Reading of words

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.000	0.000	1007	0	0	3.780	8.608	952	0	30
Treatment	0.035	0.887	1040	0	28	7.128	11.039	1047	0	30



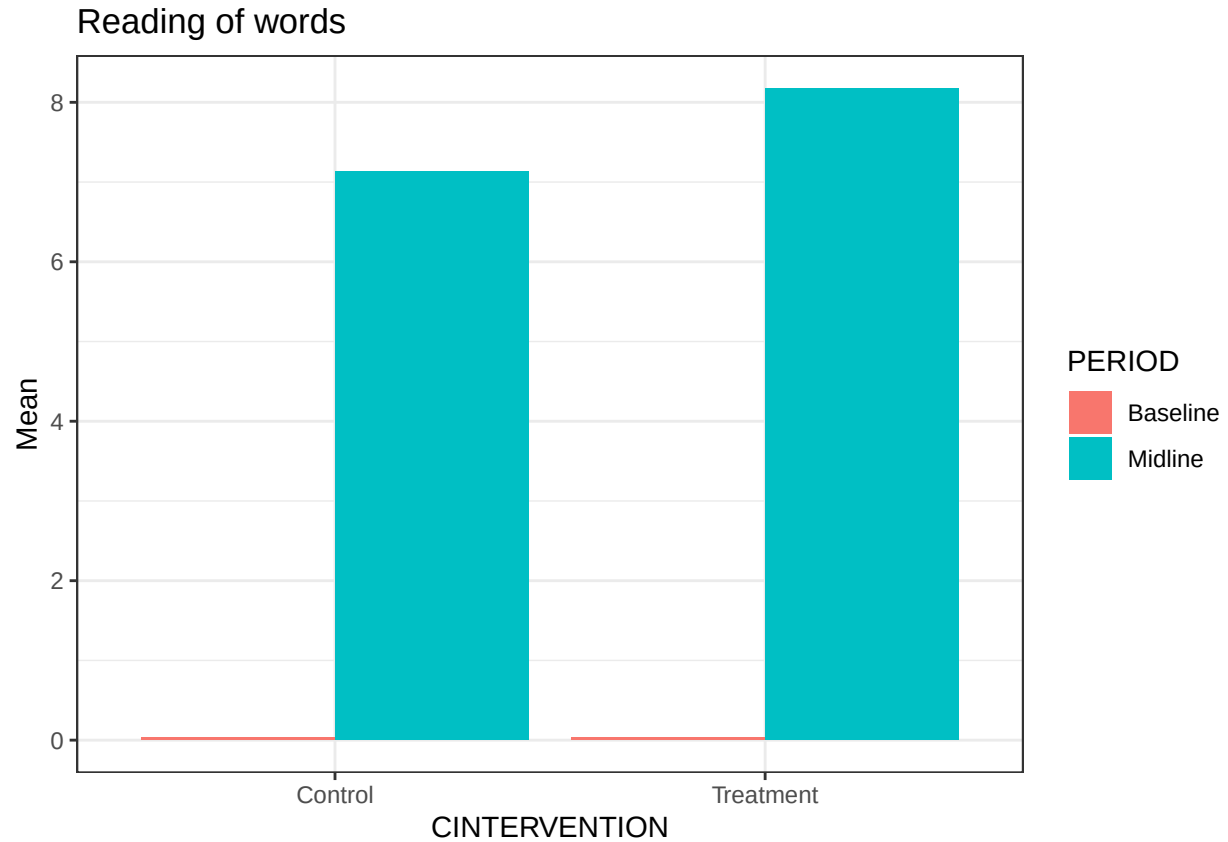


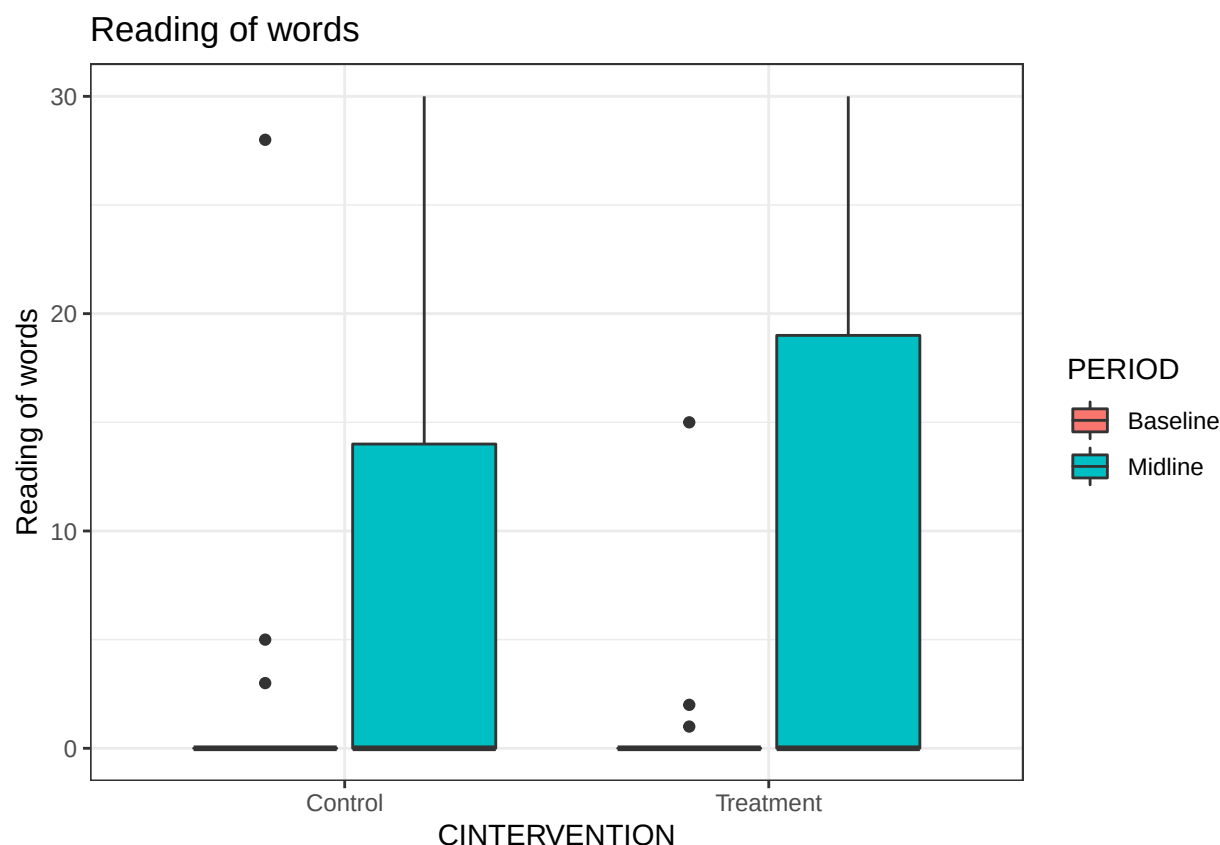
As shown in the table above, for the the Reading of words EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at baseline was 0 (SD = 0) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.03461538 (SD = 0.8866206). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.03461538 points. The p-value for this difference was 0.2085951. The mean for the Control (FFE only (Portuguese)) condition at midline was 3.780462 (SD = 8.608259) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 7.127985 (SD = 11.0389). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 3.347523 points. The p-value for this difference was 0.0001868595. The change from the baseline to the midline of 3.780462 points for the Control (FFE only (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 7.093369 points. Consequently, the change for the Control (FFE only (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 3.312907 points. The p-value for this difference was 0.0002150606. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Portuguese)) when compared to the Control (FFE only (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Portuguese)) (or some other unobserved process) impacted on the Reading of words EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

1.7.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 35: Reading of words

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.035	0.887	1040	0	28	7.128	11.039	1047	0	30
Treatment	0.033	0.654	538	0	15	8.177	11.455	761	0	30





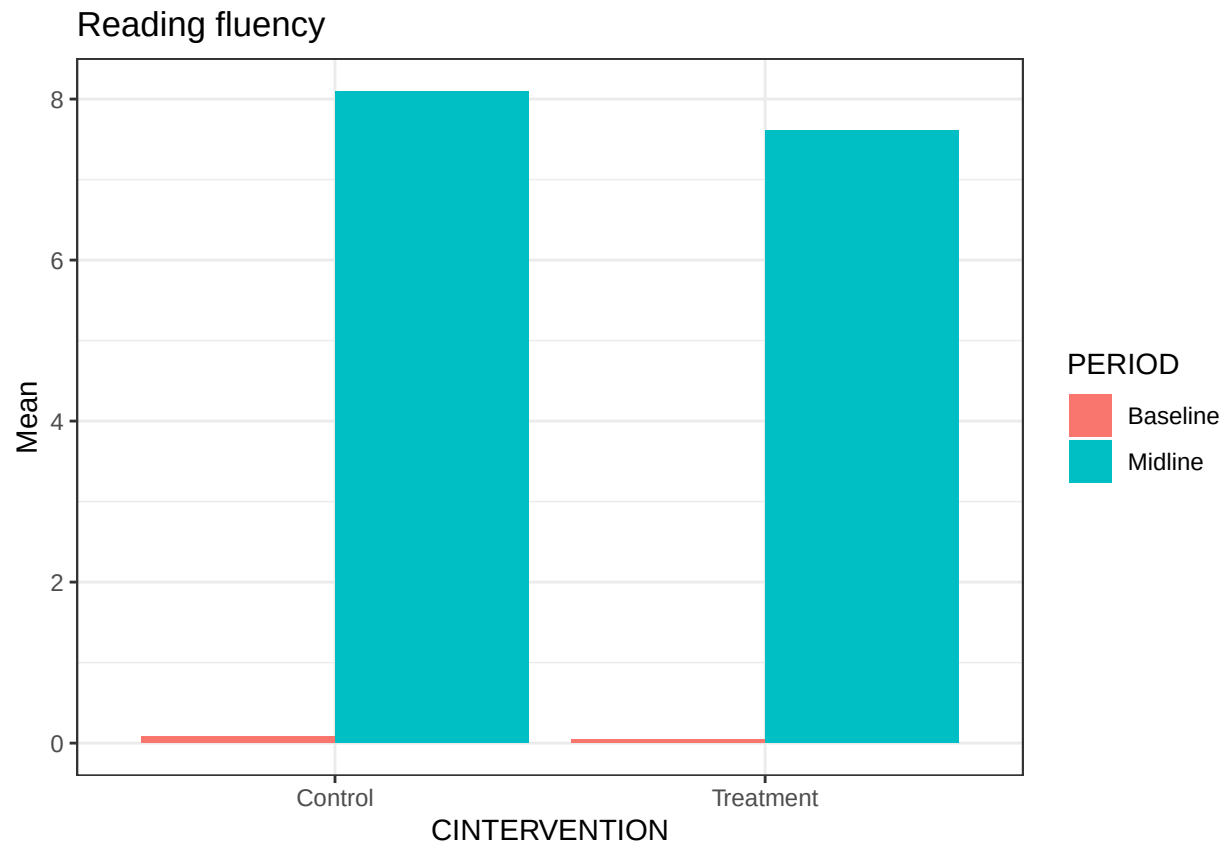
As shown in the table above, for the the Reading of words EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at baseline was 0.03461538 (SD = 0.8866206) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.03345725 (SD = 0.6535931). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.001158136 points. The p-value for this difference was 0.9761761. The mean for the Control (FFE + lit (Portuguese)) condition at midline was 7.127985 (SD = 11.0389) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 8.177398 (SD = 11.45466). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 1.049413 points. The p-value for this difference was 0.3010063. The change from the baseline to the midline of 7.093369 points for the Control (FFE + lit (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 8.143941 points. Consequently, the change for the Control (FFE + lit (Portuguese)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 1.050572 points. The p-value for this difference was 0.3004288. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Reading of words EGRA subtask across the Control (FFE + lit (Portuguese)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

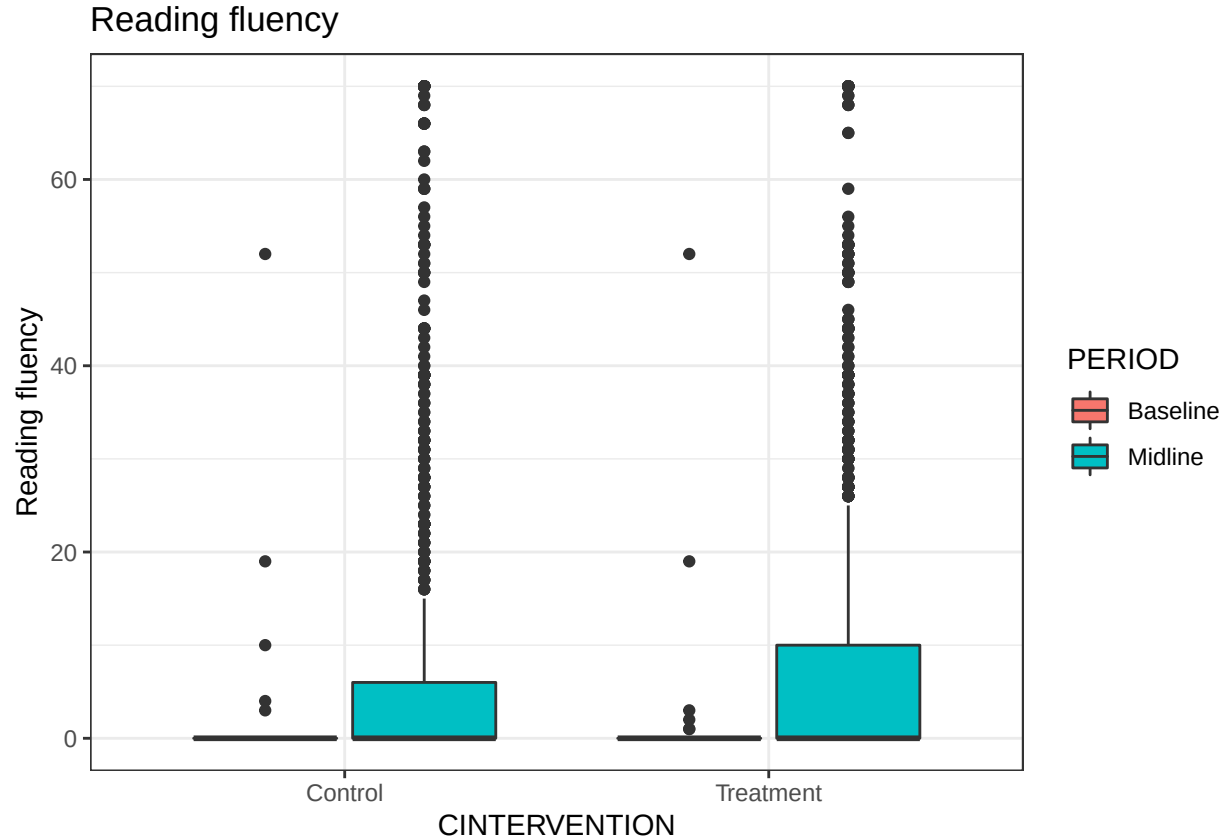
1.8 EGRA_ST8_1A: Reading fluency

1.8.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 36: Reading fluency

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.077	1.675	1136	0	52	8.098	16.495	1080	0	70
Treatment	0.049	1.397	1578	0	52	7.605	13.853	1791	0	70



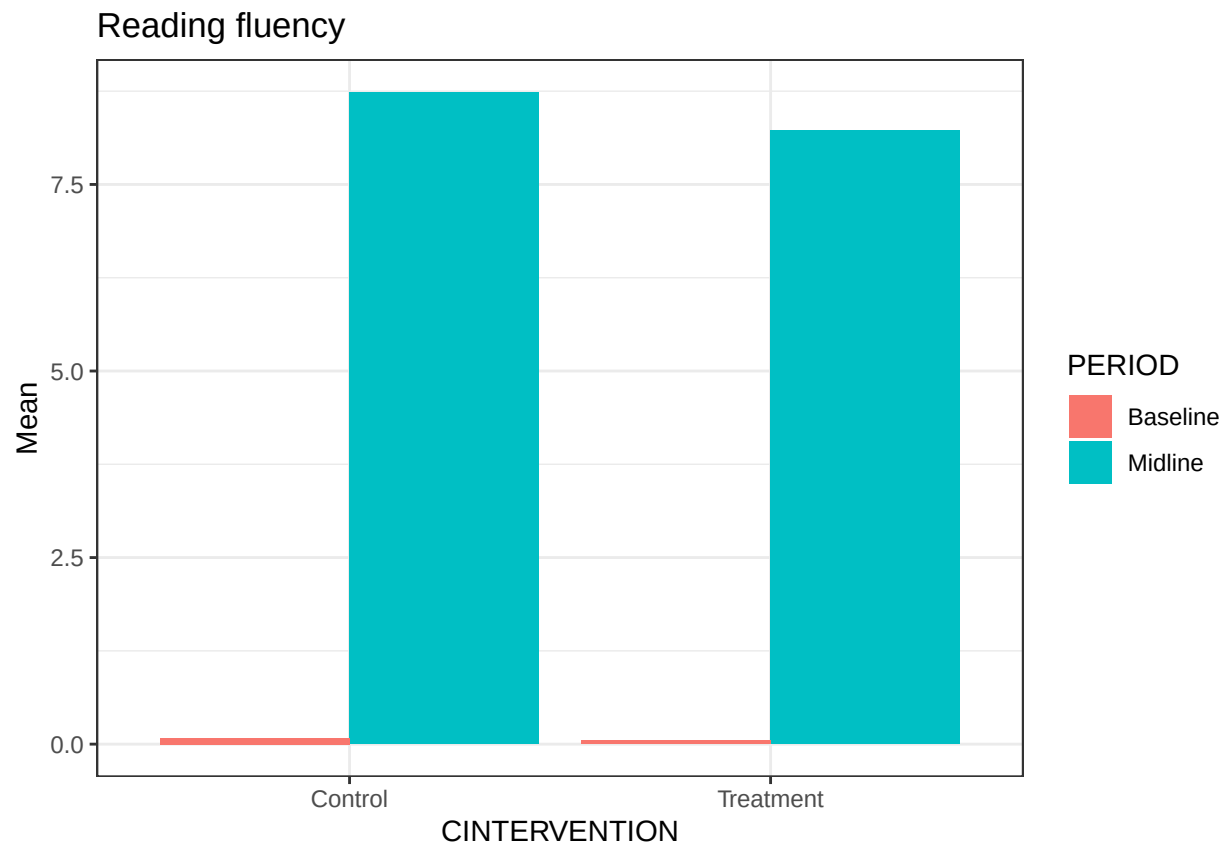


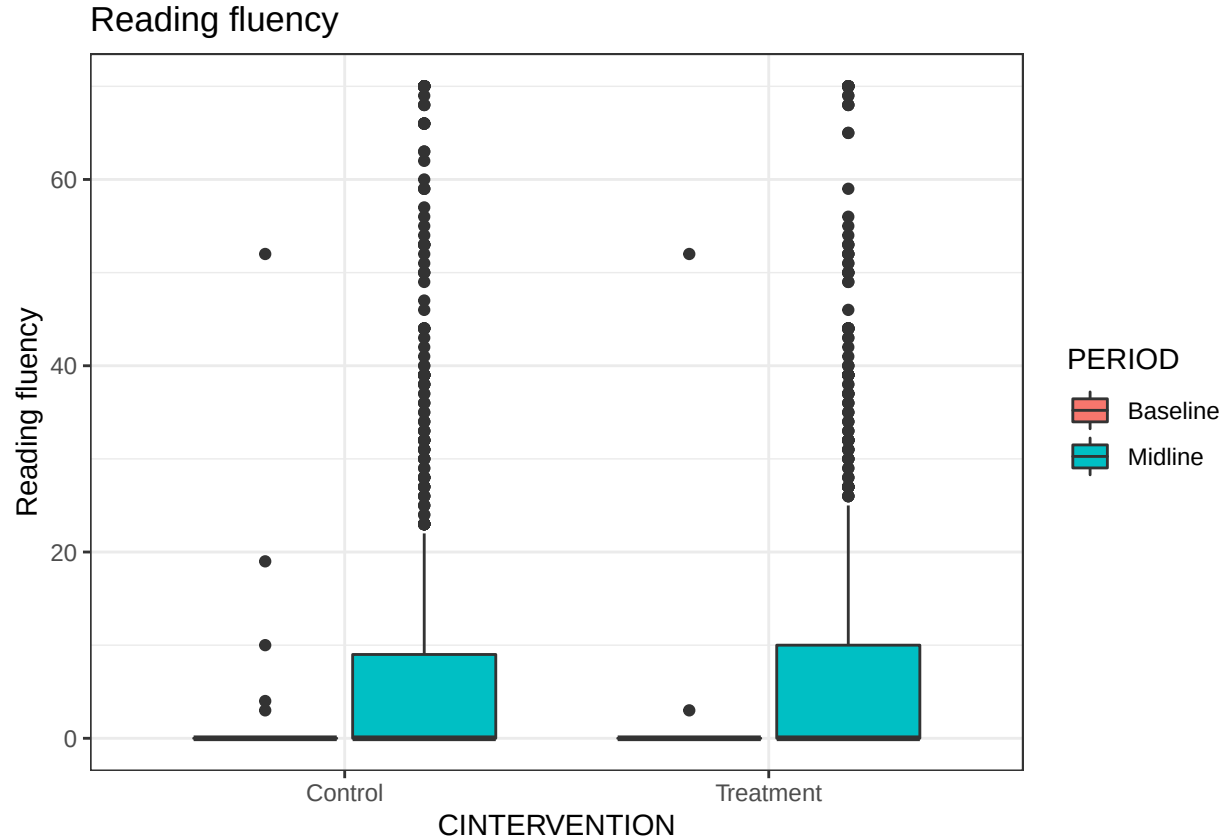
As shown in the table above, for the the Reading fluency EGRA subtask, the mean for the Control (Comparison (all)) condition at baseline was 0.07746479 (SD = 1.674684) and the mean for the Treatment (FFE + lit (all)) condition at baseline was 0.04942966 (SD = 1.39665). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at baseline was thus of 0.02803513 points. The p-value for this difference was 0.6447485. The mean for the Control (Comparison (all)) condition at midline was 8.098148 (SD = 16.49483) and the mean for the Treatment (FFE + lit (all)) condition at midline was 7.605248 (SD = 13.8533). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at midline was thus of 0.4928997 points. The p-value for this difference was 0.6686538. The change from the baseline to the midline of 8.020683 points for the Control (Comparison (all)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (all)) can be assessed against. The change for the Treatment (FFE + lit (all)) from the baseline to the midline was of 7.555819 points. Consequently, the change for the Control (Comparison (all)) relative to the change for the Treatment (FFE + lit (all)) condition was of -0.4648646 points. The p-value for this difference was 0.6859271. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Reading fluency EGRA subtask across the Control (Comparison (all)) and the Treatment (FFE + lit (all)) conditions beyond that expected from sampling error.

1.8.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 37: Reading fluency

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.084	1.744	1047	0	52	8.737	17.047	972	0	70
Treatment	0.053	1.615	1040	0	52	8.223	15.261	1039	0	70



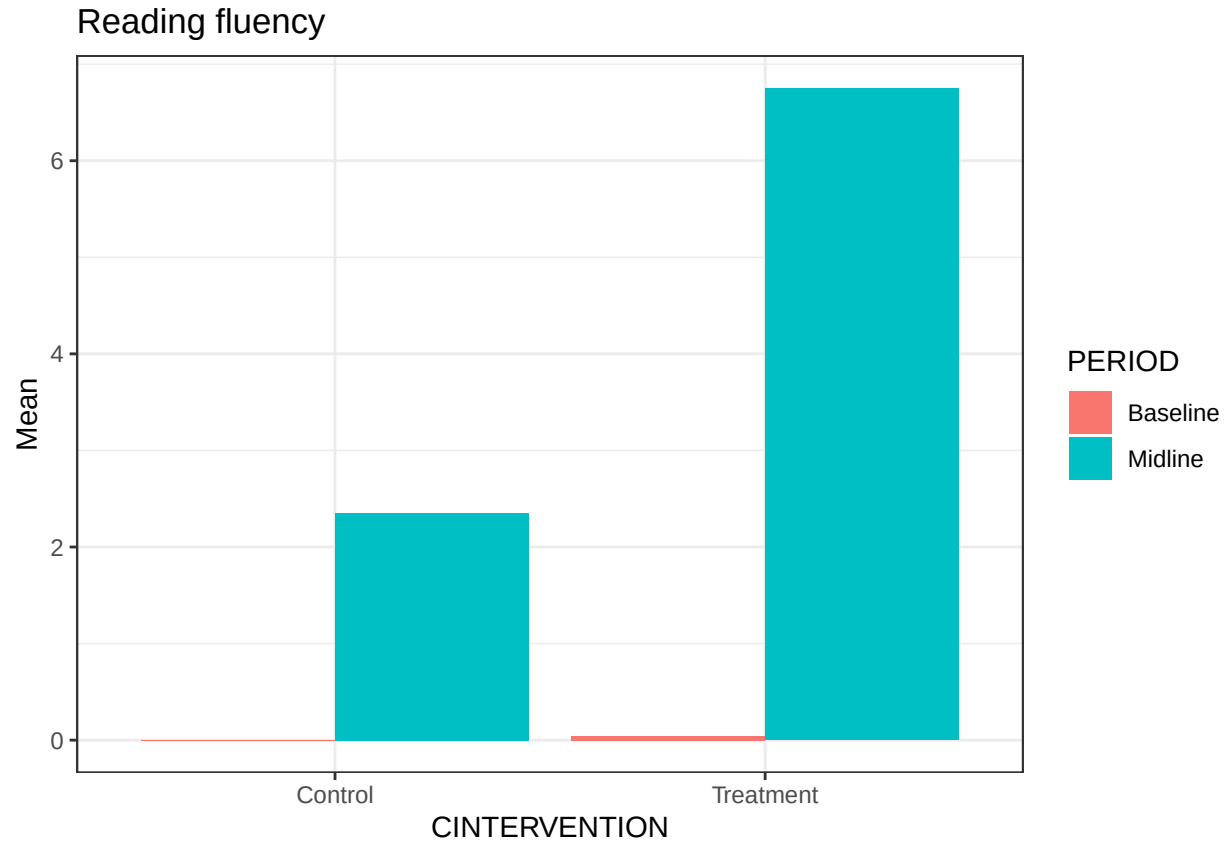


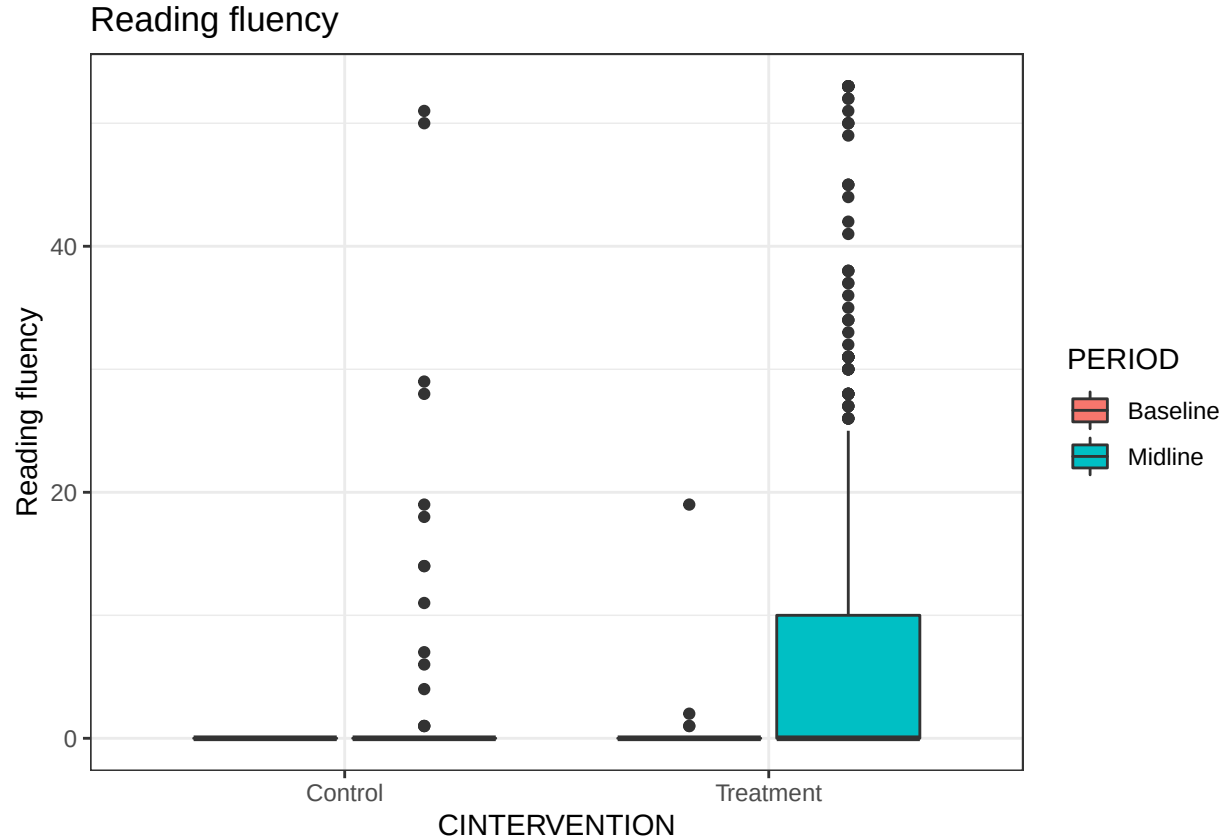
As shown in the table above, for the the Reading fluency EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at baseline was 0.08404967 (SD = 1.744317) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.05288462 (SD = 1.615043). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.03116505 points. The p-value for this difference was 0.6718402. The mean for the Control (Comparison (Portuguese)) condition at midline was 8.736626 (SD = 17.04723) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 8.223292 (SD = 15.26094). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.5133339 points. The p-value for this difference was 0.7056136. The change from the baseline to the midline of 8.652576 points for the Control (Comparison (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 8.170407 points. Consequently, the change for the Control (Comparison (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of -0.4821688 points. The p-value for this difference was 0.722265. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Reading fluency EGRA subtask across the Control (Comparison (Portuguese)) and the Treatment (FFE + lit (Portuguese)) conditions beyond that expected from sampling error.

1.8.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 38: Reading fluency

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.000	0.000	89	0	0	2.352	8.333	108	0	51
Treatment	0.043	0.826	538	0	19	6.751	11.587	752	0	53



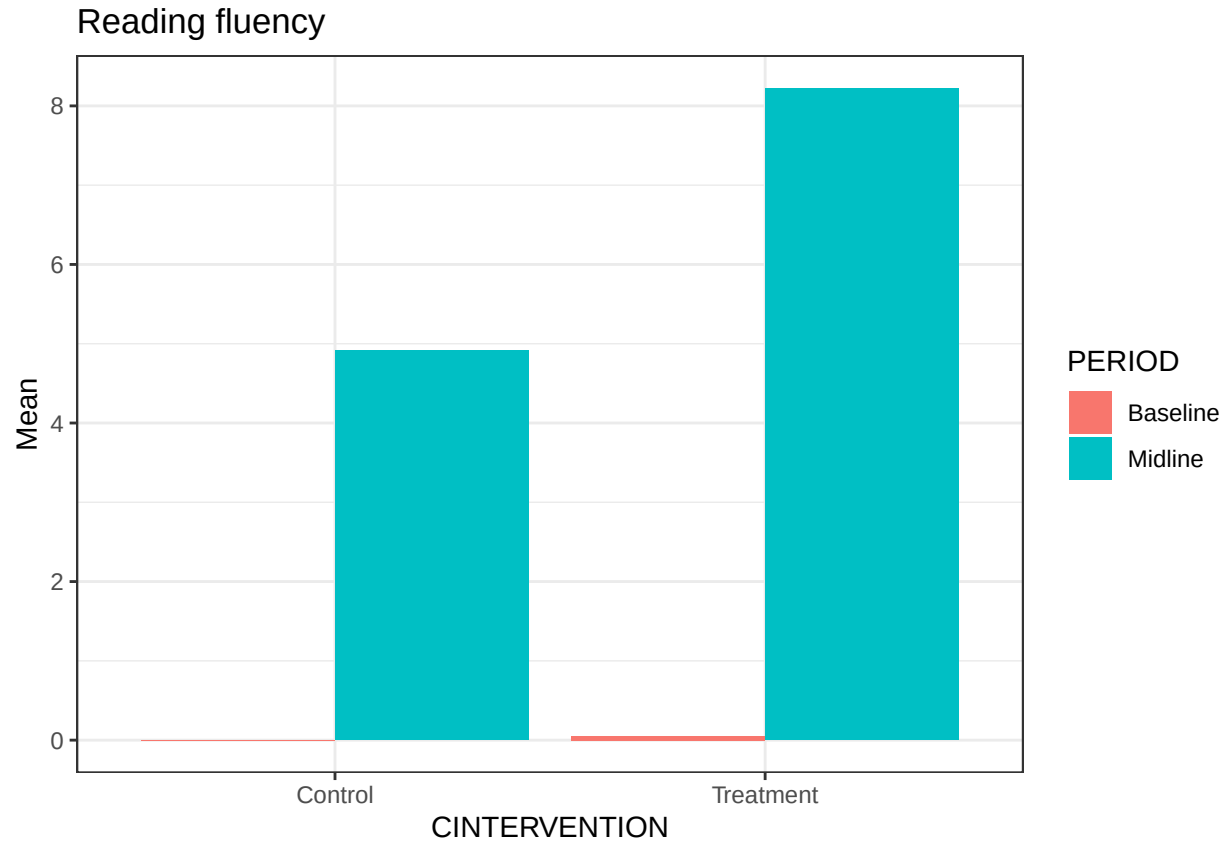


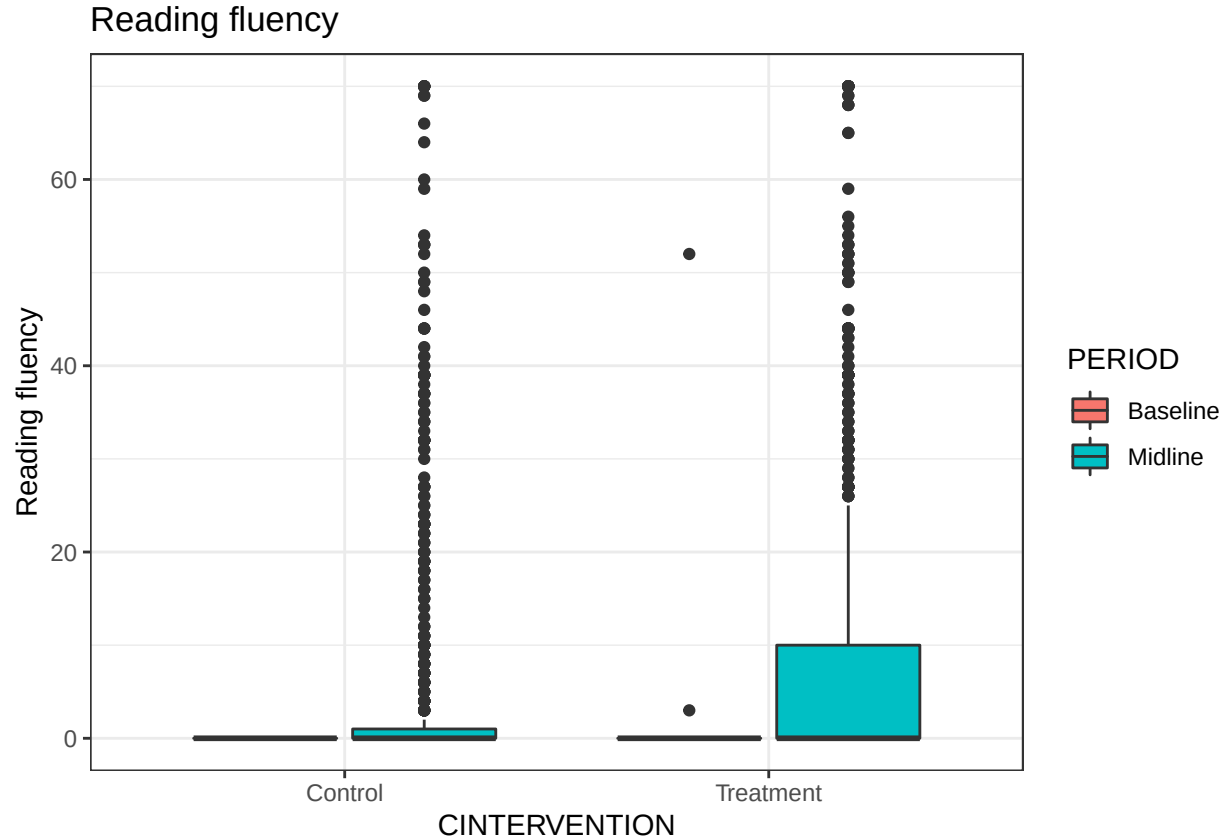
As shown in the table above, for the the Reading fluency EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at baseline was 0 (SD = 0) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.04275093 (SD = 0.8255879). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.04275093 points. The p-value for this difference was 0.2270531. The mean for the Control (Comparison (Bilingual)) condition at midline was 2.351852 (SD = 8.333375) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 6.75133 (SD = 11.58736). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 4.399478 points. The p-value for this difference was 0.01060688. The change from the baseline to the midline of 2.351852 points for the Control (Comparison (Bilingual)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 6.708579 points. Consequently, the change for the Control (Comparison (Bilingual)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 4.356727 points. The p-value for this difference was 0.01145581. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Reading fluency EGRA subtask across the Control (Comparison (Bilingual)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

1.8.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 39: Reading fluency

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.000	0.000	1007	0	0	4.916	12.433	951	0	70
Treatment	0.053	1.615	1040	0	52	8.223	15.261	1039	0	70



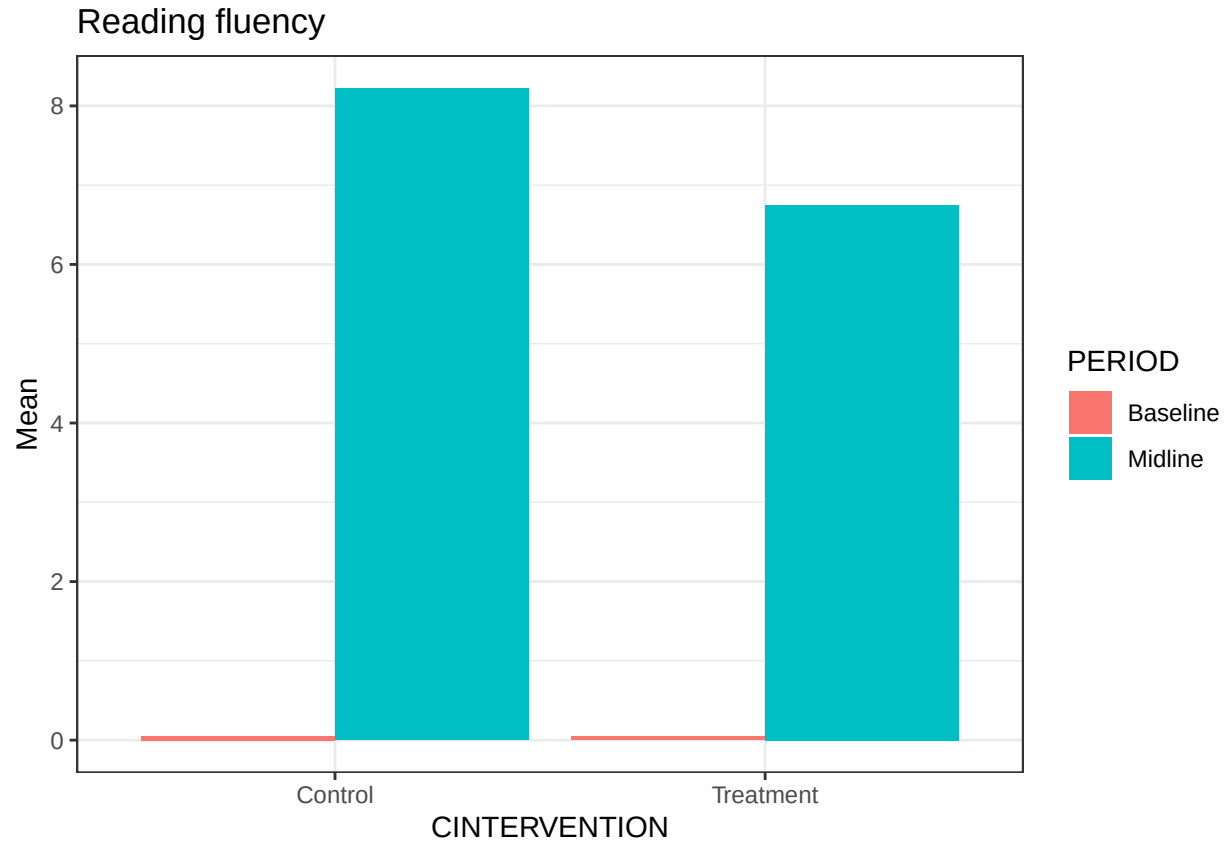


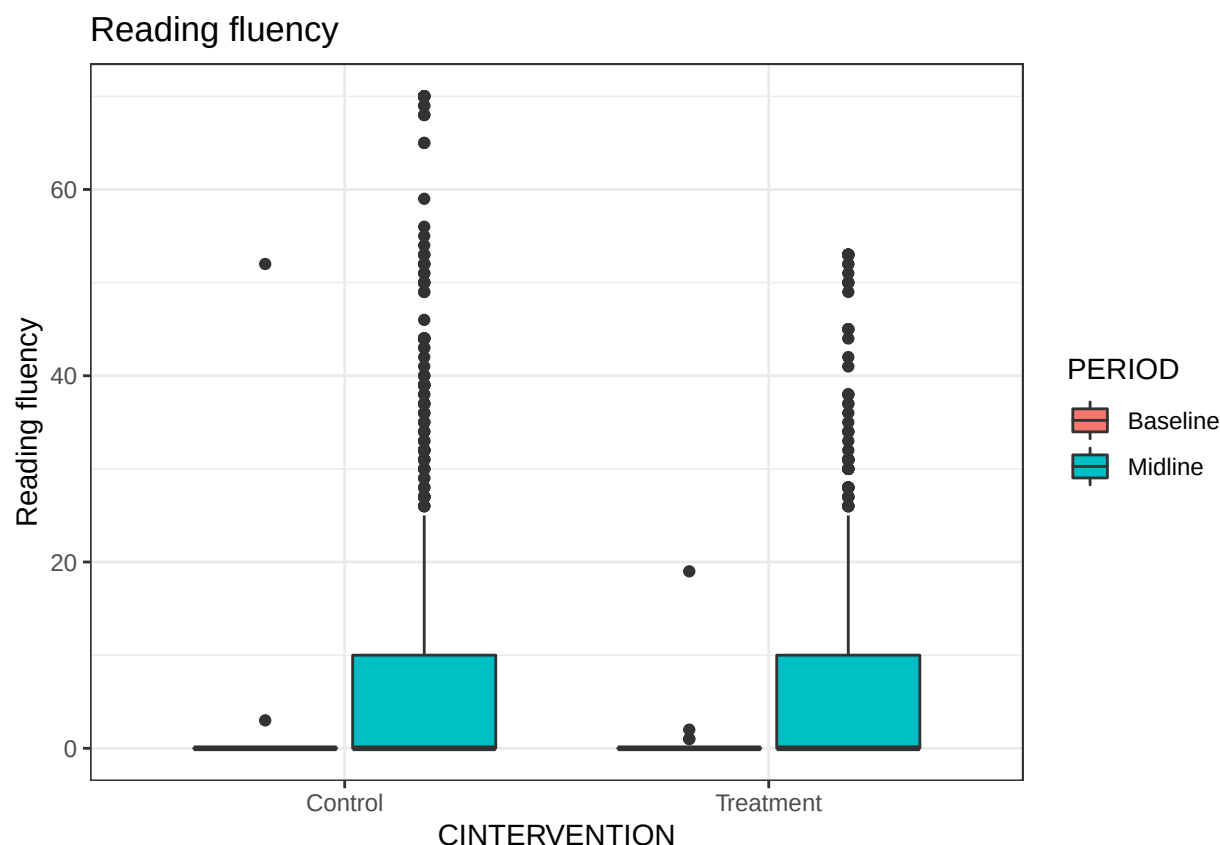
As shown in the table above, for the the Reading fluency EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at baseline was 0 (SD = 0) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.05288462 (SD = 1.615043). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.05288462 points. The p-value for this difference was 0.2920071. The mean for the Control (FFE only (Portuguese)) condition at midline was 4.915878 (SD = 12.43346) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 8.223292 (SD = 15.26094). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 3.307414 points. The p-value for this difference was 0.005798436. The change from the baseline to the midline of 4.915878 points for the Control (FFE only (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 8.170407 points. Consequently, the change for the Control (FFE only (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 3.254529 points. The p-value for this difference was 0.006581735. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Portuguese)) when compared to the Control (FFE only (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Portuguese)) (or some other unobserved process) impacted on the Reading fluency EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

1.8.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 40: Reading fluency

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.053	1.615	1040	0	52	8.223	15.261	1039	0	70
Treatment	0.043	0.826	538	0	19	6.751	11.587	752	0	53





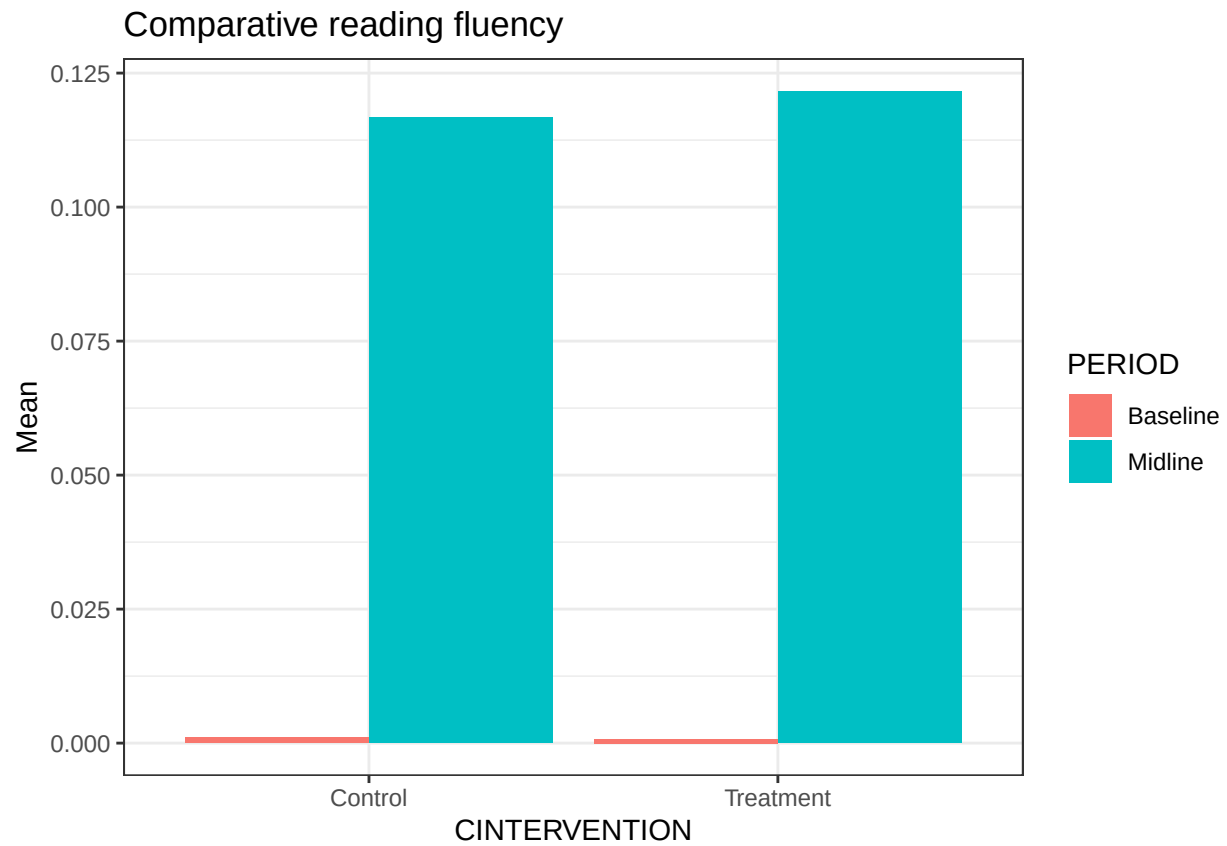
As shown in the table above, for the the Reading fluency EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at baseline was 0.05288462 (SD = 1.615043) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.04275093 (SD = 0.8255879). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.01013369 points. The p-value for this difference was 0.8687699. The mean for the Control (FFE + lit (Portuguese)) condition at midline was 8.223292 (SD = 15.26094) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 6.75133 (SD = 11.58736). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 1.471962 points. The p-value for this difference was 0.2100281. The change from the baseline to the midline of 8.170407 points for the Control (FFE + lit (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 6.708579 points. Consequently, the change for the Control (FFE + lit (Portuguese)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of -1.461828 points. The p-value for this difference was 0.2135764. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Reading fluency EGRA subtask across the Control (FFE + lit (Portuguese)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

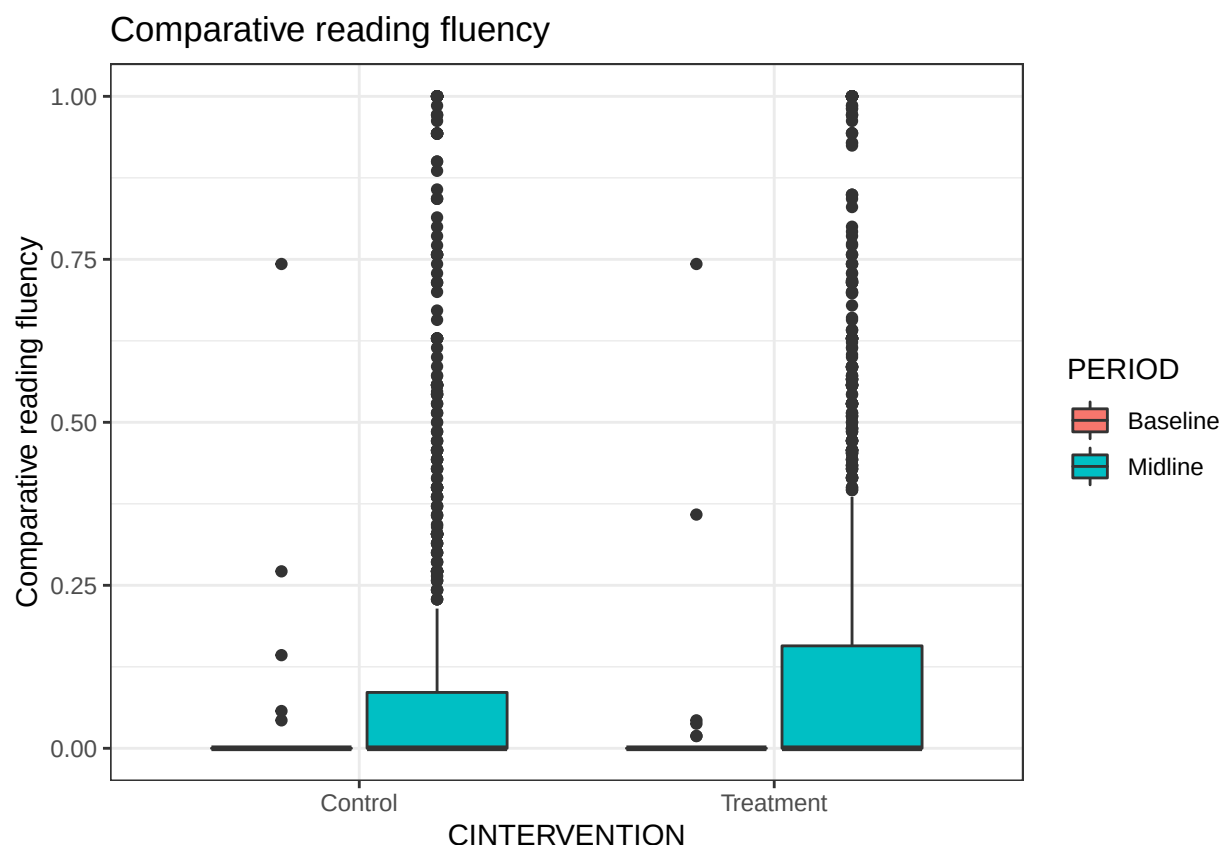
1.9 EGRA_ST8_1AS: Comparative reading fluency

1.9.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 41: Comparative reading fluency

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.001	0.024	1136	0	0.743	0.117	0.237	1080	0	1
Treatment	0.001	0.021	1578	0	0.743	0.122	0.218	1791	0	1



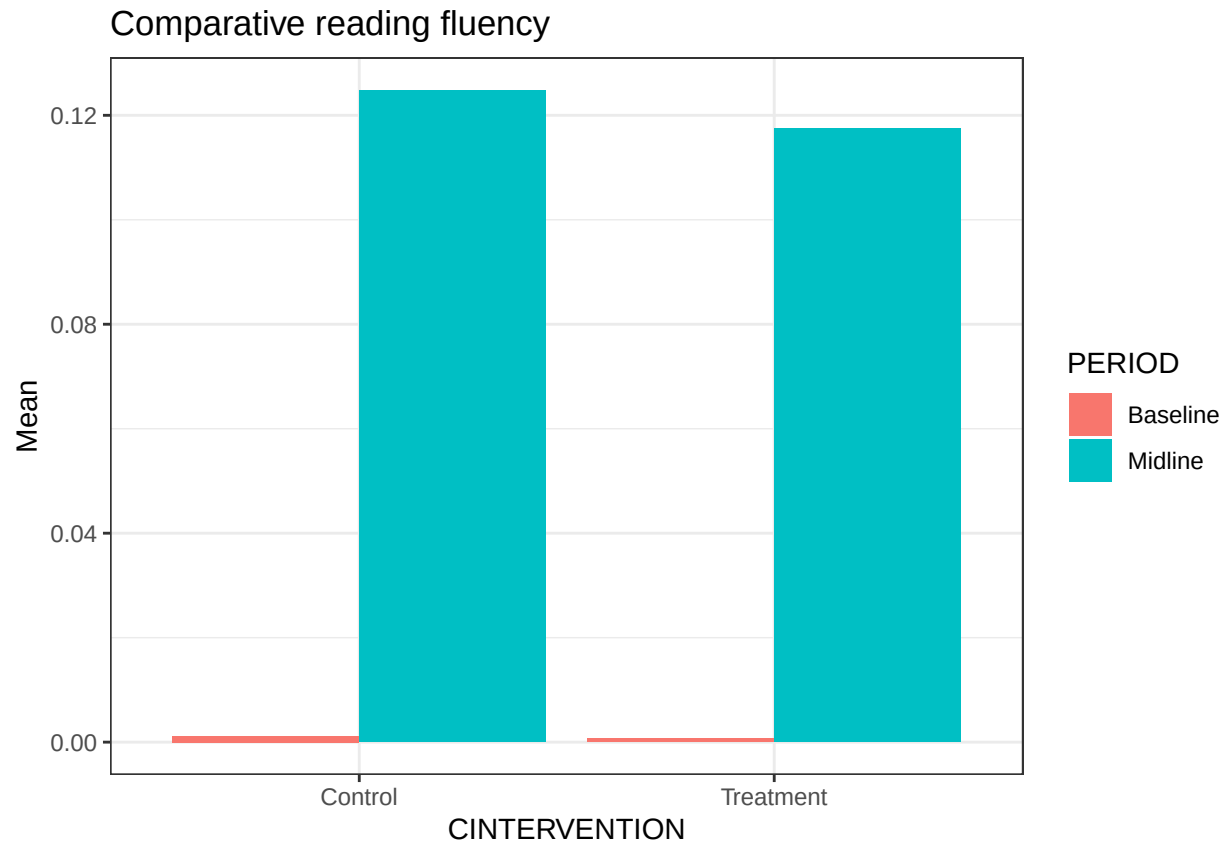


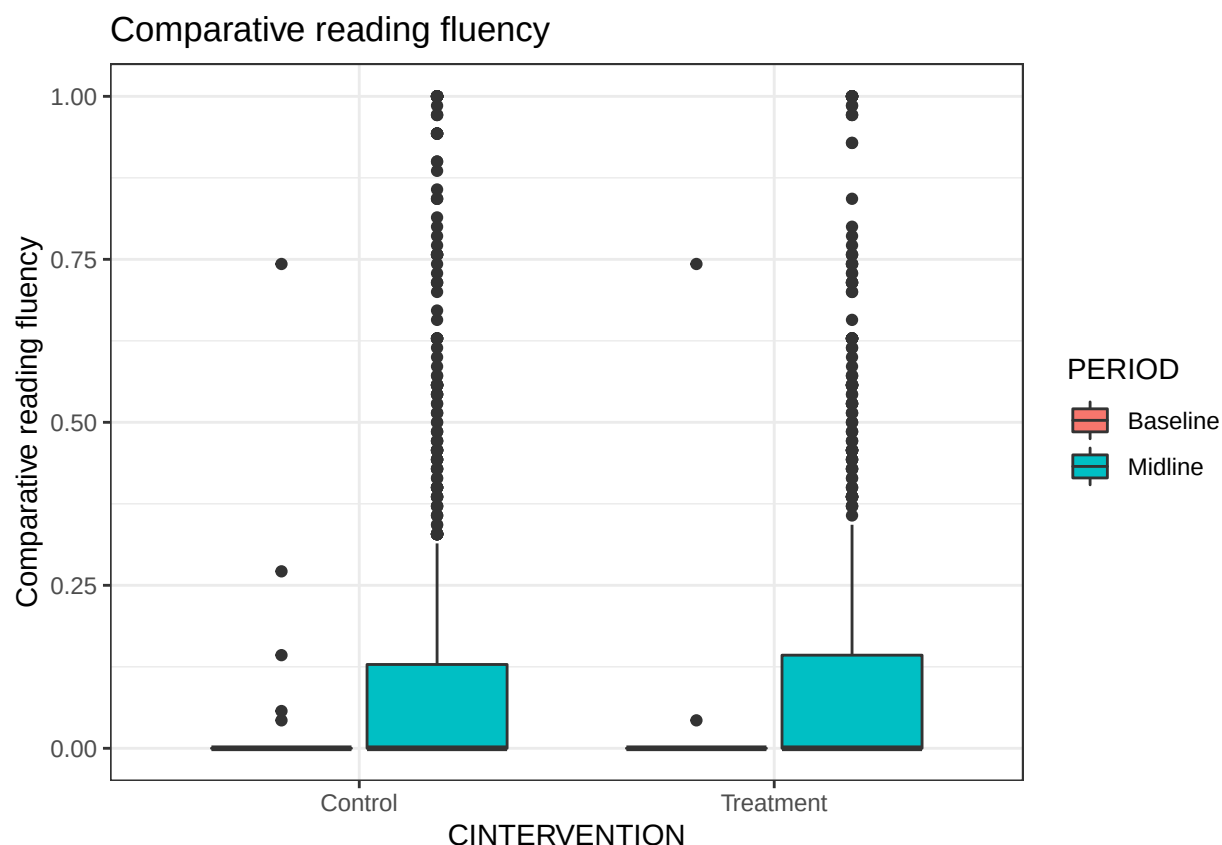
As shown in the table above, for the the Comparative reading fluency EGRA subtask, the mean for the Control (Comparison (all)) condition at baseline was 0.00110664 (SD = 0.02392406) and the mean for the Treatment (FFE + lit (all)) condition at baseline was 0.0007729256 (SD = 0.0208169). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at baseline was thus of 0.0003337143 points. The p-value for this difference was 0.7049342. The mean for the Control (Comparison (all)) condition at midline was 0.1167655 (SD = 0.2374994) and the mean for the Treatment (FFE + lit (all)) condition at midline was 0.1216281 (SD = 0.2182686). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at midline was thus of 0.004862558 points. The p-value for this difference was 0.7742896. The change from the baseline to the midline of 0.1156589 points for the Control (Comparison (all)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (all)) can be assessed against. The change for the Treatment (FFE + lit (all)) from the baseline to the midline was of 0.1208551 points. Consequently, the change for the Control (Comparison (all)) relative to the change for the Treatment (FFE + lit (all)) condition was of 0.005196272 points. The p-value for this difference was 0.758846. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Comparative reading fluency EGRA subtask across the Control (Comparison (all)) and the Treatment (FFE + lit (all)) conditions beyond that expected from sampling error.

1.9.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 42: Comparative reading fluency

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.001	0.025	1047	0	0.743	0.125	0.244	972	0	1
Treatment	0.001	0.023	1040	0	0.743	0.117	0.218	1039	0	1



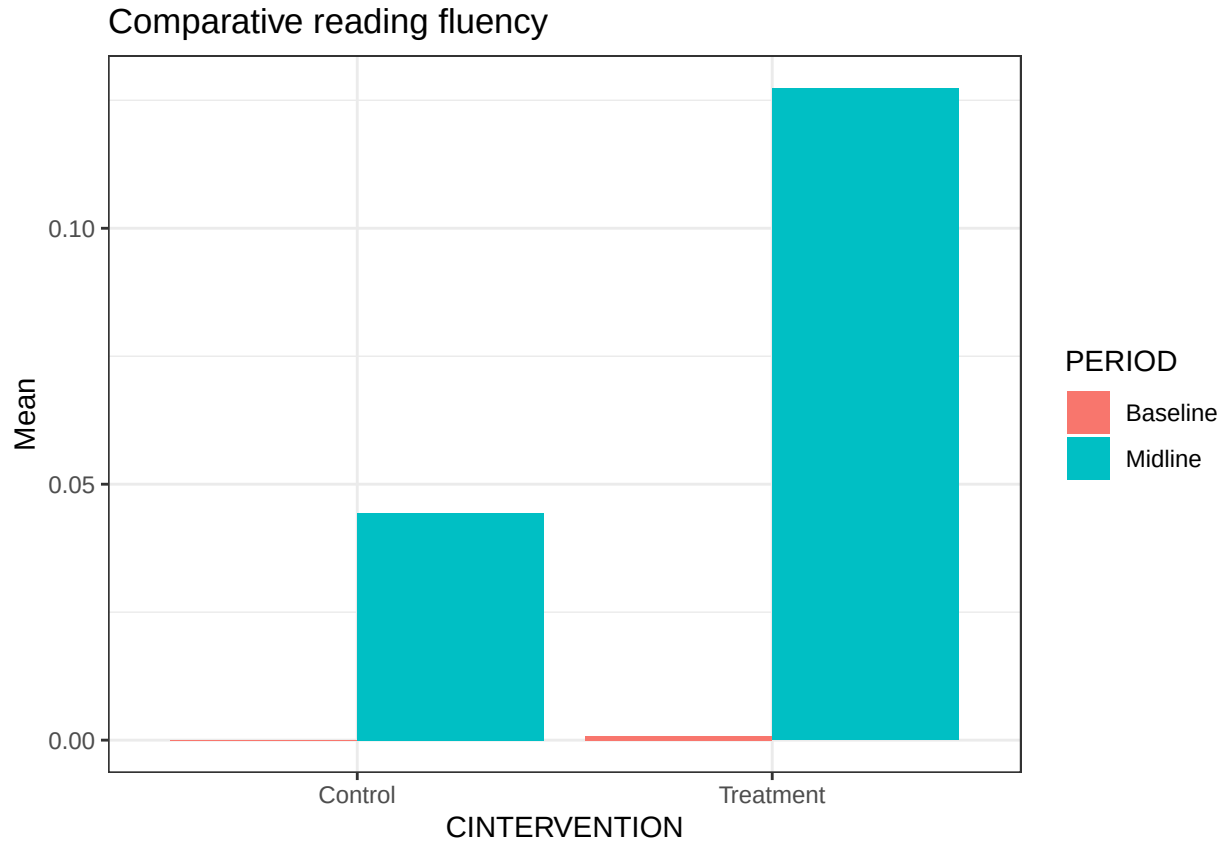


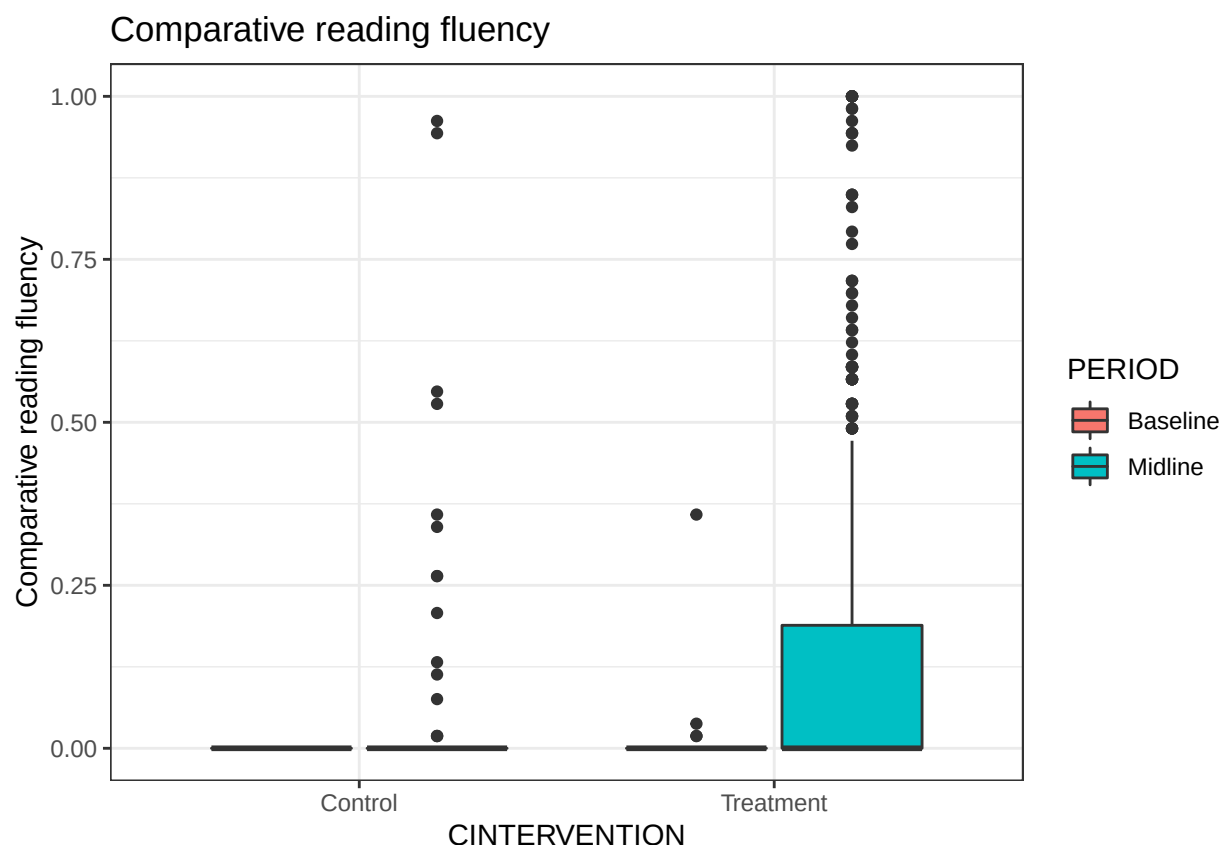
As shown in the table above, for the the Comparative reading fluency EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at baseline was 0.00120071 (SD = 0.02491882) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.0007554945 (SD = 0.02307205). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.000445215 points. The p-value for this difference was 0.6718402. The mean for the Control (Comparison (Portuguese)) condition at midline was 0.1248089 (SD = 0.2435319) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 0.1174756 (SD = 0.2180134). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.007333341 points. The p-value for this difference was 0.7056136. The change from the baseline to the midline of 0.1236082 points for the Control (Comparison (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 0.1167201 points. Consequently, the change for the Control (Comparison (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of -0.00688126 points. The p-value for this difference was 0.722265. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Comparative reading fluency EGRA subtask across the Control (Comparison (Portuguese)) and the Treatment (FFE + lit (Portuguese)) conditions beyond that expected from sampling error.

1.9.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 43: Comparative reading fluency

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.000	0.000	89	0	0.000	0.044	0.157	108	0	0.962
Treatment	0.001	0.016	538	0	0.358	0.127	0.219	752	0	1.000



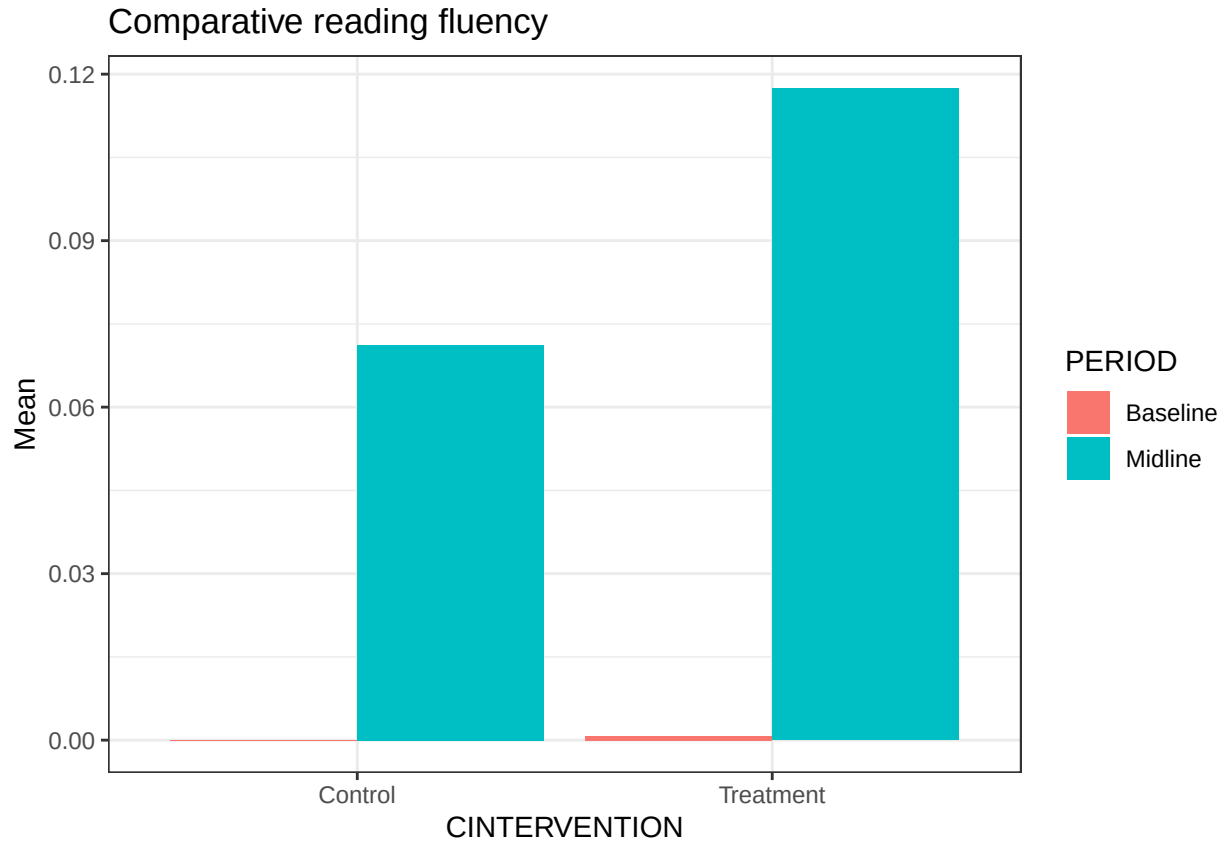


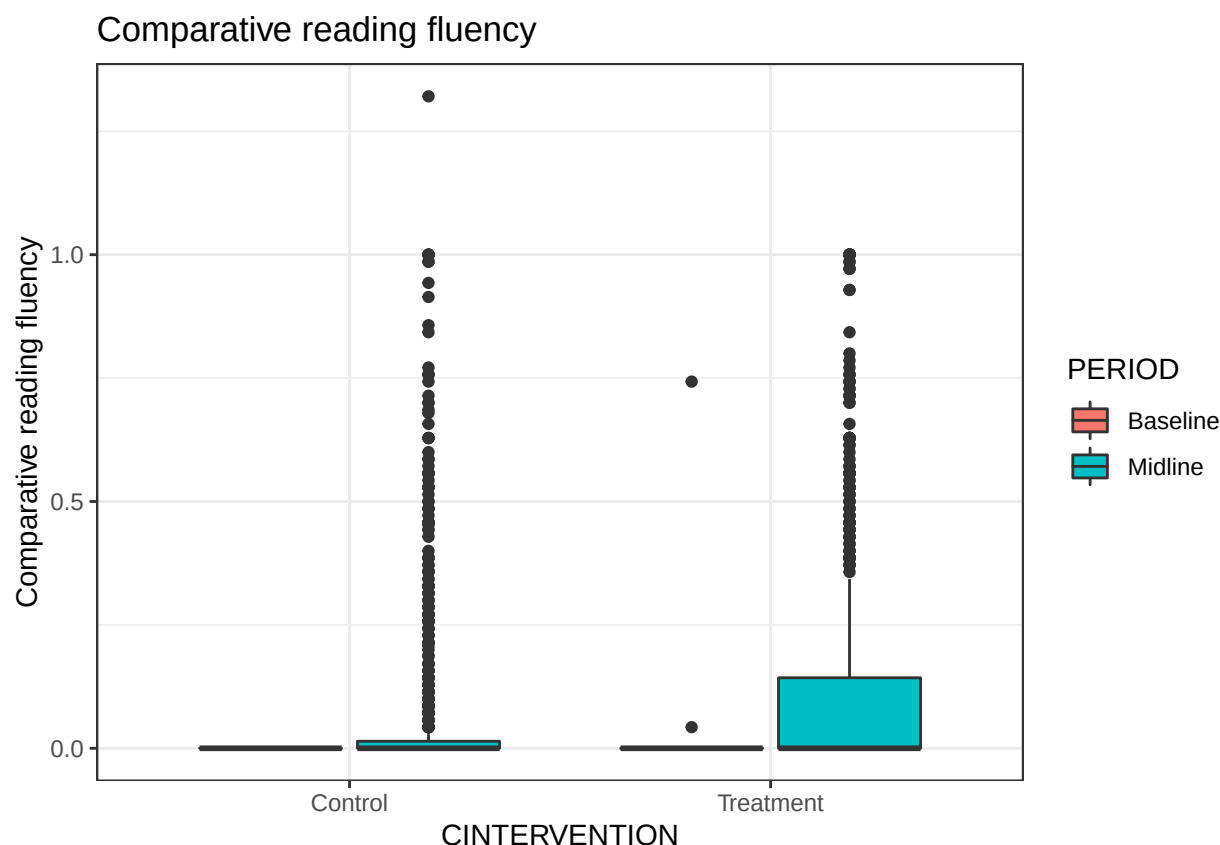
As shown in the table above, for the the Comparative reading fluency EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at baseline was 0 (SD = 0) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.0008066213 (SD = 0.01557713). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.0008066213 points. The p-value for this difference was 0.2270531. The mean for the Control (Comparison (Bilingual)) condition at midline was 0.04437456 (SD = 0.1572335) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 0.1273653 (SD = 0.218636). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.08299074 points. The p-value for this difference was 0.01062726. The change from the baseline to the midline of 0.04437456 points for the Control (Comparison (Bilingual)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 0.1265587 points. Consequently, the change for the Control (Comparison (Bilingual)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 0.08218412 points. The p-value for this difference was 0.01147754. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Comparative reading fluency EGRA subtask across the Control (Comparison (Bilingual)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

1.9.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 44: Comparative reading fluency

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.000	0.000	1007	0	0.000	0.071	0.181	951	0	1.321
Treatment	0.001	0.023	1040	0	0.743	0.117	0.218	1039	0	1.000



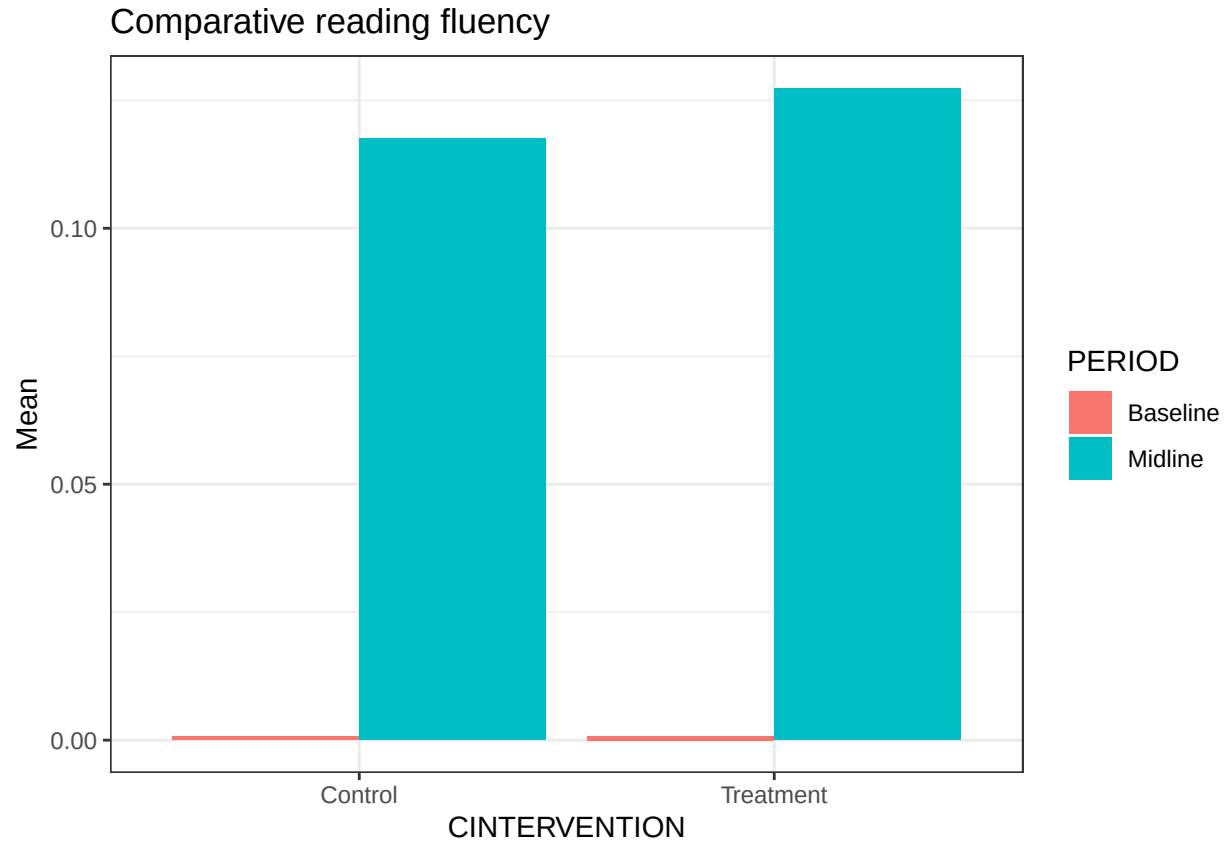


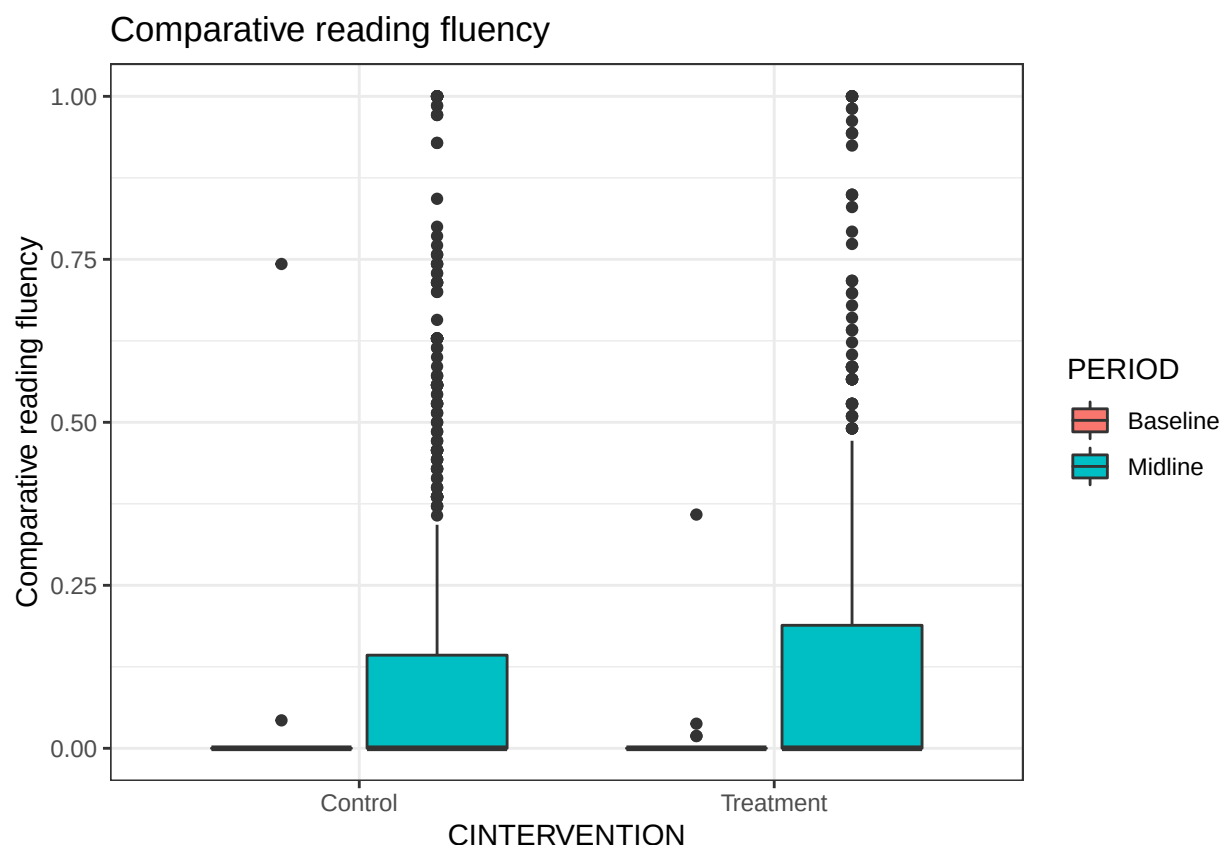
As shown in the table above, for the the Comparative reading fluency EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at baseline was 0 (SD = 0) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.0007554945 (SD = 0.02307205). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.0007554945 points. The p-value for this difference was 0.2920071. The mean for the Control (FFE only (Portuguese)) condition at midline was 0.07120013 (SD = 0.1807224) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 0.1174756 (SD = 0.2180134). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.04627547 points. The p-value for this difference was 0.00734767. The change from the baseline to the midline of 0.07120013 points for the Control (FFE only (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 0.1167201 points. Consequently, the change for the Control (FFE only (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 0.04551997 points. The p-value for this difference was 0.008307649. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Portuguese)) when compared to the Control (FFE only (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Portuguese)) (or some other unobserved process) impacted on the Comparative reading fluency EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

1.9.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 45: Comparative reading fluency

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.001	0.023	1040	0	0.743	0.117	0.218	1039	0	1
Treatment	0.001	0.016	538	0	0.358	0.127	0.219	752	0	1





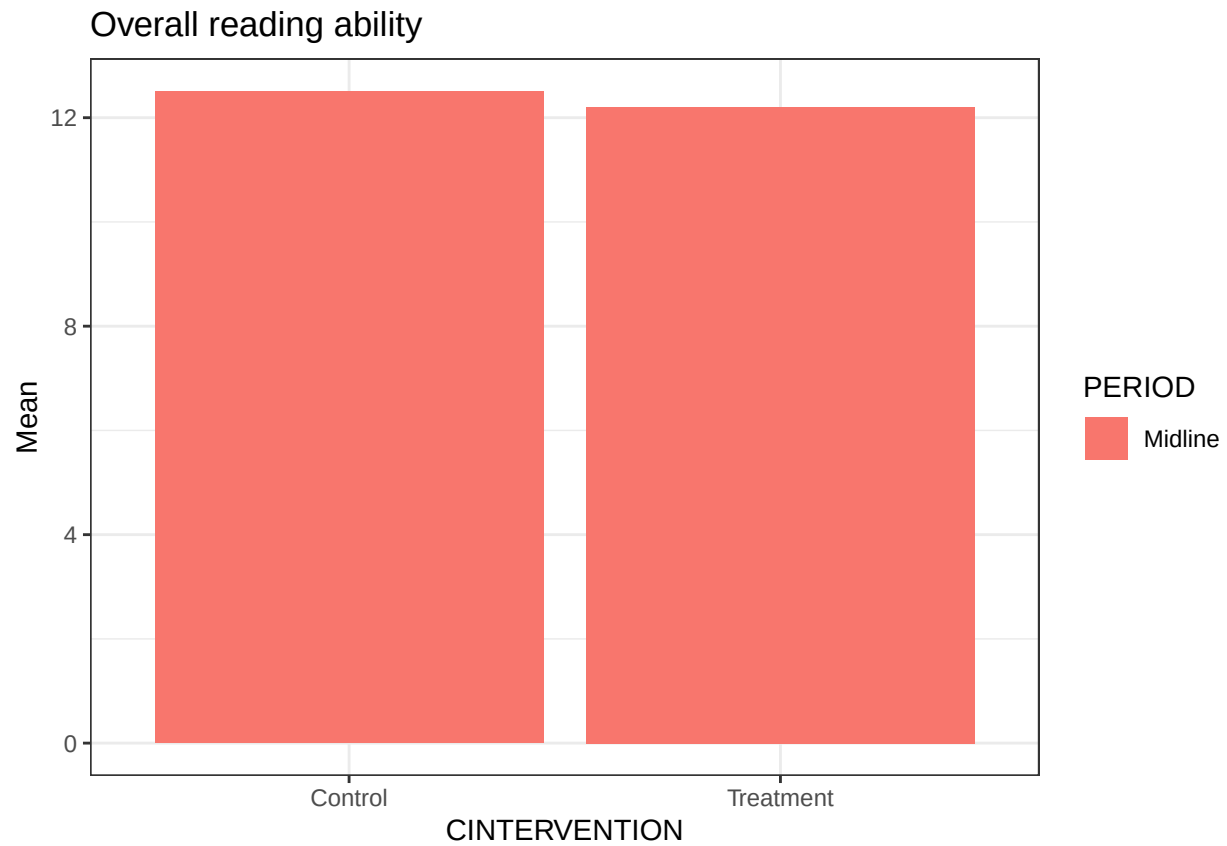
As shown in the table above, for the the Comparative reading fluency EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at baseline was 0.0007554945 (SD = 0.02307205) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.0008066213 (SD = 0.01557713). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 5.11268e-05 points. The p-value for this difference was 0.9582923. The mean for the Control (FFE + lit (Portuguese)) condition at midline was 0.1174756 (SD = 0.2180134) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 0.1273653 (SD = 0.218636). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.009889706 points. The p-value for this difference was 0.6060341. The change from the baseline to the midline of 0.1167201 points for the Control (FFE + lit (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 0.1265587 points. Consequently, the change for the Control (FFE + lit (Portuguese)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 0.009838579 points. The p-value for this difference was 0.6083137. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Comparative reading fluency EGRA subtask across the Control (FFE + lit (Portuguese)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

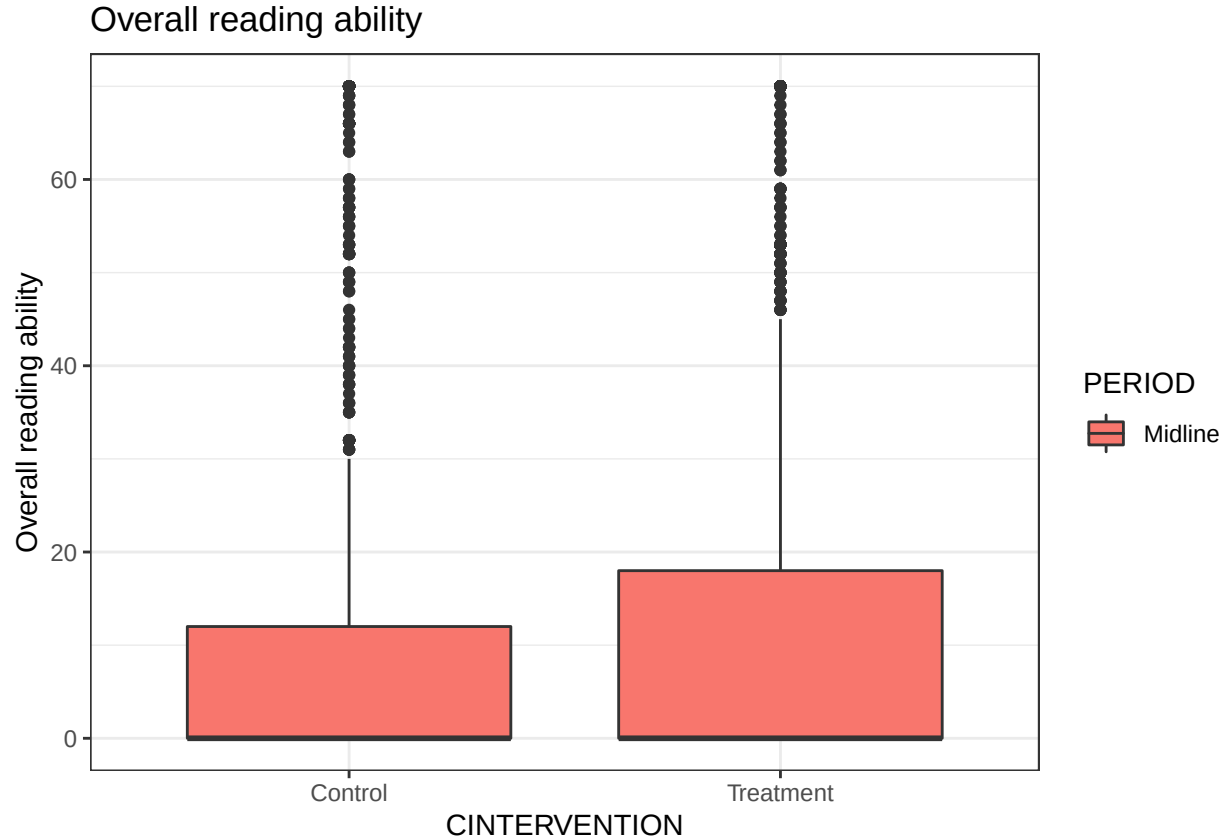
1.10 EGRA_ST8_1B: Overall reading ability

1.10.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 46: Overall reading ability

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	12.509	22.893	1078	0	70
Treatment	NA	NA	0	NA	NA	12.204	20.815	1759	0	70



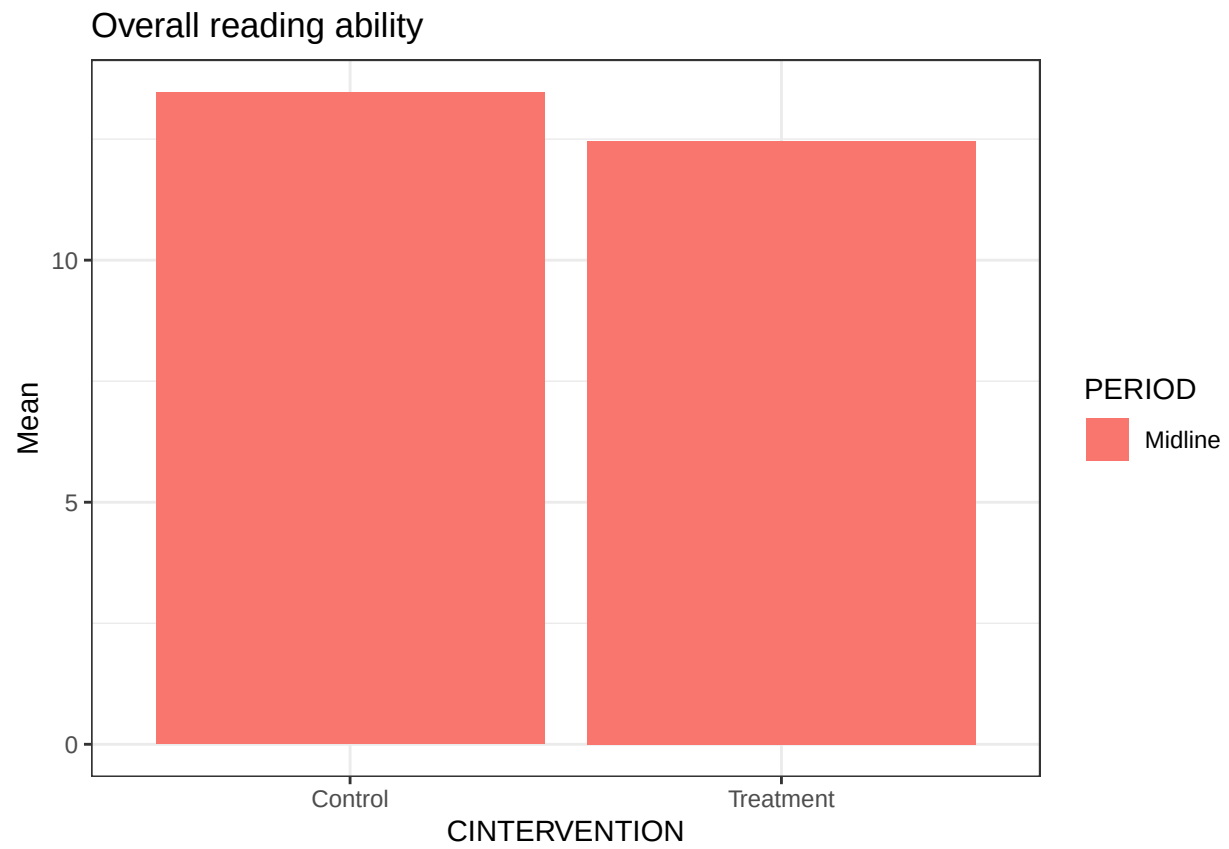


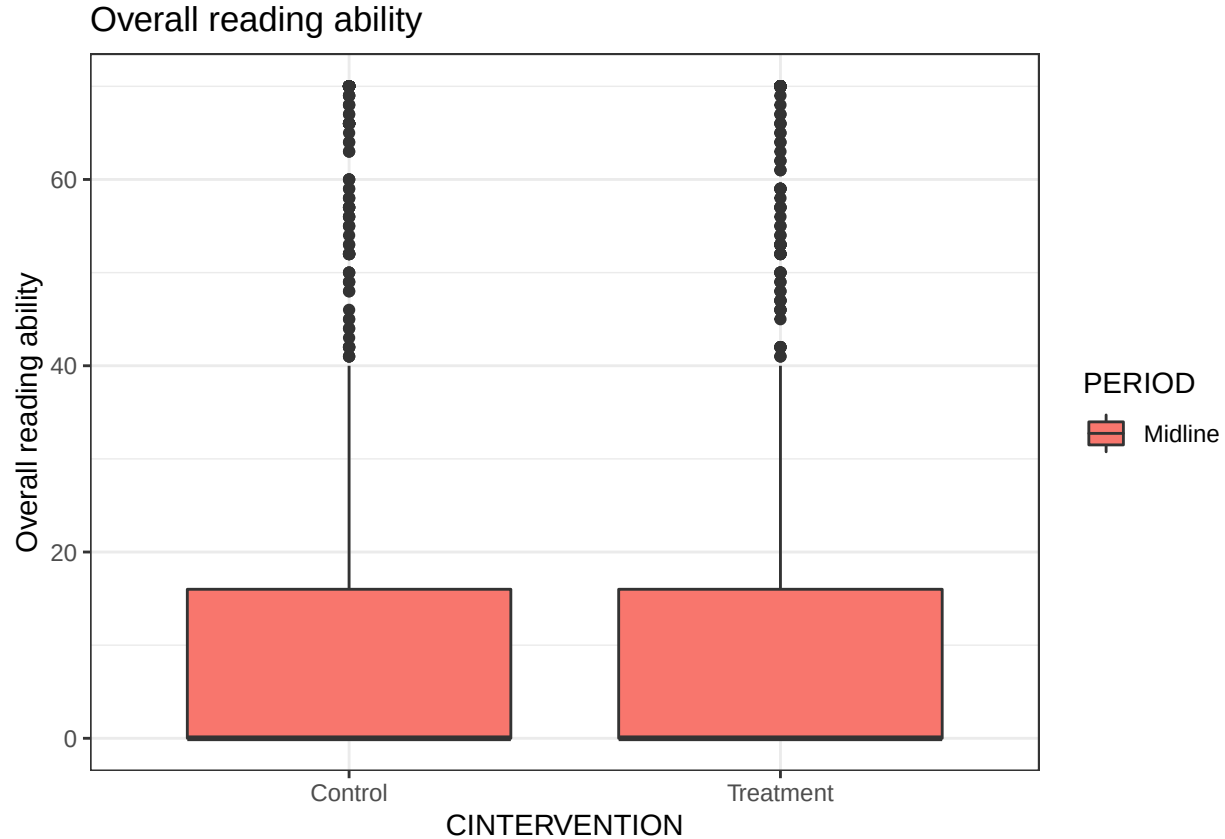
As shown in the table above, for the Overall reading ability EGRA subtask, the mean for the Control (Comparison (all)) condition at midline was 12.50928 (SD = 22.89327) and the mean for the Treatment (FFE + lit (all)) condition at midline was 12.20409 (SD = 20.81483). The difference for the Overall reading ability across the Control (Comparison (all)) and Treatment (FFE + lit (all)) conditions at midline was thus -0.3051832 points (there was no baseline measurement for this variable). The p-value for this difference was 0.8609184. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the levels of Overall reading ability between the Control (Comparison (all)) and Treatment (FFE + lit (all)) conditions at midline beyond that which could result from sampling error.

1.10.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 47: Overall reading ability

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	13.466	23.617	970	0	70
Treatment	NA	NA	0	NA	NA	12.455	22.366	1017	0	70



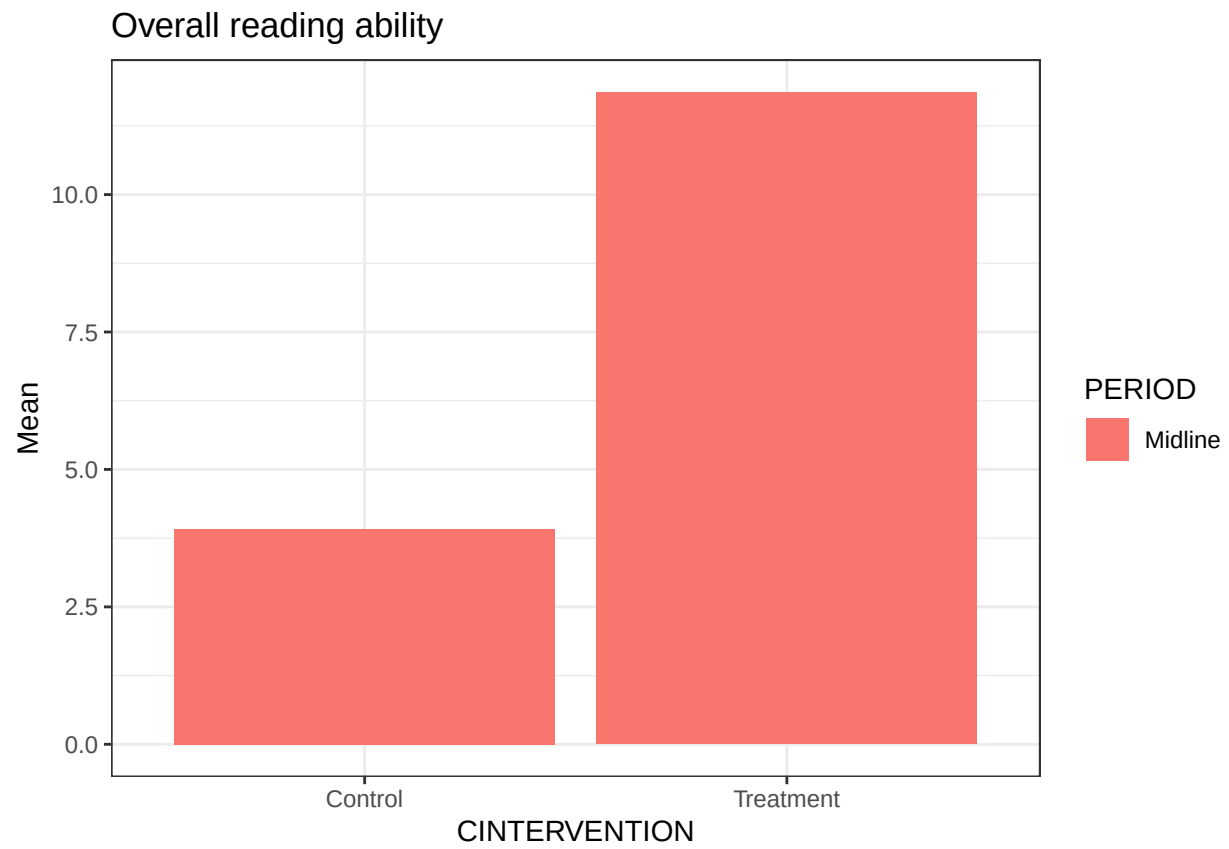


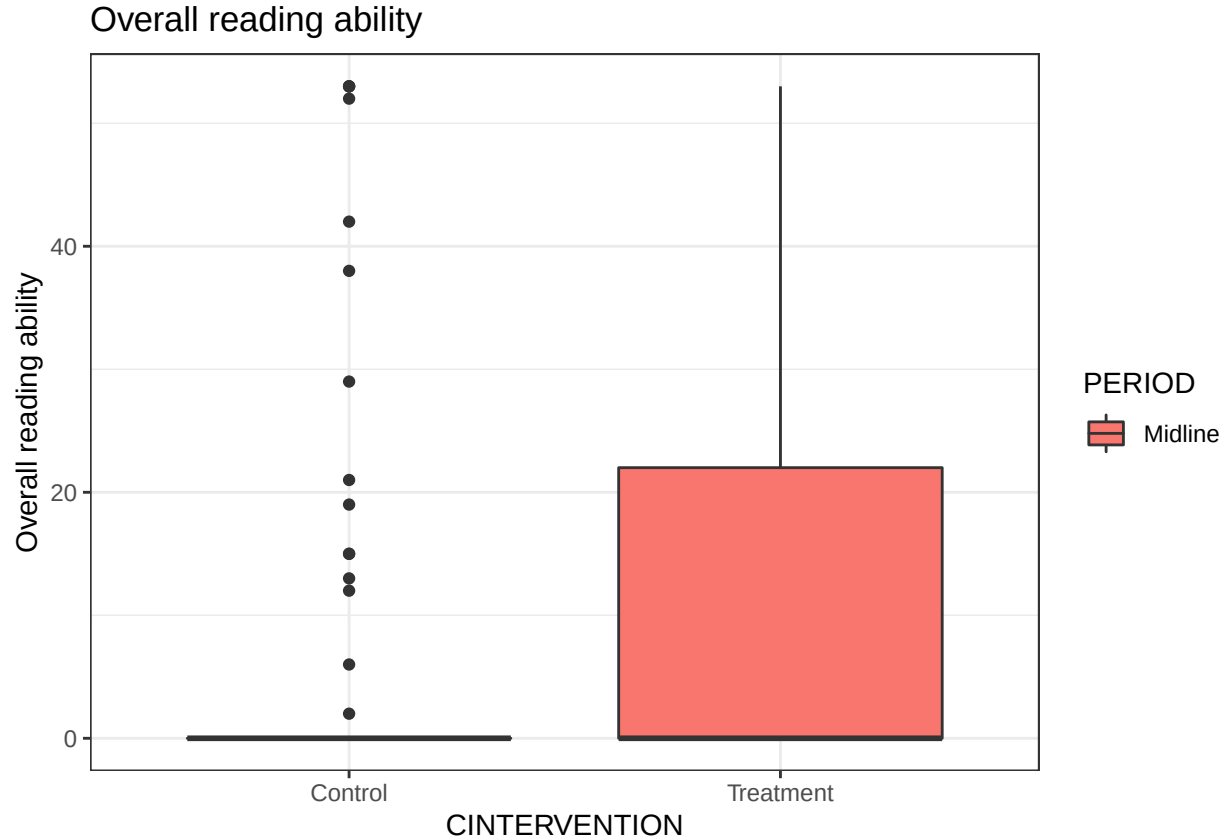
As shown in the table above, for the Overall reading ability EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at midline was 13.46598 (SD = 23.6169) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 12.45526 (SD = 22.36627). The difference for the Overall reading ability across the Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline was thus -1.010719 points (there was no baseline measurement for this variable). The p-value for this difference was 0.6293656. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the levels of Overall reading ability between the Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline beyond that which could result from sampling error.

1.10.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 48: Overall reading ability

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	3.917	11.891	108	0	53
Treatment	NA	NA	0	NA	NA	11.860	18.487	742	0	53



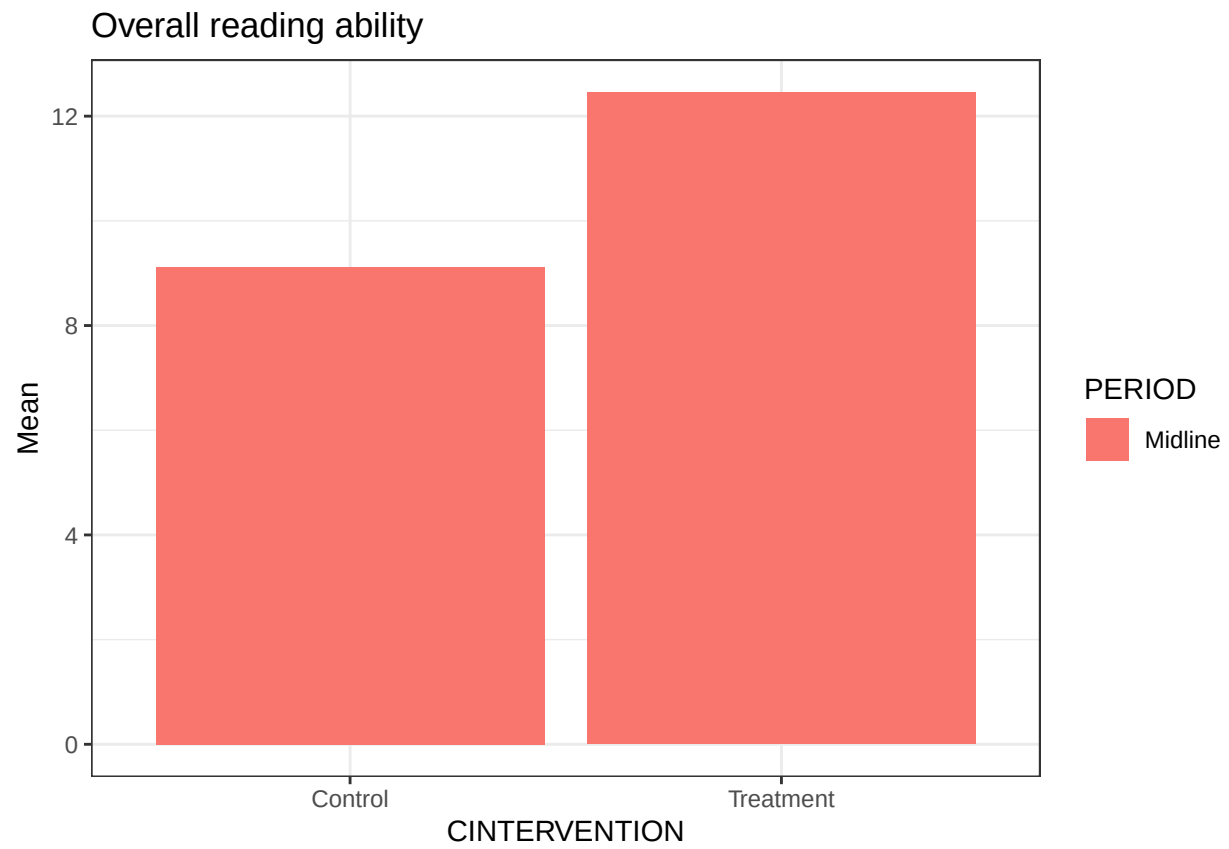


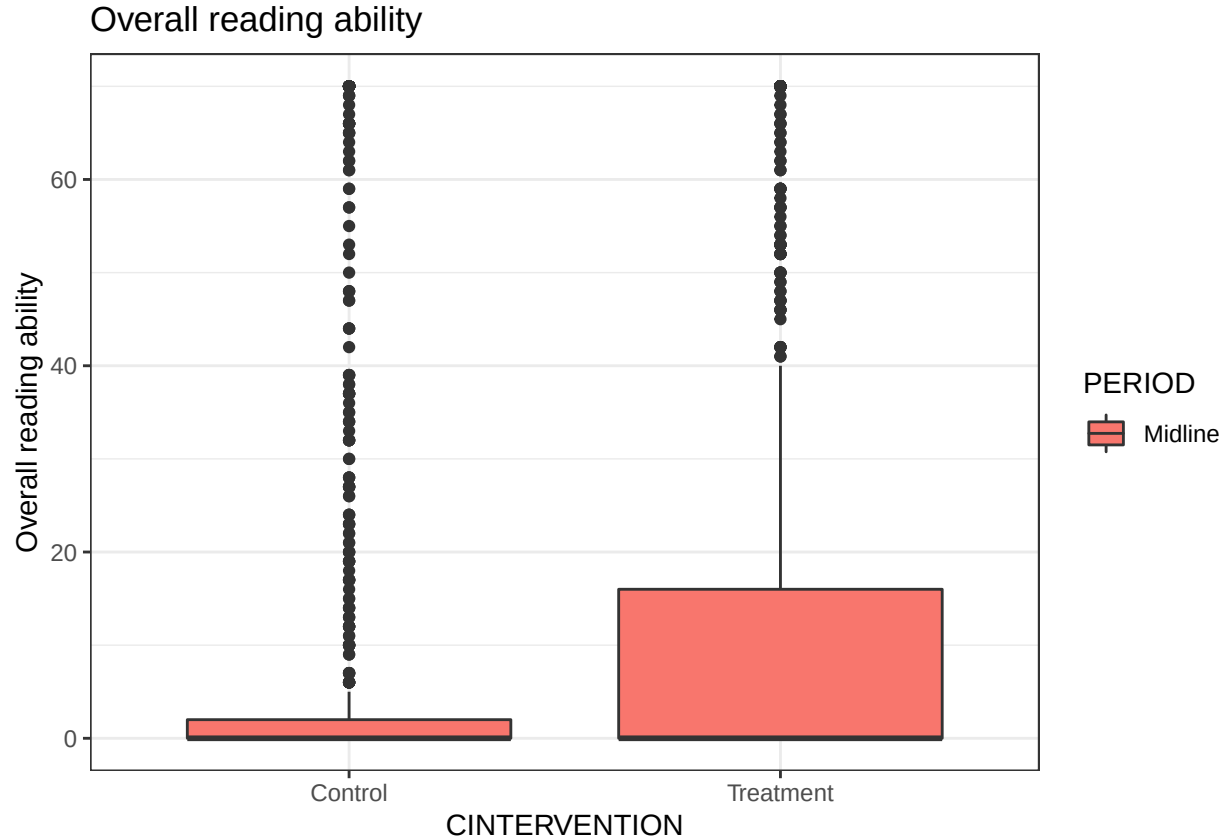
As shown in the table above, for the Overall reading ability EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at midline was 3.916667 (SD = 11.89056) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 11.85984 (SD = 18.48731). The difference for the Overall reading ability across the Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) conditions at midline was thus 7.943172 points (there was no baseline measurement for this variable). The p-value for this difference was 0.001604491. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts). Nevertheless, due to the absence of baseline values and lack of random allocation to the treatment and control conditions, a causal attribution of this effect to the treatment is not possible

1.10.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 49: Overall reading ability

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	9.120	20.498	950	0	70
Treatment	NA	NA	0	NA	NA	12.455	22.366	1017	0	70



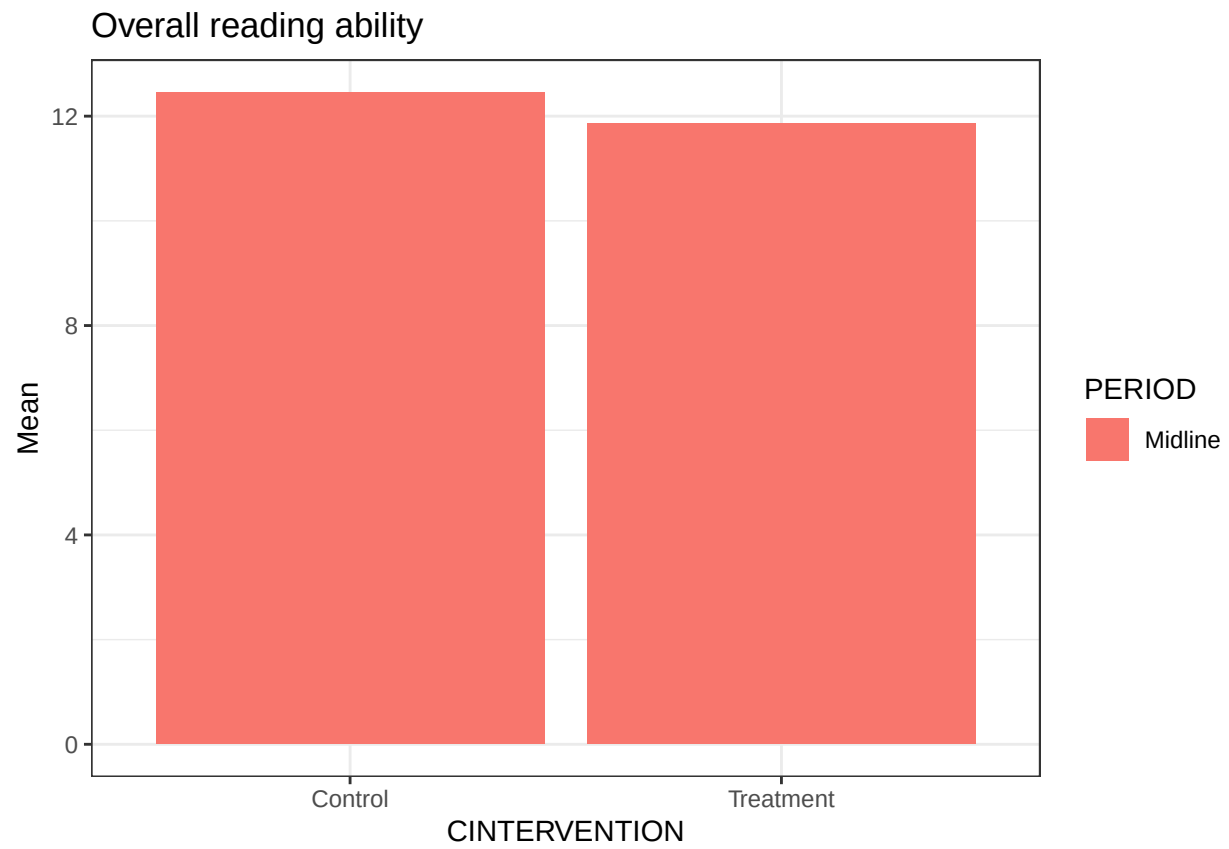


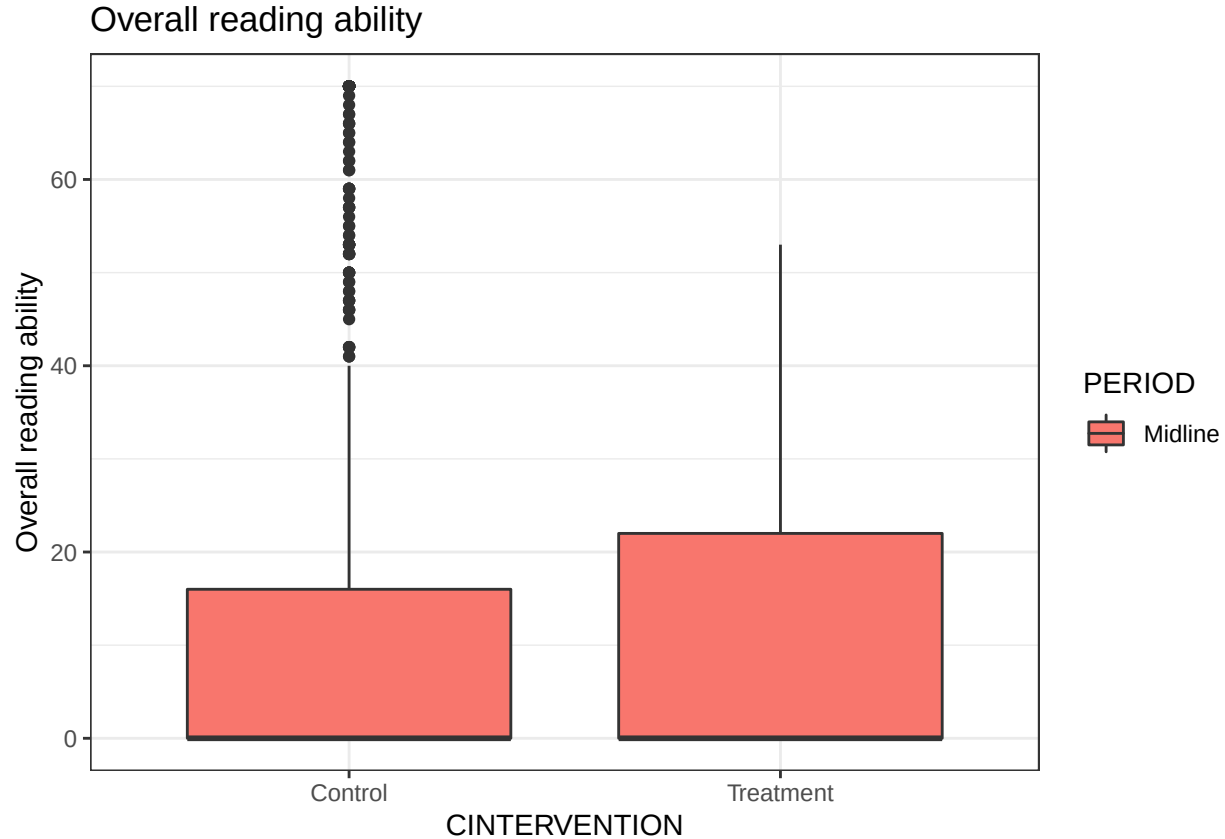
As shown in the table above, for the Overall reading ability EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at midline was 9.12 (SD = 20.49843) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 12.45526 (SD = 22.36627). The difference for the Overall reading ability across the Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline was thus 3.335261 points (there was no baseline measurement for this variable). The p-value for this difference was 0.1007097. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the levels of Overall reading ability between the Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline beyond that which could result from sampling error.

1.10.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 50: Overall reading ability

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	12.455	22.366	1017	0	70
Treatment	NA	NA	0	NA	NA	11.860	18.487	742	0	53





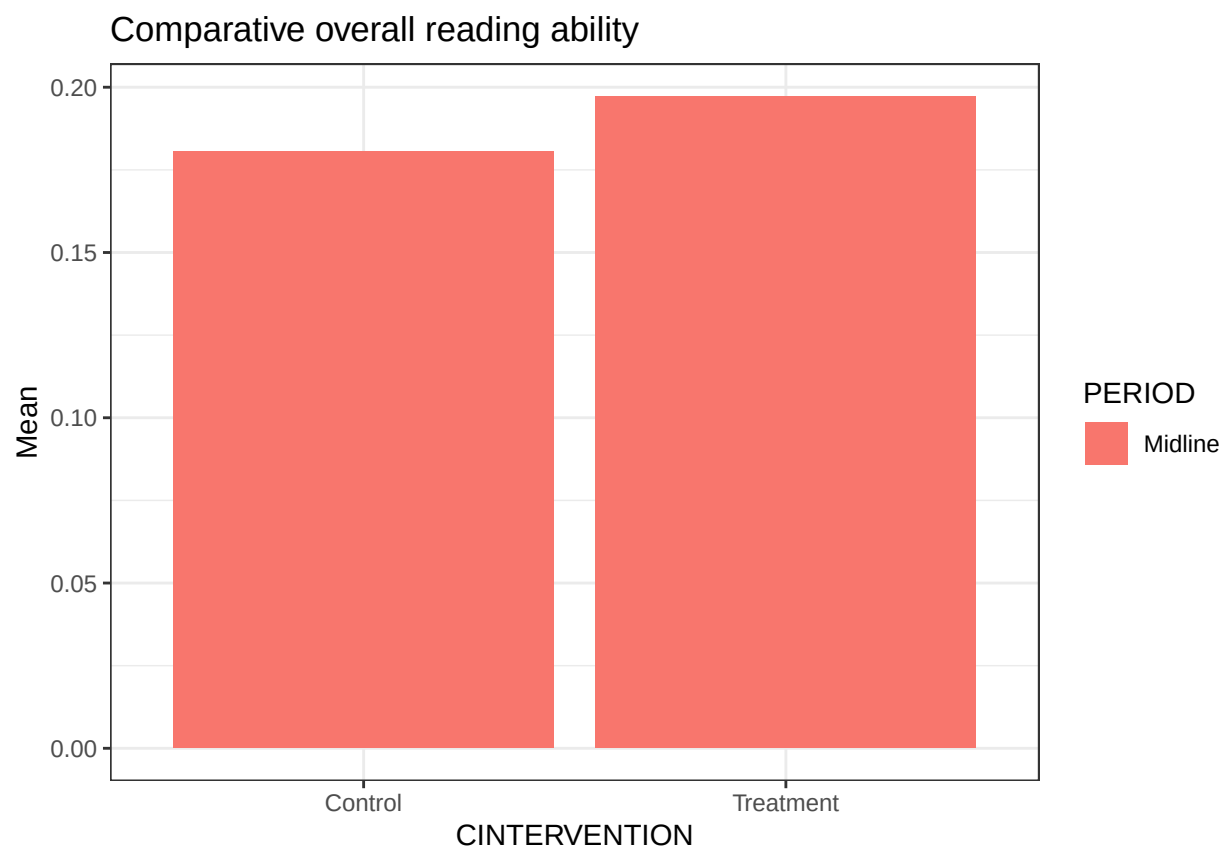
As shown in the table above, for the Overall reading ability EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at midline was 12.45526 (SD = 22.36627) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 11.85984 (SD = 18.48731). The difference for the Overall reading ability across the Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) conditions at midline was thus -0.5954223 points (there was no baseline measurement for this variable). The p-value for this difference was 0.7526363. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the levels of Overall reading ability between the Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) conditions at midline beyond that which could result from sampling error.

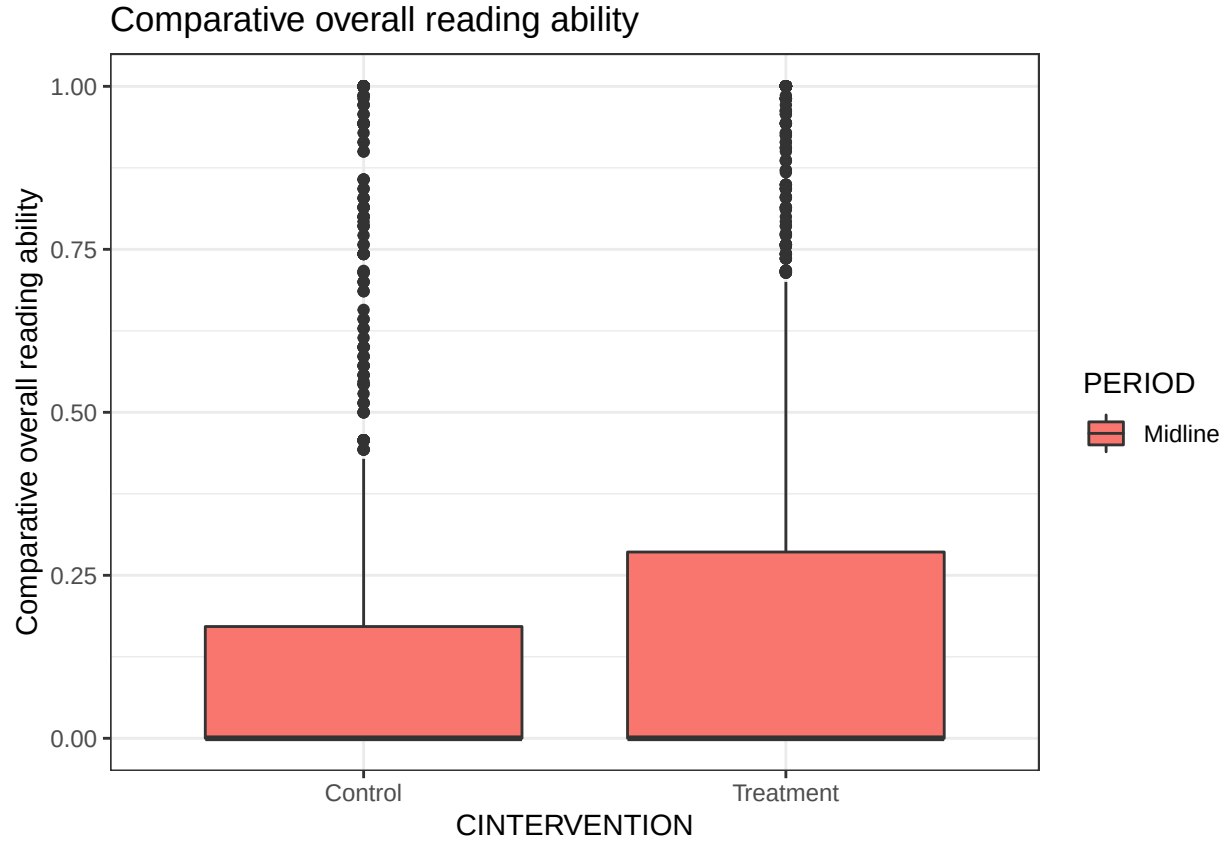
1.11 EGRA_ST8_1BS: Comparative overall reading ability

1.11.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 51: Comparative overall reading ability

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	0.181	0.330	1078	0	1
Treatment	NA	NA	0	NA	NA	0.197	0.333	1759	0	1



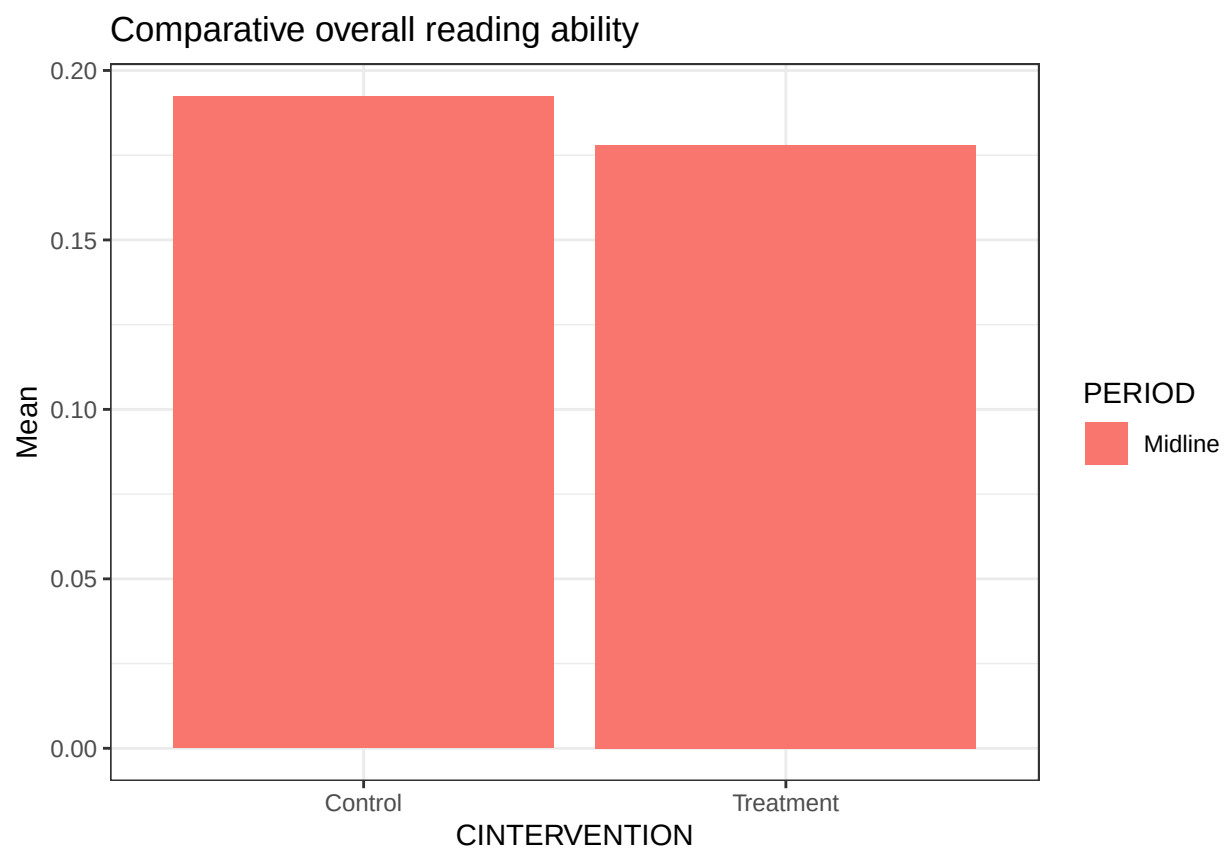


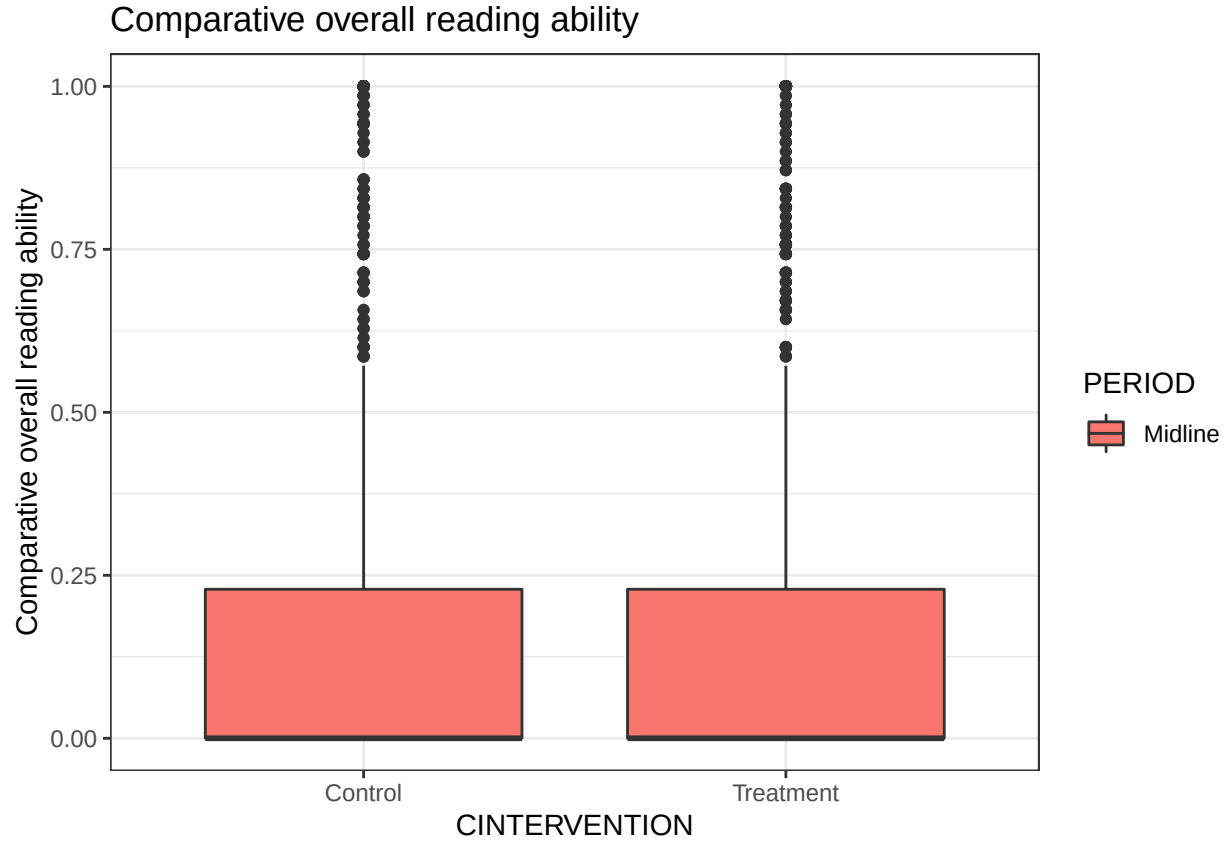
As shown in the table above, for the Comparative overall reading ability EGRA subtask, the mean for the Control (Comparison (all)) condition at midline was 0.180502 (SD = 0.3296675) and the mean for the Treatment (FFE + lit (all)) condition at midline was 0.1972422 (SD = 0.3328691). The difference for the Comparative overall reading ability across the Control (Comparison (all)) and Treatment (FFE + lit (all)) conditions at midline was thus 0.01674024 points (there was no baseline measurement for this variable). The p-value for this difference was 0.5162973. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the levels of Comparative overall reading ability between the Control (Comparison (all)) and Treatment (FFE + lit (all)) conditions at midline beyond that which could result from sampling error.

1.11.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 52: Comparative overall reading ability

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	0.192	0.337	970	0	1
Treatment	NA	NA	0	NA	NA	0.178	0.320	1017	0	1



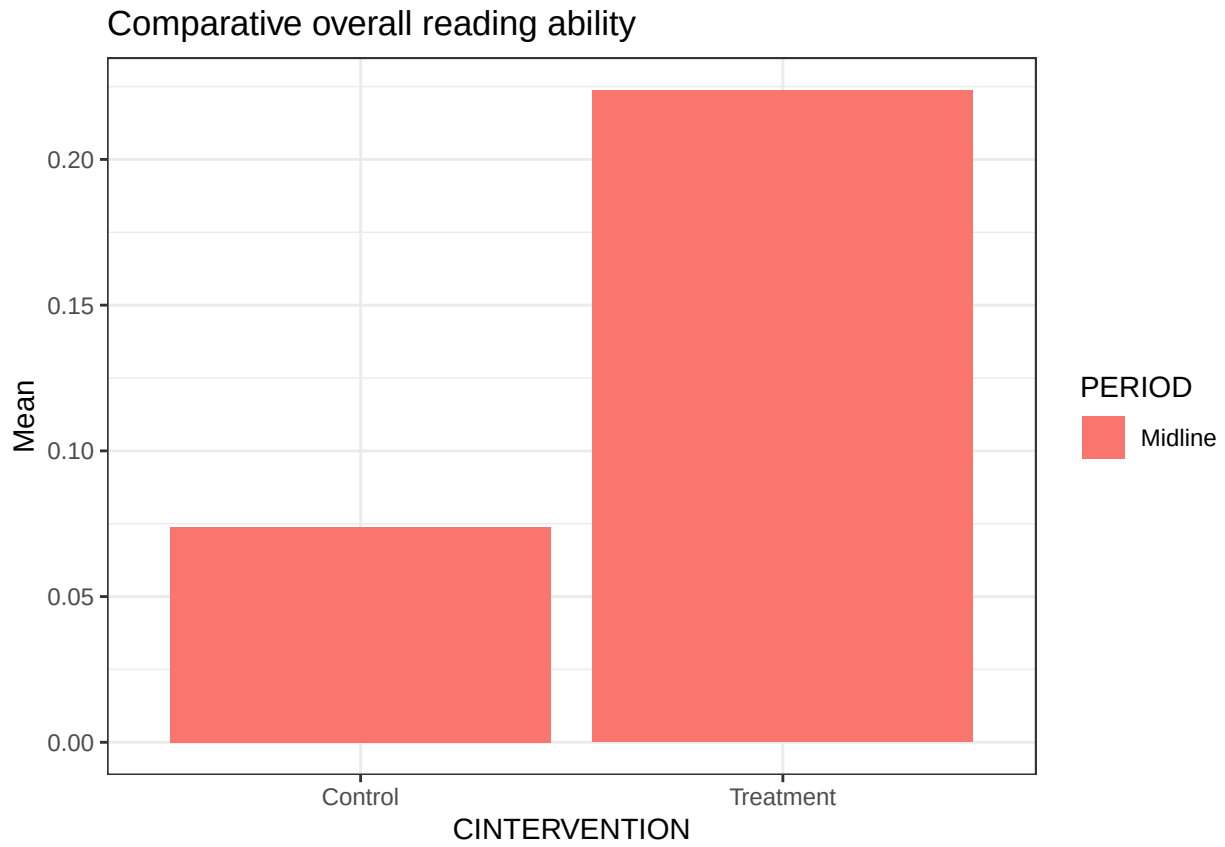


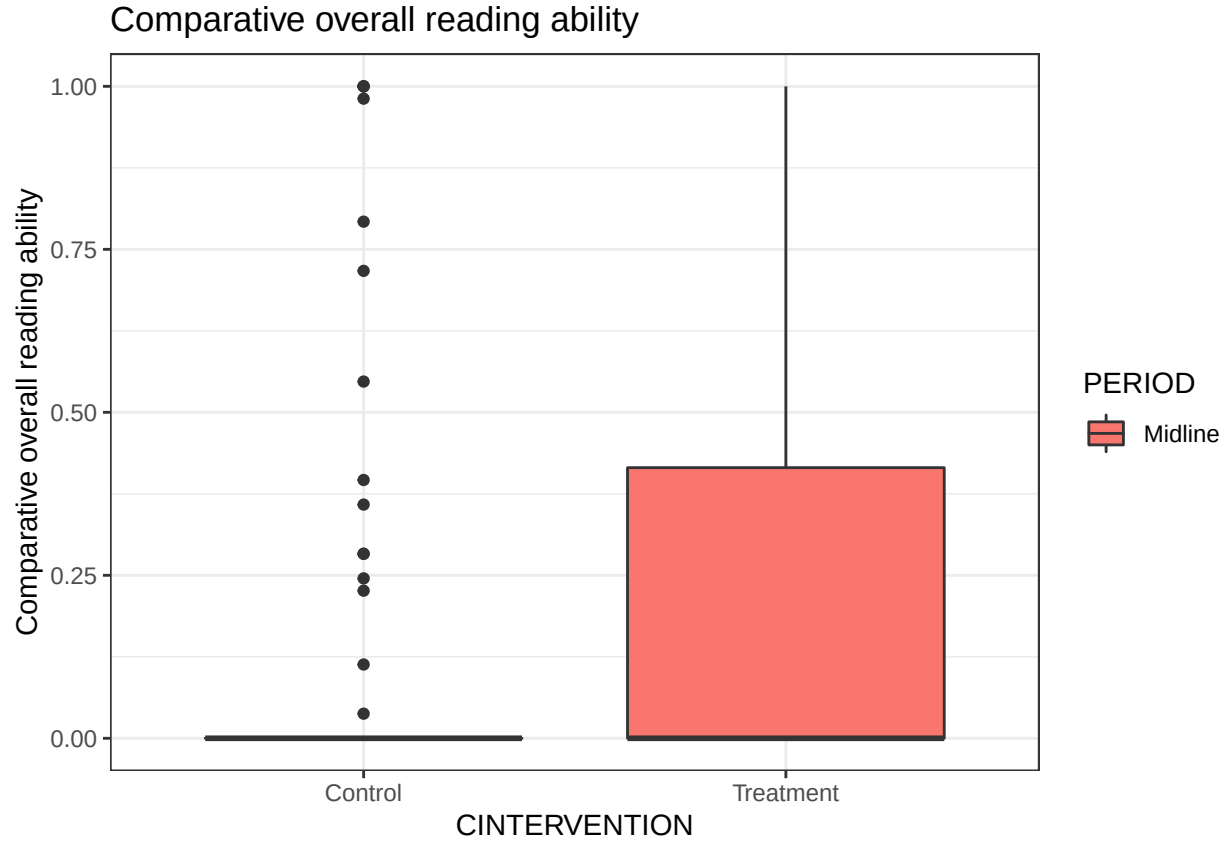
As shown in the table above, for the Comparative overall reading ability EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at midline was 0.1923711 (SD = 0.3373843) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 0.1779323 (SD = 0.3195182). The difference for the Comparative overall reading ability across the Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline was thus -0.01443884 points (there was no baseline measurement for this variable). The p-value for this difference was 0.6293656. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the levels of Comparative overall reading ability between the Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline beyond that which could result from sampling error.

1.11.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 53: Comparative overall reading ability

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	0.074	0.224	108	0	1
Treatment	NA	NA	0	NA	NA	0.224	0.349	742	0	1



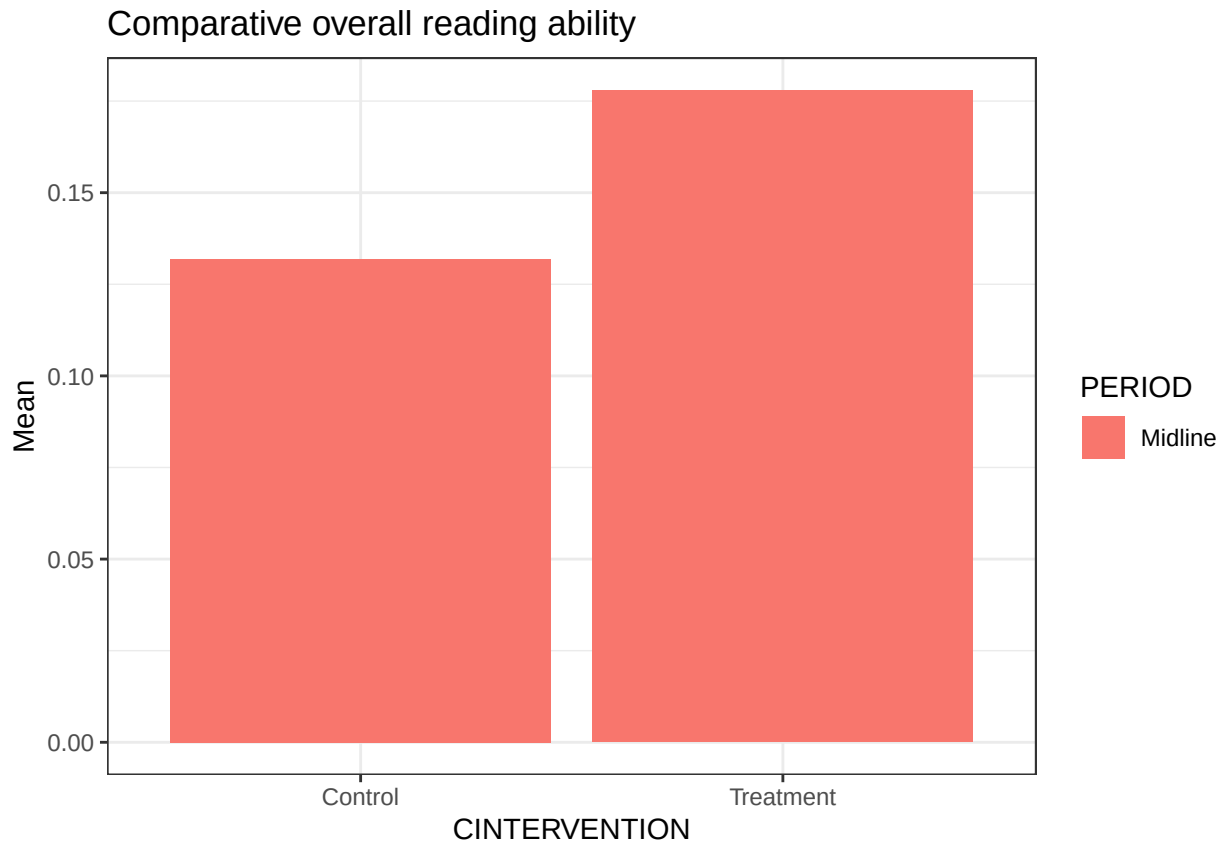


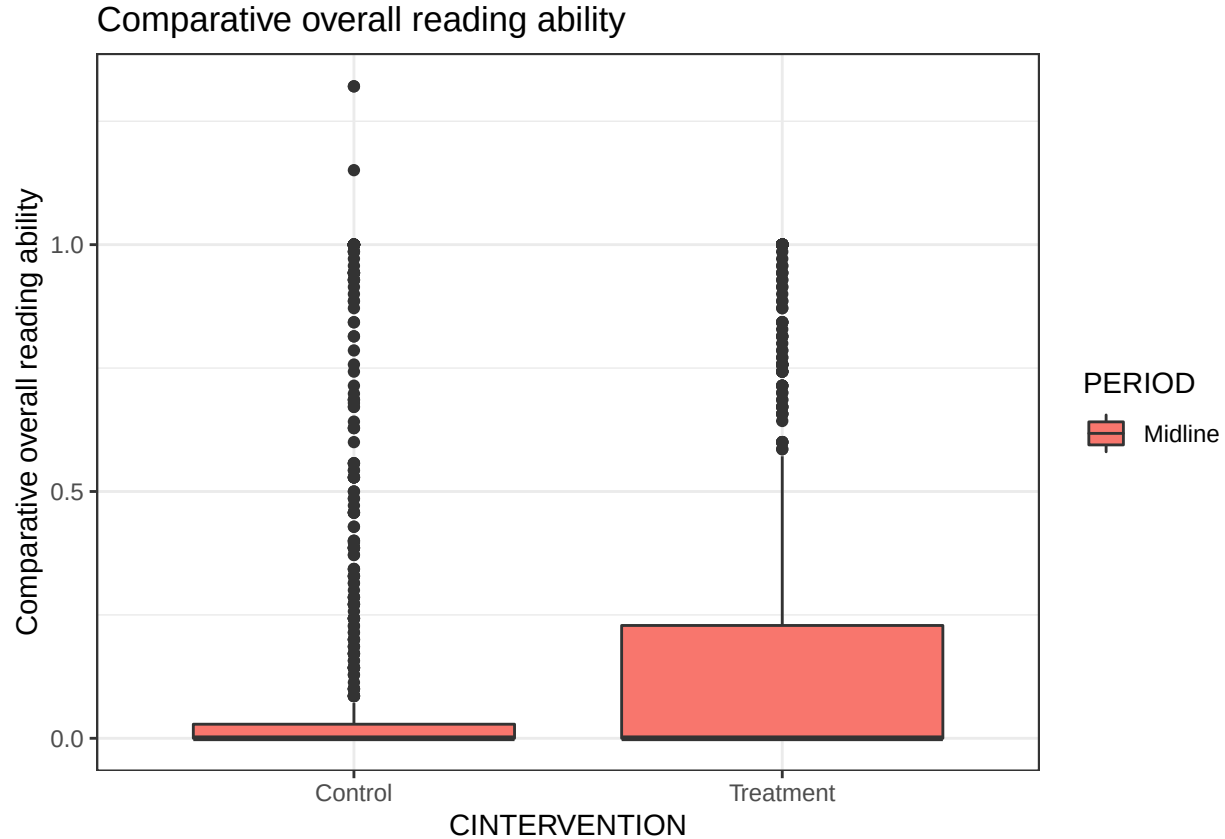
As shown in the table above, for the Comparative overall reading ability EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at midline was 0.07389937 (SD = 0.2243503) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 0.2237088 (SD = 0.3488274). The difference for the Comparative overall reading ability across the Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) conditions at midline was thus 0.1498094 points (there was no baseline measurement for this variable). The p-value for this difference was 0.001613896. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts). Nevertheless, due to the absence of baseline values and lack of random allocation to the treatment and control conditions, a causal attribution of this effect to the treatment is not possible

1.11.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 54: Comparative overall reading ability

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	0.132	0.297	950	0	1.321
Treatment	NA	NA	0	NA	NA	0.178	0.320	1017	0	1.000



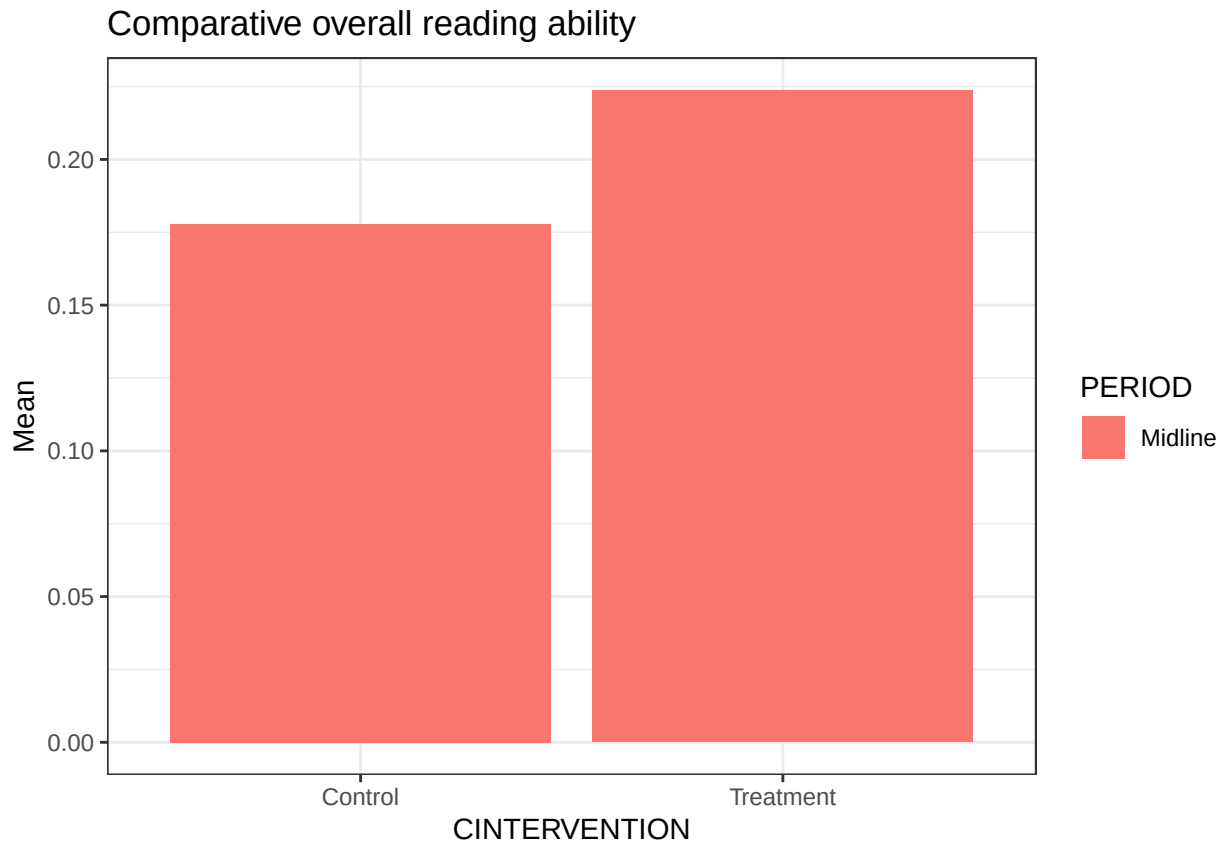


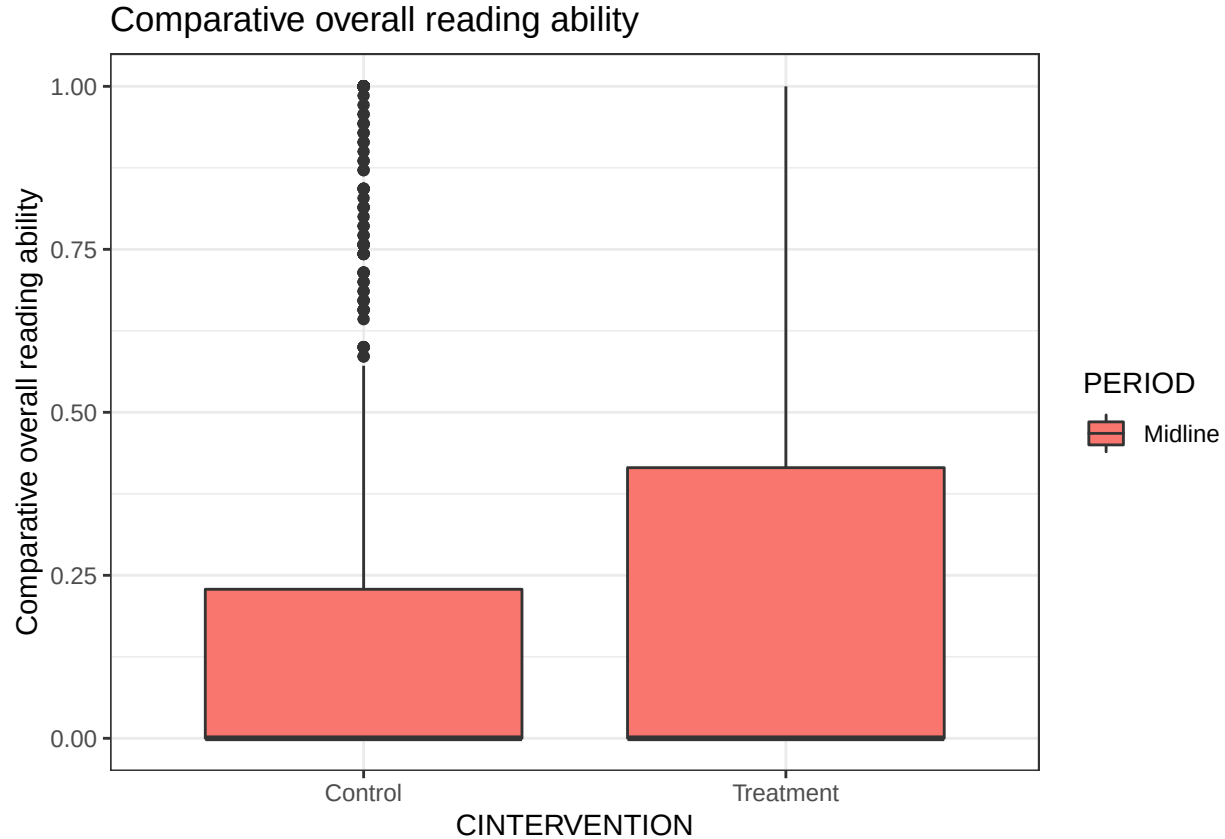
As shown in the table above, for the Comparative overall reading ability EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at midline was 0.1319594 (SD = 0.2969652) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 0.1779323 (SD = 0.3195182). The difference for the Comparative overall reading ability across the Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline was thus 0.04597287 points (there was no baseline measurement for this variable). The p-value for this difference was 0.1161117. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the levels of Comparative overall reading ability between the Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline beyond that which could result from sampling error.

1.11.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 55: Comparative overall reading ability

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	0.178	0.320	1017	0	1
Treatment	NA	NA	0	NA	NA	0.224	0.349	742	0	1





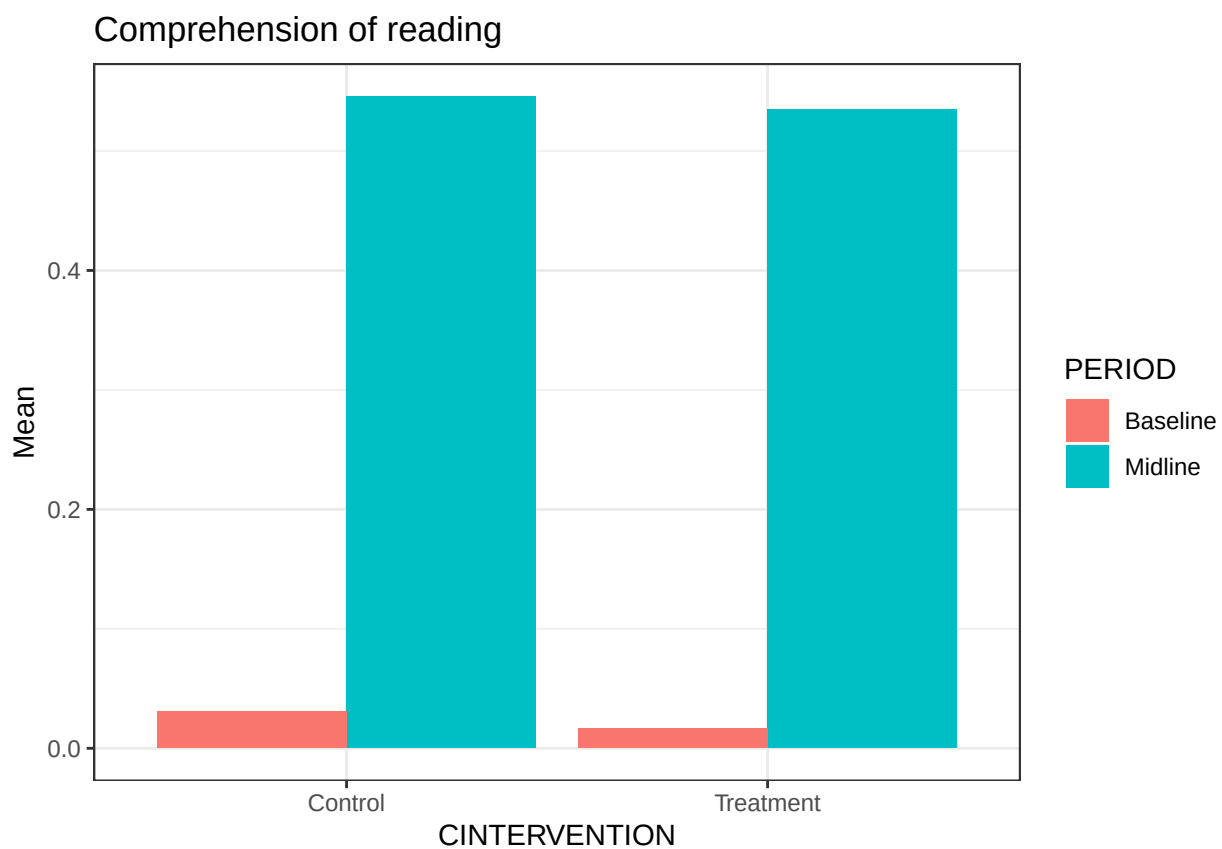
As shown in the table above, for the Comparative overall reading ability EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at midline was 0.1779323 (SD = 0.3195182) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 0.2237088 (SD = 0.3488274). The difference for the Comparative overall reading ability across the Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) conditions at midline was thus 0.04577648 points (there was no baseline measurement for this variable). The p-value for this difference was 0.1376108. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the levels of Comparative overall reading ability between the Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) conditions at midline beyond that which could result from sampling error.

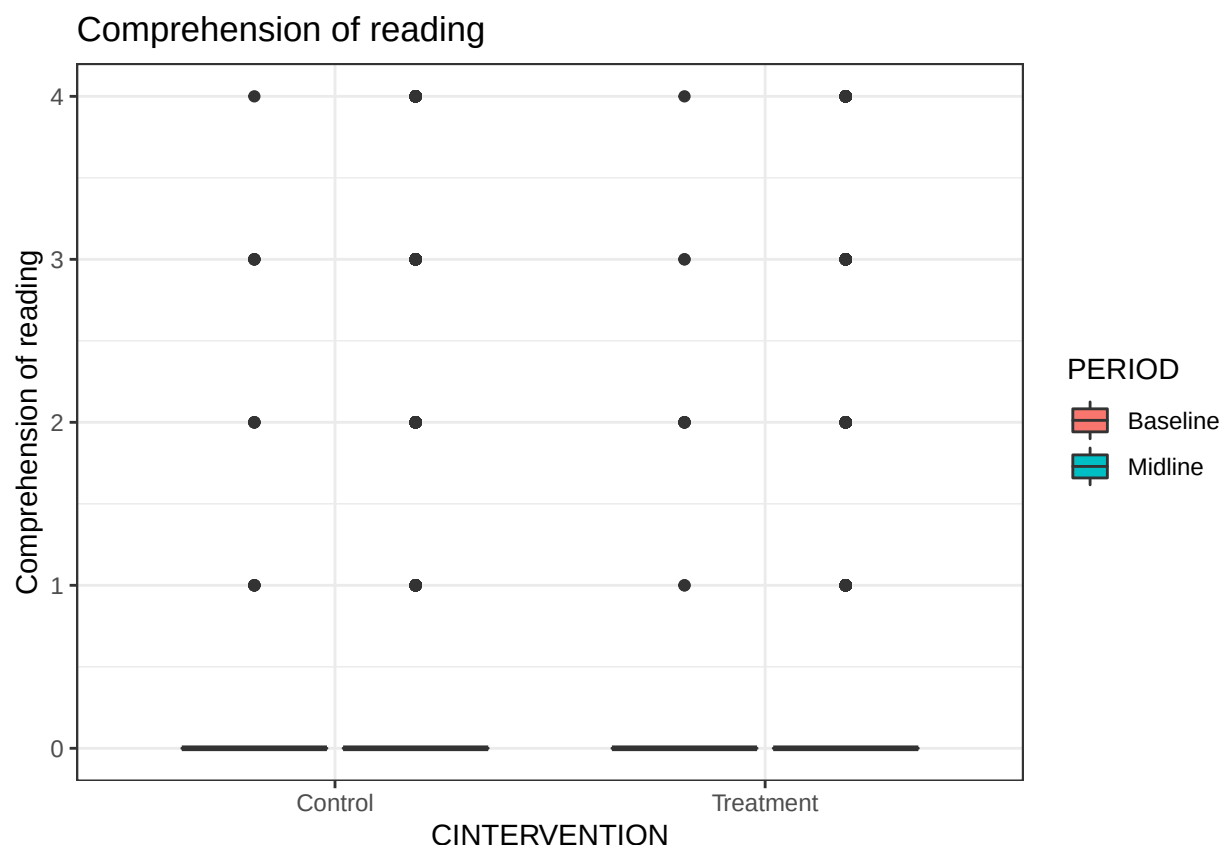
1.12 EGRA_ST8_2: Comprehension of reading

1.12.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 56: Comprehension of reading

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.031	0.275	1136	0	4	0.546	1.180	1081	0	4
Treatment	0.016	0.201	1578	0	4	0.535	1.143	1808	0	4



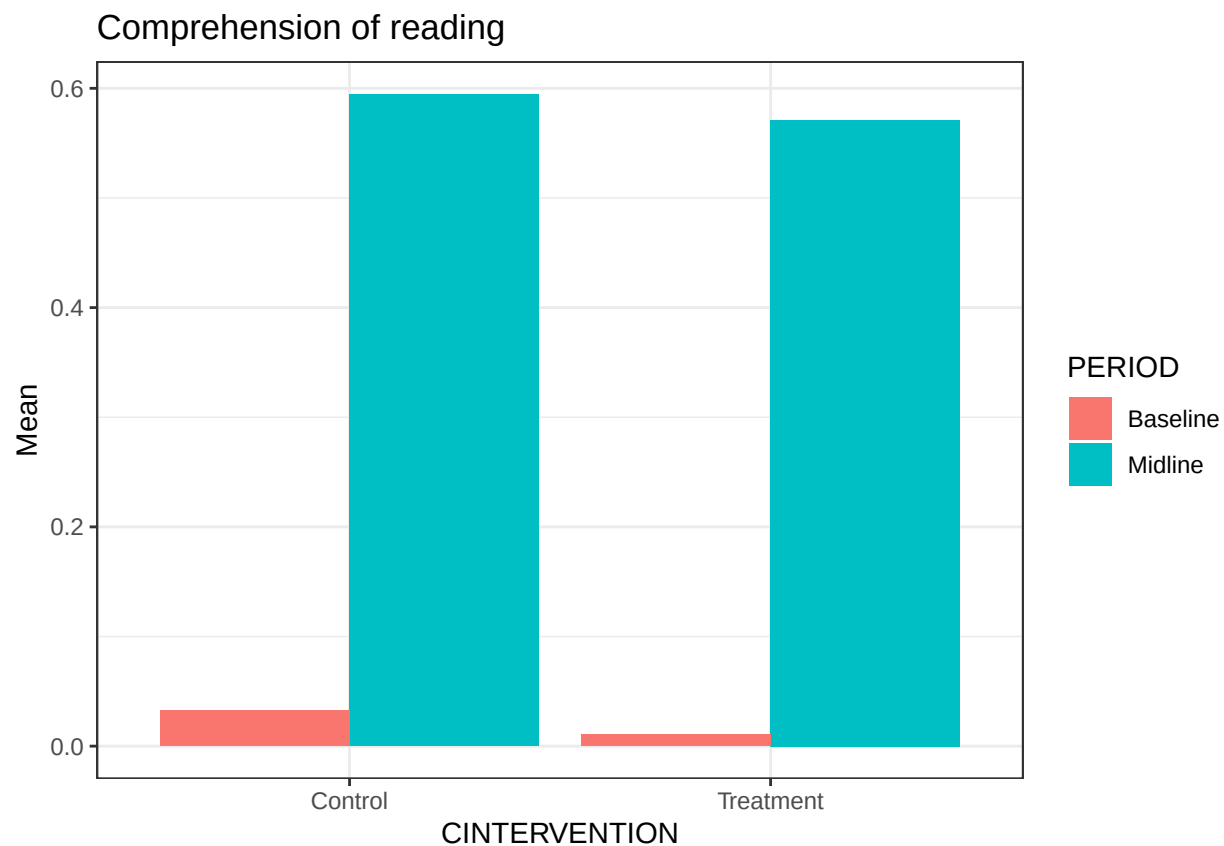


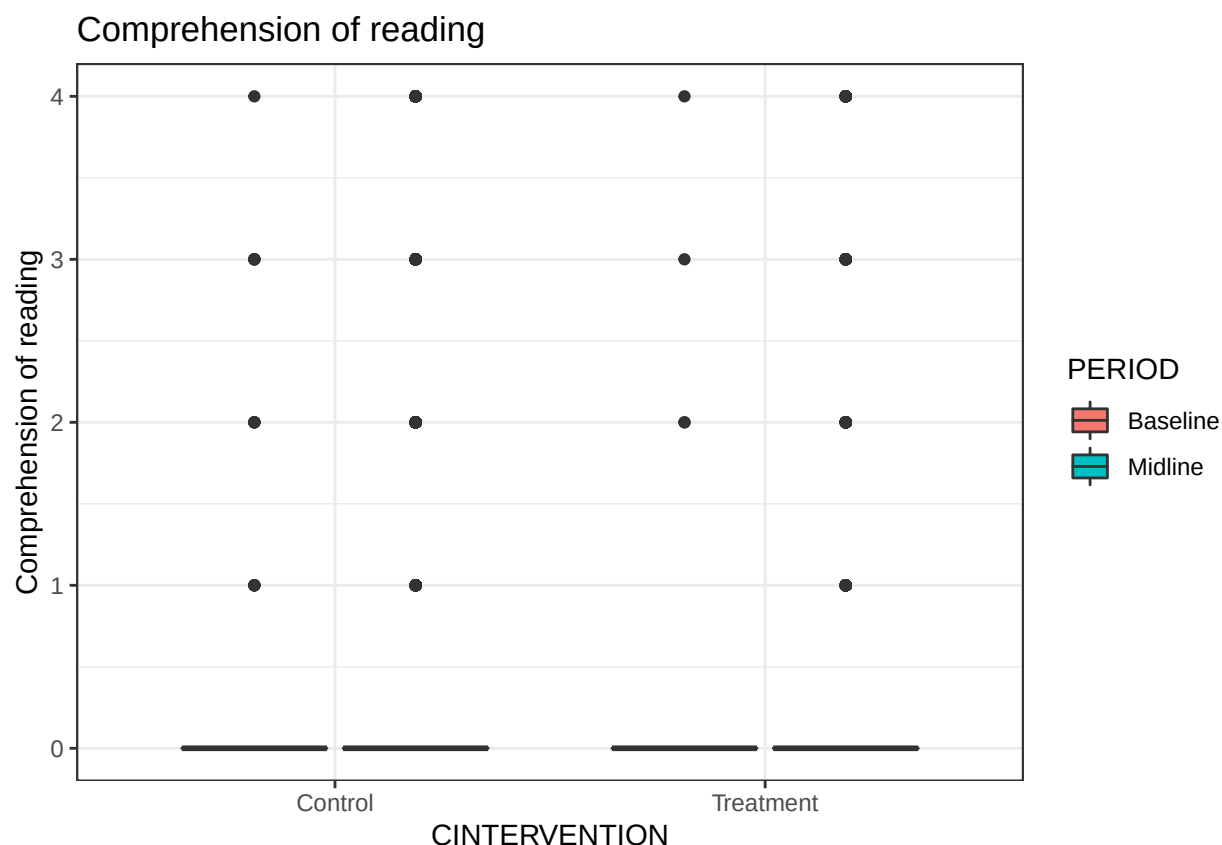
As shown in the table above, for the the Comprehension of reading EGRA subtask, the mean for the Control (Comparison (all)) condition at baseline was 0.03080986 (SD = 0.2751398) and the mean for the Treatment (FFE + lit (all)) condition at baseline was 0.01647655 (SD = 0.2007778). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at baseline was thus of 0.01433331 points. The p-value for this difference was 0.3039468. The mean for the Control (Comparison (all)) condition at midline was 0.5457909 (SD = 1.180075) and the mean for the Treatment (FFE + lit (all)) condition at midline was 0.5348451 (SD = 1.14277). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at midline was thus of 0.0109458 points. The p-value for this difference was 0.8947624. The change from the baseline to the midline of 0.5149811 points for the Control (Comparison (all)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (all)) can be assessed against. The change for the Treatment (FFE + lit (all)) from the baseline to the midline was of 0.5183686 points. Consequently, the change for the Control (Comparison (all)) relative to the change for the Treatment (FFE + lit (all)) condition was of 0.003387505 points. The p-value for this difference was 0.967264. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Comprehension of reading EGRA subtask across the Control (Comparison (all)) and the Treatment (FFE + lit (all)) conditions beyond that expected from sampling error.

1.12.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 57: Comprehension of reading

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.032	0.285	1047	0	4	0.595	1.223	972	0	4
Treatment	0.011	0.178	1040	0	4	0.571	1.201	1047	0	4



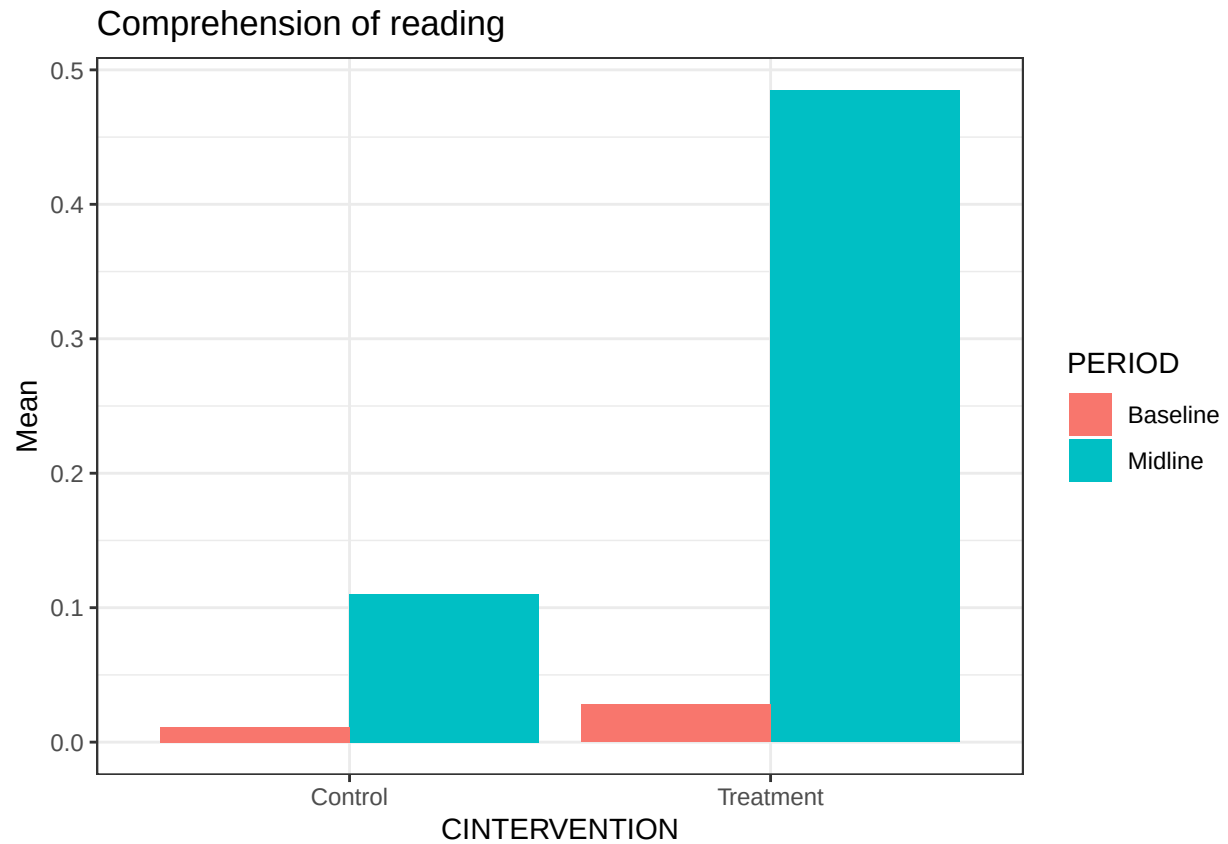


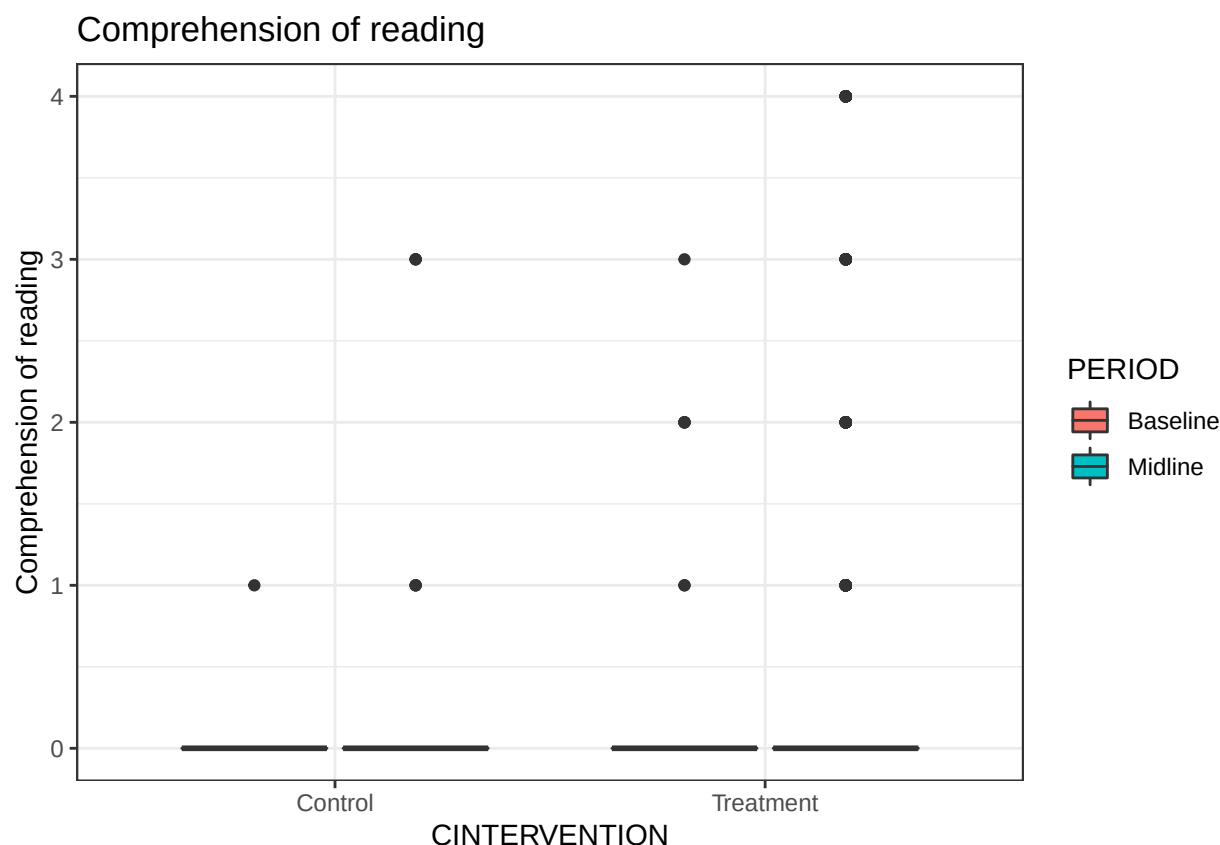
As shown in the table above, for the the Comprehension of reading EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at baseline was 0.03247373 (SD = 0.2848902) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.01057692 (SD = 0.1779026). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.02189681 points. The p-value for this difference was 0.1397576. The mean for the Control (Comparison (Portuguese)) condition at midline was 0.5946502 (SD = 1.222974) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 0.5711557 (SD = 1.20088). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.02349452 points. The p-value for this difference was 0.8094379. The change from the baseline to the midline of 0.5621765 points for the Control (Comparison (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 0.5605788 points. Consequently, the change for the Control (Comparison (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of -0.001597711 points. The p-value for this difference was 0.9868452. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Comprehension of reading EGRA subtask across the Control (Comparison (Portuguese)) and the Treatment (FFE + lit (Portuguese)) conditions beyond that expected from sampling error.

1.12.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 58: Comprehension of reading

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.011	0.106	89	0	1	0.110	0.515	109	0	3
Treatment	0.028	0.239	538	0	3	0.485	1.056	761	0	4



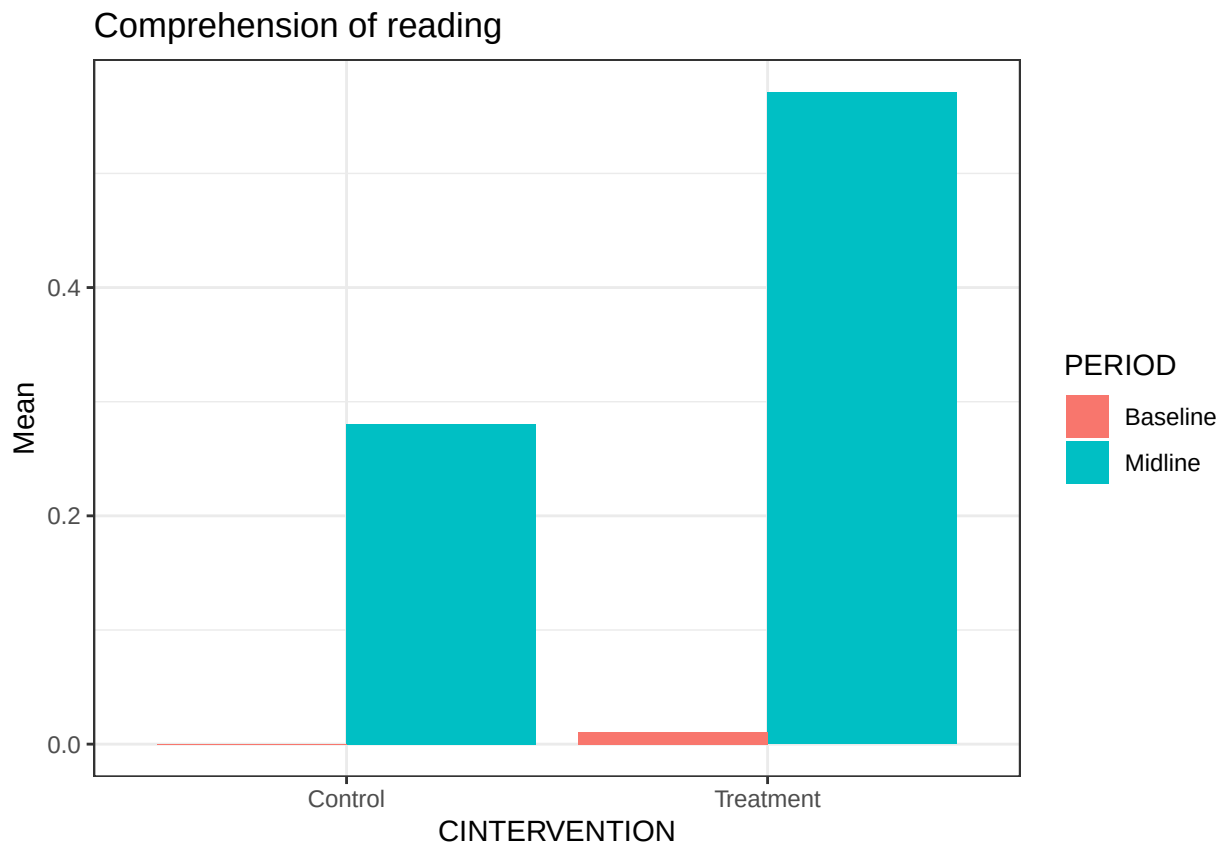


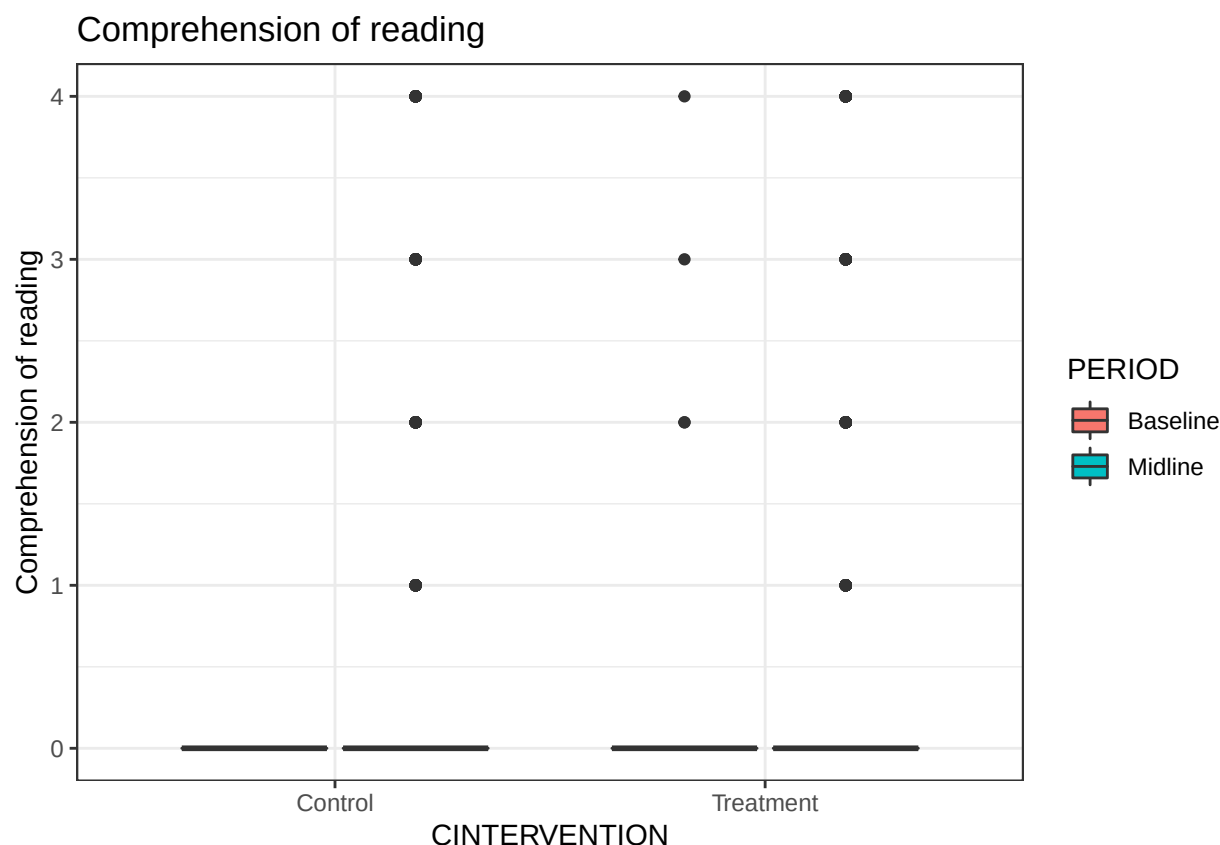
As shown in the table above, for the the Comprehension of reading EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at baseline was 0.01123596 (SD = 0.1059998) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.02788104 (SD = 0.2386406). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.01664509 points. The p-value for this difference was 0.3408817. The mean for the Control (Comparison (Bilingual)) condition at midline was 0.1100917 (SD = 0.5153109) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 0.4848883 (SD = 1.056357). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.3747966 points. The p-value for this difference was 0.000555817. The change from the baseline to the midline of 0.09885579 points for the Control (Comparison (Bilingual)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 0.4570073 points. Consequently, the change for the Control (Comparison (Bilingual)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 0.3581515 points. The p-value for this difference was 0.001274507. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Bilingual)) when compared to the Control (Comparison (Bilingual)) condition. This provides evidence that the Treatment (FFE + lit (Bilingual)) (or some other unobserved process) impacted on the Comprehension of reading EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

1.12.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 59: Comprehension of reading

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.000	0.000	1007	0	0	0.280	0.879	952	0	4
Treatment	0.011	0.178	1040	0	4	0.571	1.201	1047	0	4



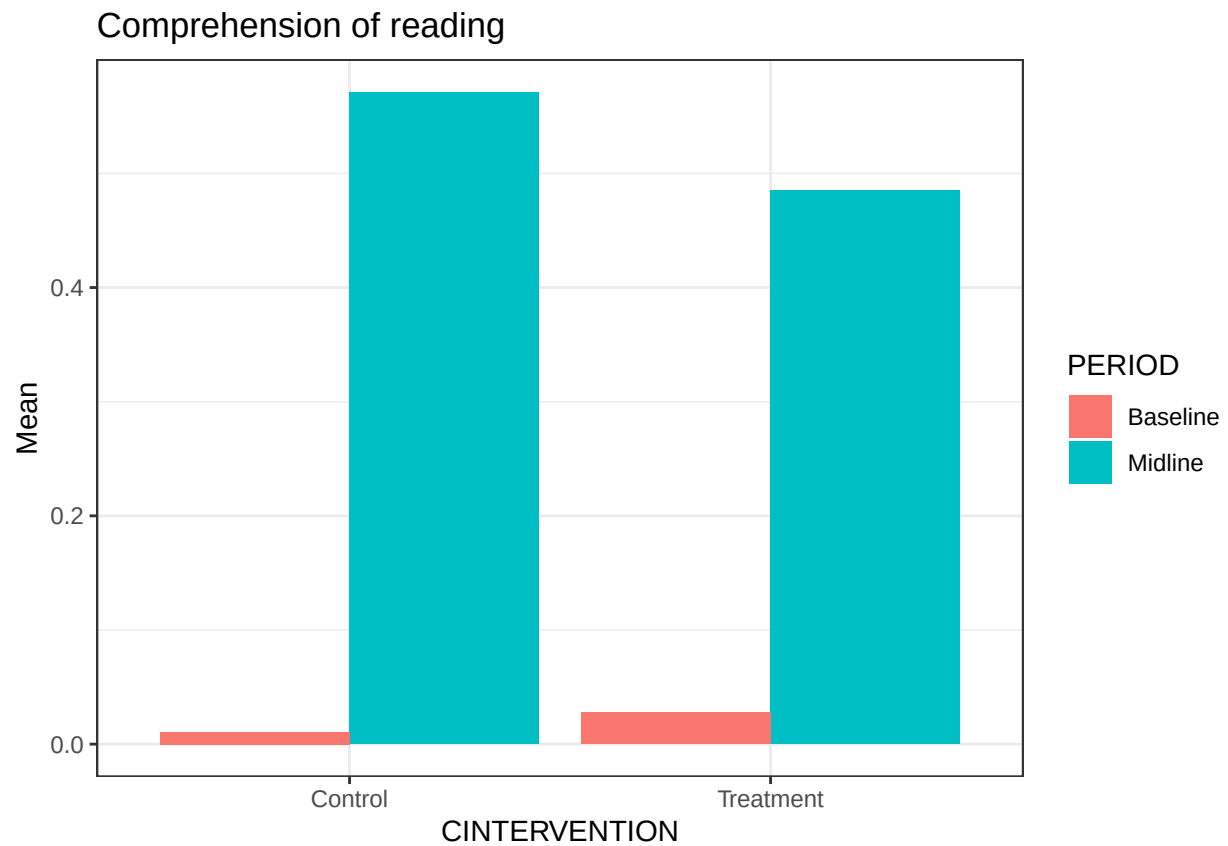


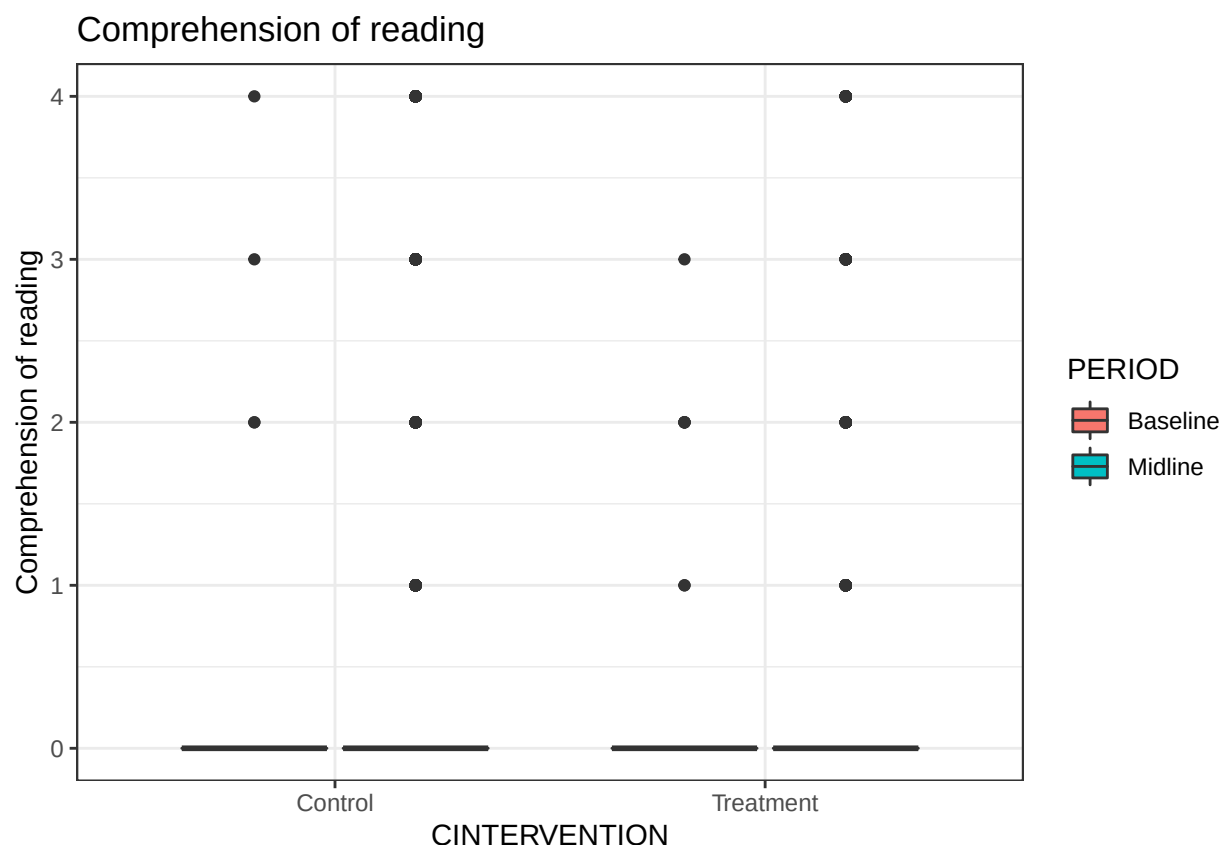
As shown in the table above, for the the Comprehension of reading EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at baseline was 0 (SD = 0) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.01057692 (SD = 0.1779026). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.01057692 points. The p-value for this difference was 0.08336649. The mean for the Control (FFE only (Portuguese)) condition at midline was 0.2804622 (SD = 0.8786022) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 0.5711557 (SD = 1.20088). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.2906935 points. The p-value for this difference was 0.001308768. The change from the baseline to the midline of 0.2804622 points for the Control (FFE only (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 0.5605788 points. Consequently, the change for the Control (FFE only (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 0.2801166 points. The p-value for this difference was 0.001853958. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Portuguese)) when compared to the Control (FFE only (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Portuguese)) (or some other unobserved process) impacted on the Comprehension of reading EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

1.12.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 60: Comprehension of reading

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.011	0.178	1040	0	4	0.571	1.201	1047	0	4
Treatment	0.028	0.239	538	0	3	0.485	1.056	761	0	4





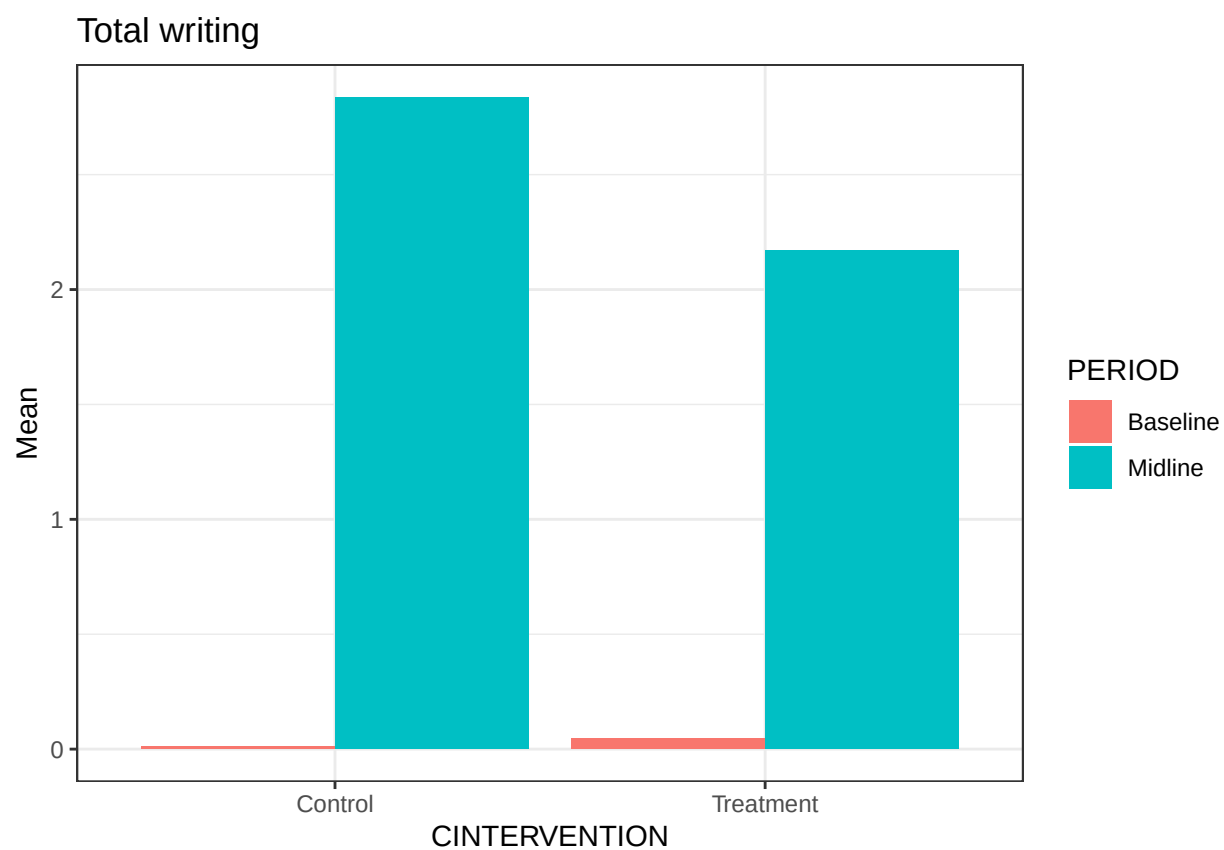
As shown in the table above, for the the Comprehension of reading EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at baseline was 0.01057692 (SD = 0.1779026) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.02788104 (SD = 0.2386406). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.01730412 points. The p-value for this difference was 0.2648872. The mean for the Control (FFE + lit (Portuguese)) condition at midline was 0.5711557 (SD = 1.20088) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 0.4848883 (SD = 1.056357). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.08626738 points. The p-value for this difference was 0.366519. The change from the baseline to the midline of 0.5605788 points for the Control (FFE + lit (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 0.4570073 points. Consequently, the change for the Control (FFE + lit (Portuguese)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of -0.1035715 points. The p-value for this difference was 0.2842989. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Comprehension of reading EGRA subtask across the Control (FFE + lit (Portuguese)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

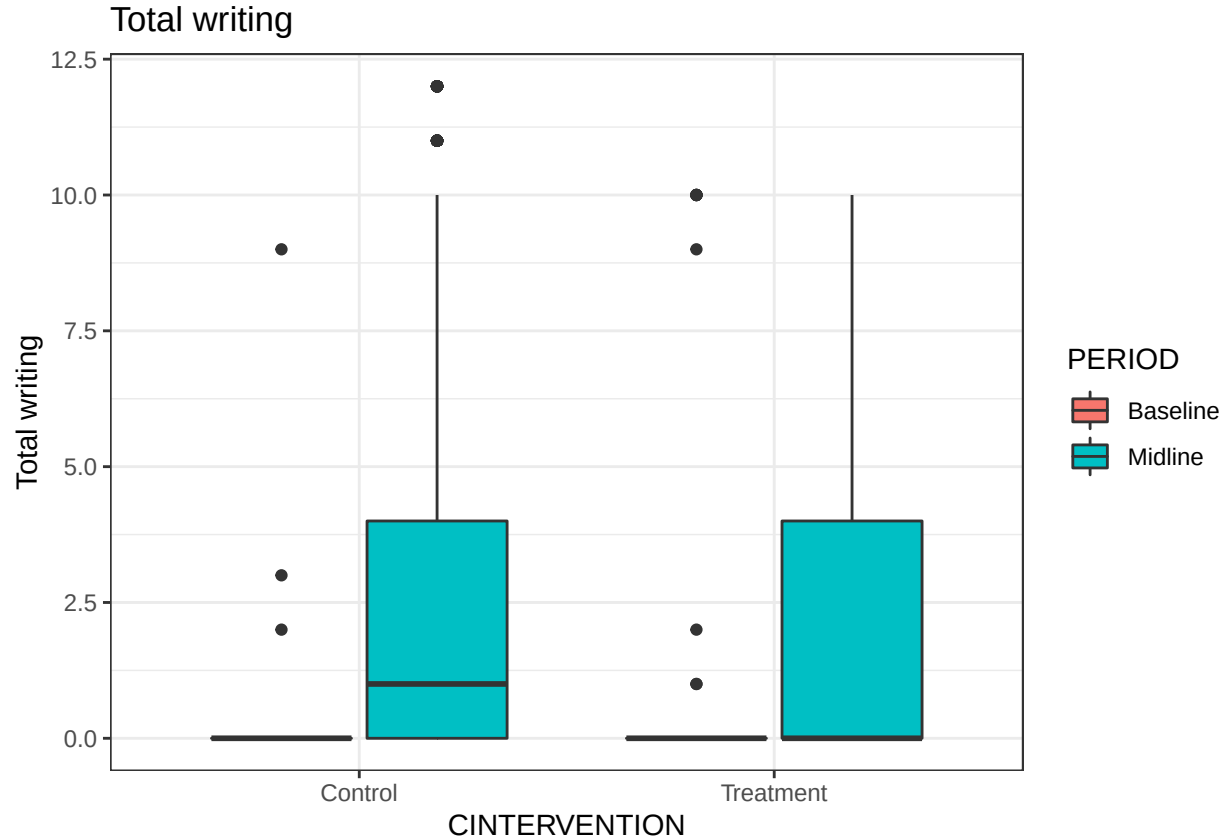
1.13 EGRA_ST9_2A: Total writing

1.13.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 61: Total writing

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.012	0.288	1136	0	9	2.837	3.935	1081	0	12
Treatment	0.046	0.658	1578	0	10	2.170	3.653	1808	0	10



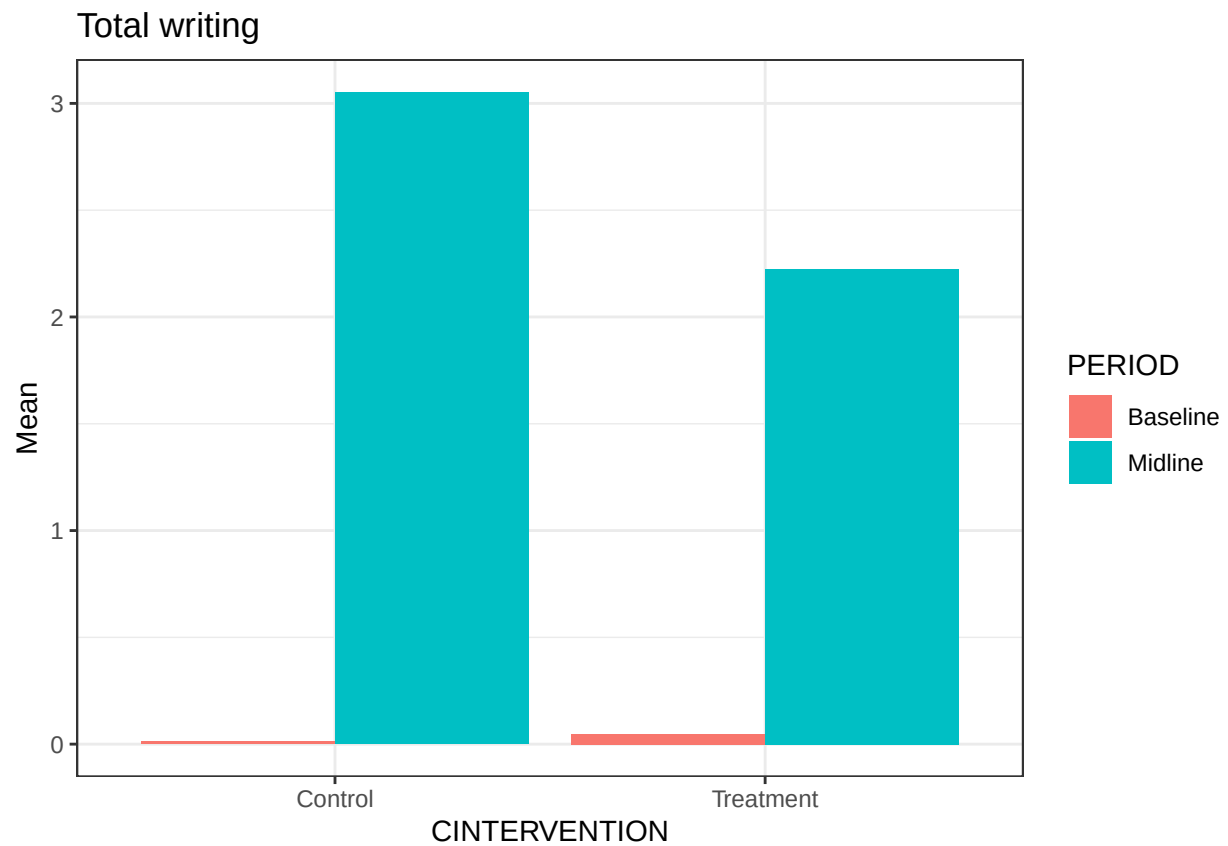


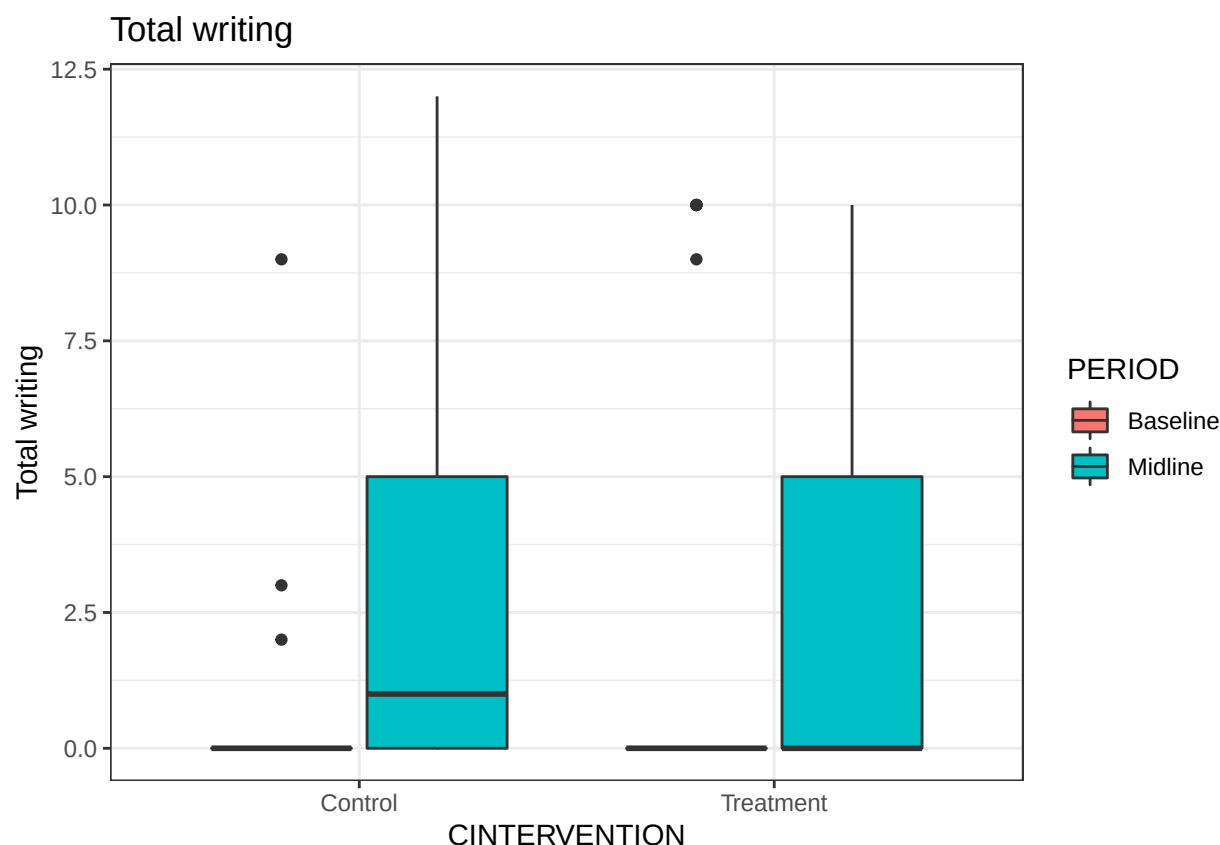
As shown in the table above, for the the Total writing EGRA subtask, the mean for the Control (Comparison (all)) condition at baseline was 0.01232394 (SD = 0.2875193) and the mean for the Treatment (FFE + lit (all)) condition at baseline was 0.04626109 (SD = 0.658404). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at baseline was thus of 0.03393715 points. The p-value for this difference was 0.06935504. The mean for the Control (Comparison (all)) condition at midline was 2.837188 (SD = 3.934574) and the mean for the Treatment (FFE + lit (all)) condition at midline was 2.170354 (SD = 3.653165). The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at midline was thus of 0.6668338 points. The p-value for this difference was 0.01980337. The change from the baseline to the midline of 2.824864 points for the Control (Comparison (all)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (all)) can be assessed against. The change for the Treatment (FFE + lit (all)) from the baseline to the midline was of 2.124093 points. Consequently, the change for the Control (Comparison (all)) relative to the change for the Treatment (FFE + lit (all)) condition was of -0.700771 points. The p-value for this difference was 0.01444353. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Total writing EGRA subtask across the Control (Comparison (all)) and the Treatment (FFE + lit (all)) conditions beyond that expected from sampling error.

1.13.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 62: Total writing

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.013	0.299	1047	0	9	3.052	4.026	972	0	12
Treatment	0.047	0.679	1040	0	10	2.224	3.641	1047	0	10



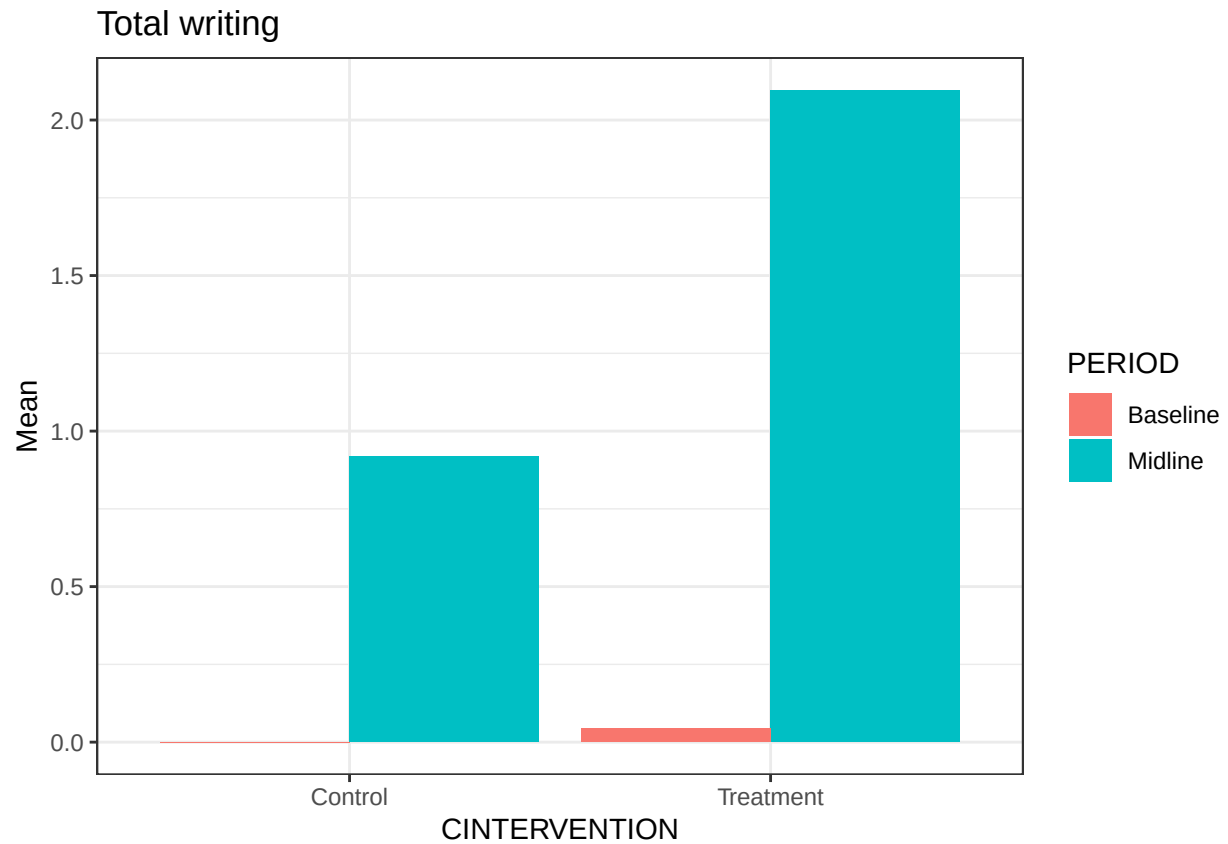


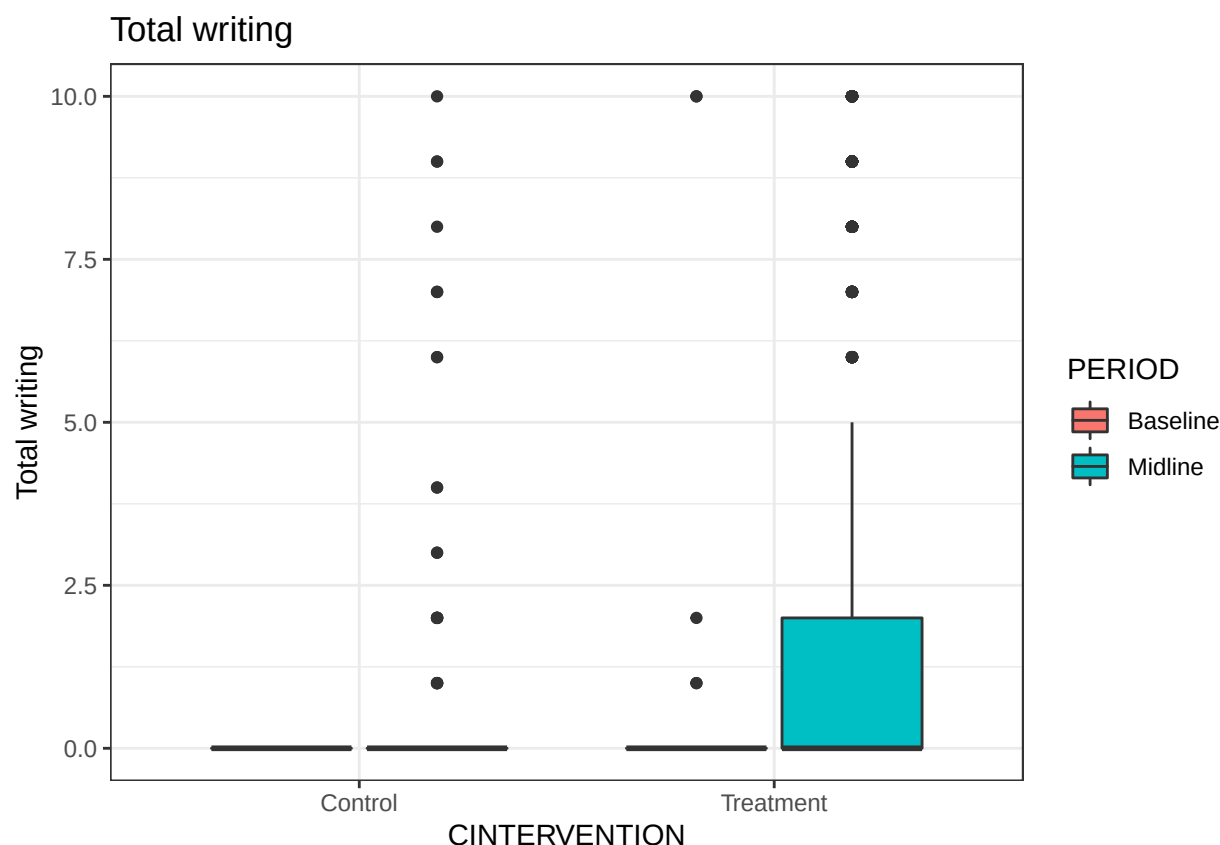
As shown in the table above, for the the Total writing EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at baseline was 0.01337154 (SD = 0.2994782) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.04711538 (SD = 0.6787659). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.03374385 points. The p-value for this difference was 0.136983. The mean for the Control (Comparison (Portuguese)) condition at midline was 3.052469 (SD = 4.026217) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 2.224451 (SD = 3.641204). The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.8280183 points. The p-value for this difference was 0.009876457. The change from the baseline to the midline of 3.039098 points for the Control (Comparison (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 2.177335 points. Consequently, the change for the Control (Comparison (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of -0.8617622 points. The p-value for this difference was 0.006910867. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Portuguese)) when compared to the Control (Comparison (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Portuguese)) (or some other unobserved process) impacted on the Total writing EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed. It should be pointed out that the difference in the rate of change of the Treatment (FFE + lit (Portuguese)) relative to the Control (Comparison (Portuguese)) was negative and thus suggests that students in the Control (Comparison (Portuguese)) schools performed significantly better than those in the Treatment (FFE + lit (Portuguese)) .

1.13.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 63: Total writing

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.000	0.000	89	0	0	0.917	2.220	109	0	10
Treatment	0.045	0.618	538	0	10	2.096	3.671	761	0	10



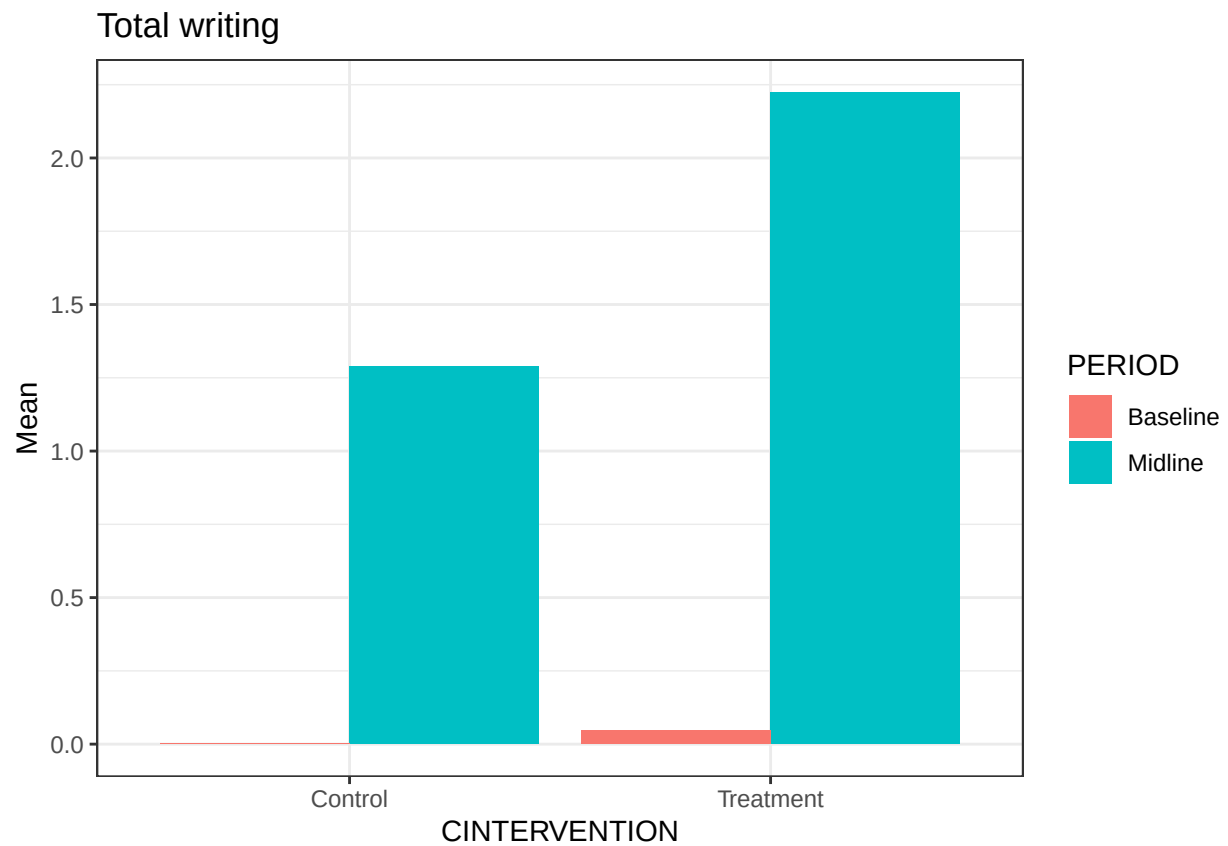


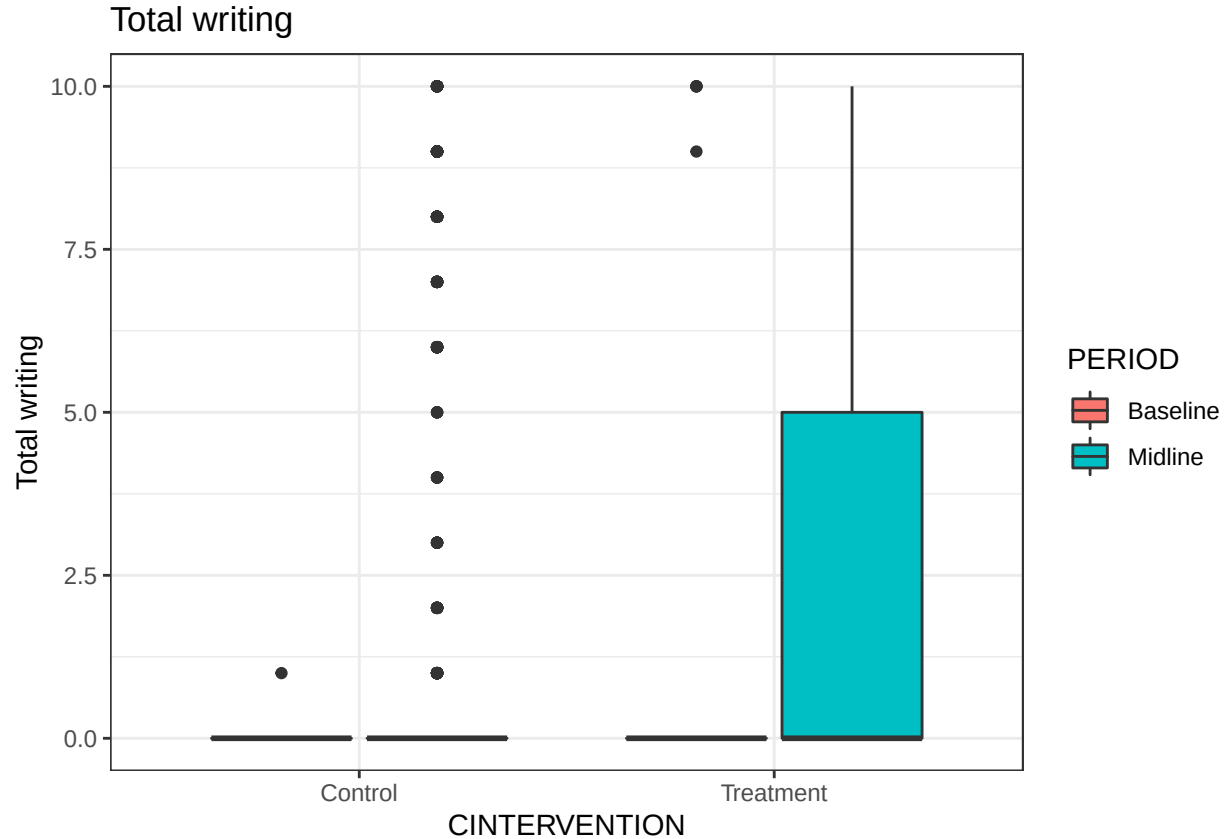
As shown in the table above, for the the Total writing EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at baseline was 0 (SD = 0) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.04460967 (SD = 0.6177531). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.04460967 points. The p-value for this difference was 0.1183386. The mean for the Control (Comparison (Bilingual)) condition at midline was 0.9174312 (SD = 2.219978) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 2.095926 (SD = 3.670651). The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 1.178495 points. The p-value for this difference was 0.008536447. The change from the baseline to the midline of 0.9174312 points for the Control (Comparison (Bilingual)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 2.051317 points. Consequently, the change for the Control (Comparison (Bilingual)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 1.133886 points. The p-value for this difference was 0.01217518. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Total writing EGRA subtask across the Control (Comparison (Bilingual)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

1.13.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 64: Total writing

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.001	0.032	1007	0	1	1.289	2.920	952	0	10
Treatment	0.047	0.679	1040	0	10	2.224	3.641	1047	0	10



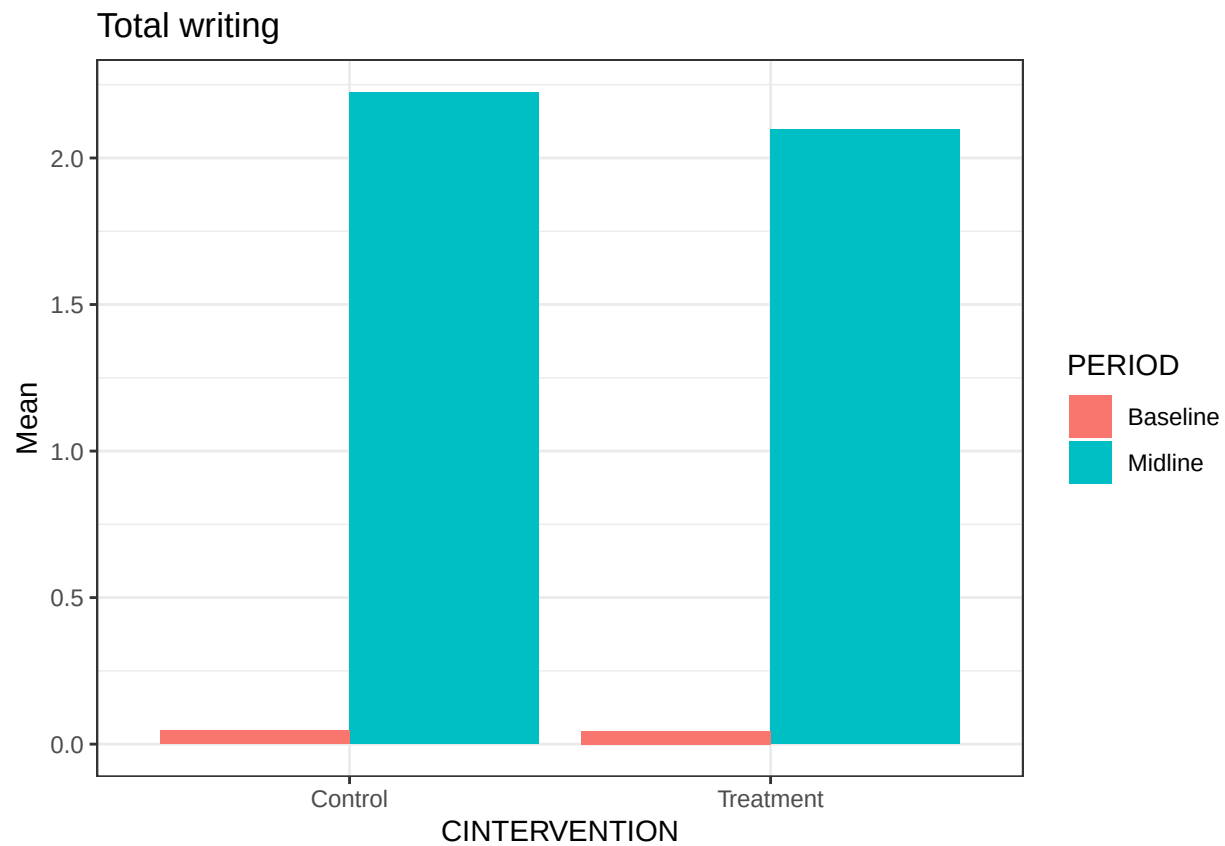


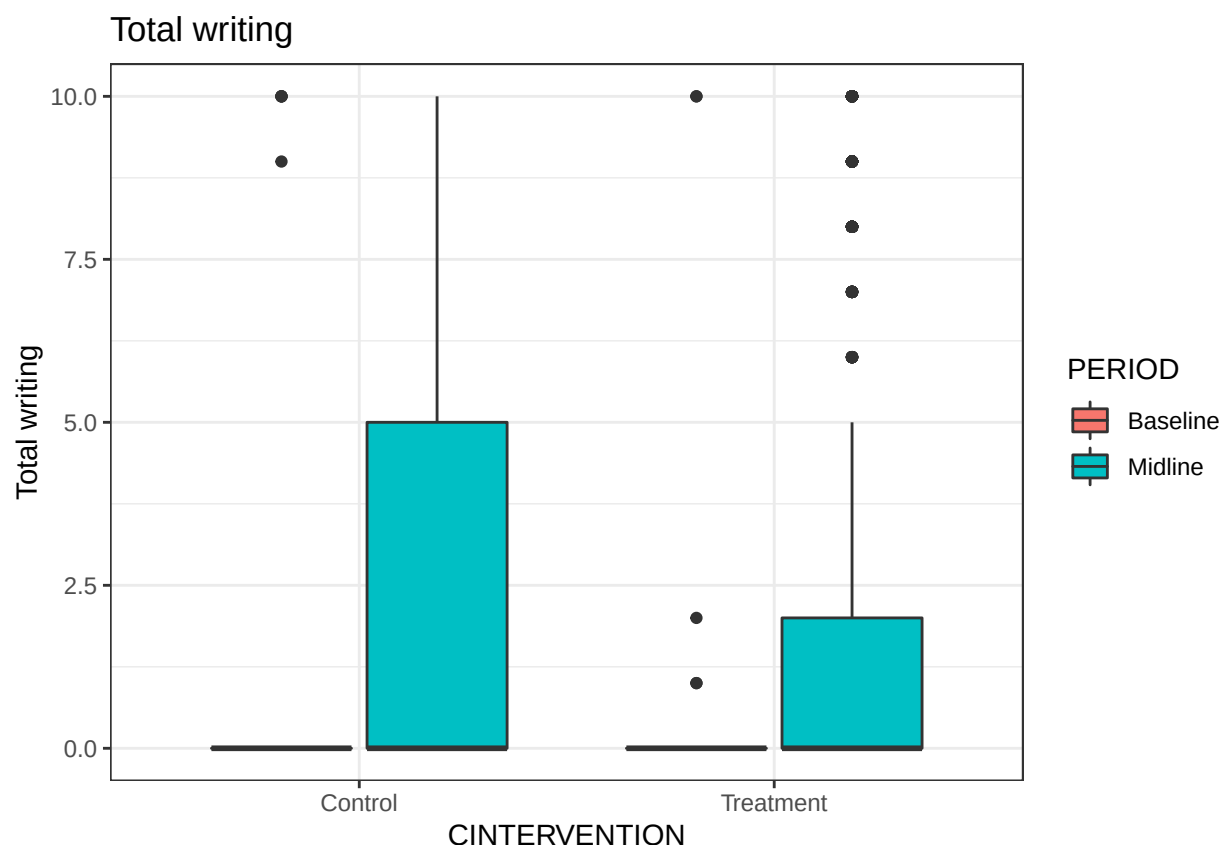
As shown in the table above, for the the Total writing EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at baseline was 0.0009930487 (SD = 0.03151267) and the mean for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.04711538 (SD = 0.6787659). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.04612234 points. The p-value for this difference was 0.02613085. The mean for the Control (FFE only (Portuguese)) condition at midline was 1.288866 (SD = 2.919812) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 2.224451 (SD = 3.641204). The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.9355853 points. The p-value for this difference was 0.0009320036. The change from the baseline to the midline of 1.287872 points for the Control (FFE only (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 2.177335 points. Consequently, the change for the Control (FFE only (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 0.8894629 points. The p-value for this difference was 0.001531693. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Portuguese)) when compared to the Control (FFE only (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Portuguese)) (or some other unobserved process) impacted on the Total writing EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

1.13.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 65: Total writing

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	0.047	0.679	1040	0	10	2.224	3.641	1047	0	10
Treatment	0.045	0.618	538	0	10	2.096	3.671	761	0	10





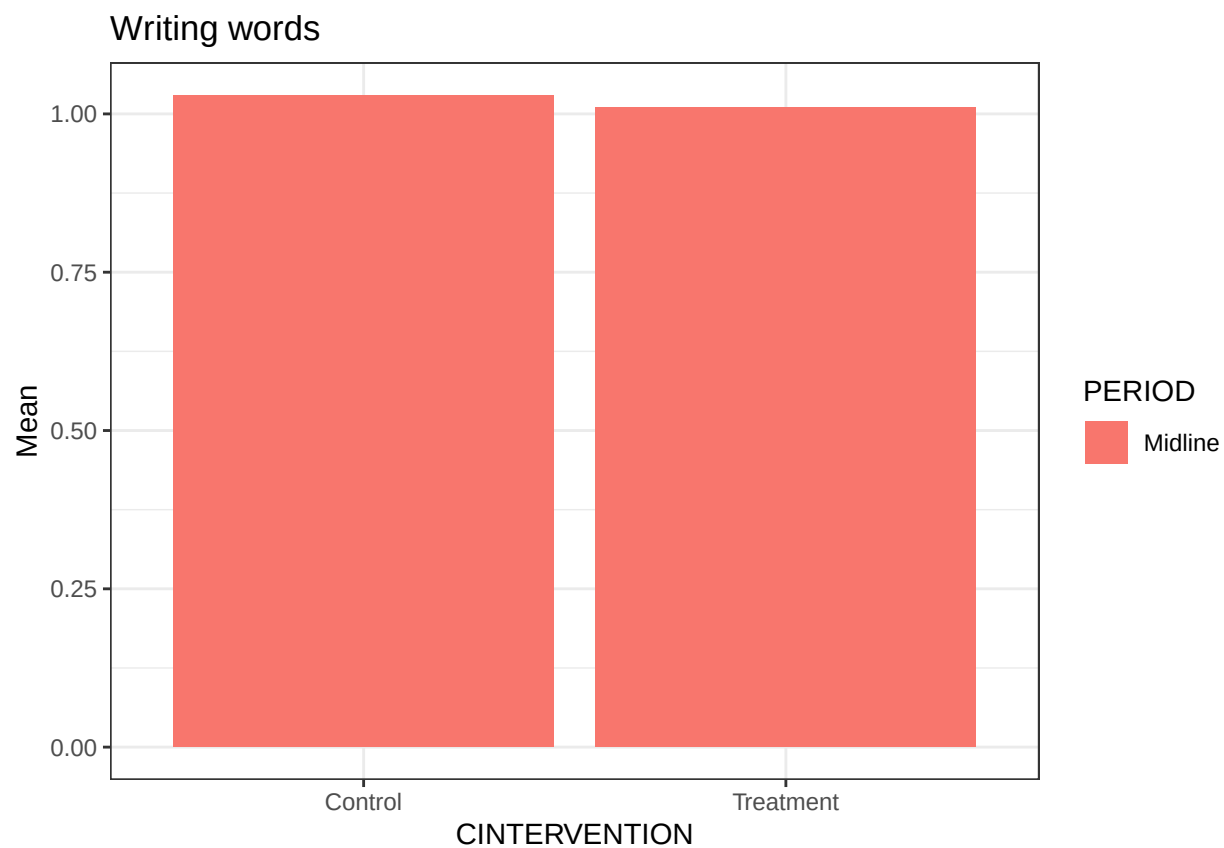
As shown in the table above, for the the Total writing EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at baseline was 0.04711538 (SD = 0.6787659) and the mean for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.04460967 (SD = 0.6177531). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.002505719 points. The p-value for this difference was 0.943553. The mean for the Control (FFE + lit (Portuguese)) condition at midline was 2.224451 (SD = 3.641204) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 2.095926 (SD = 3.670651). The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.1285244 points. The p-value for this difference was 0.6731649. The change from the baseline to the midline of 2.177335 points for the Control (FFE + lit (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 2.051317 points. Consequently, the change for the Control (FFE + lit (Portuguese)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of -0.1260187 points. The p-value for this difference was 0.6817533. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Total writing EGRA subtask across the Control (FFE + lit (Portuguese)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

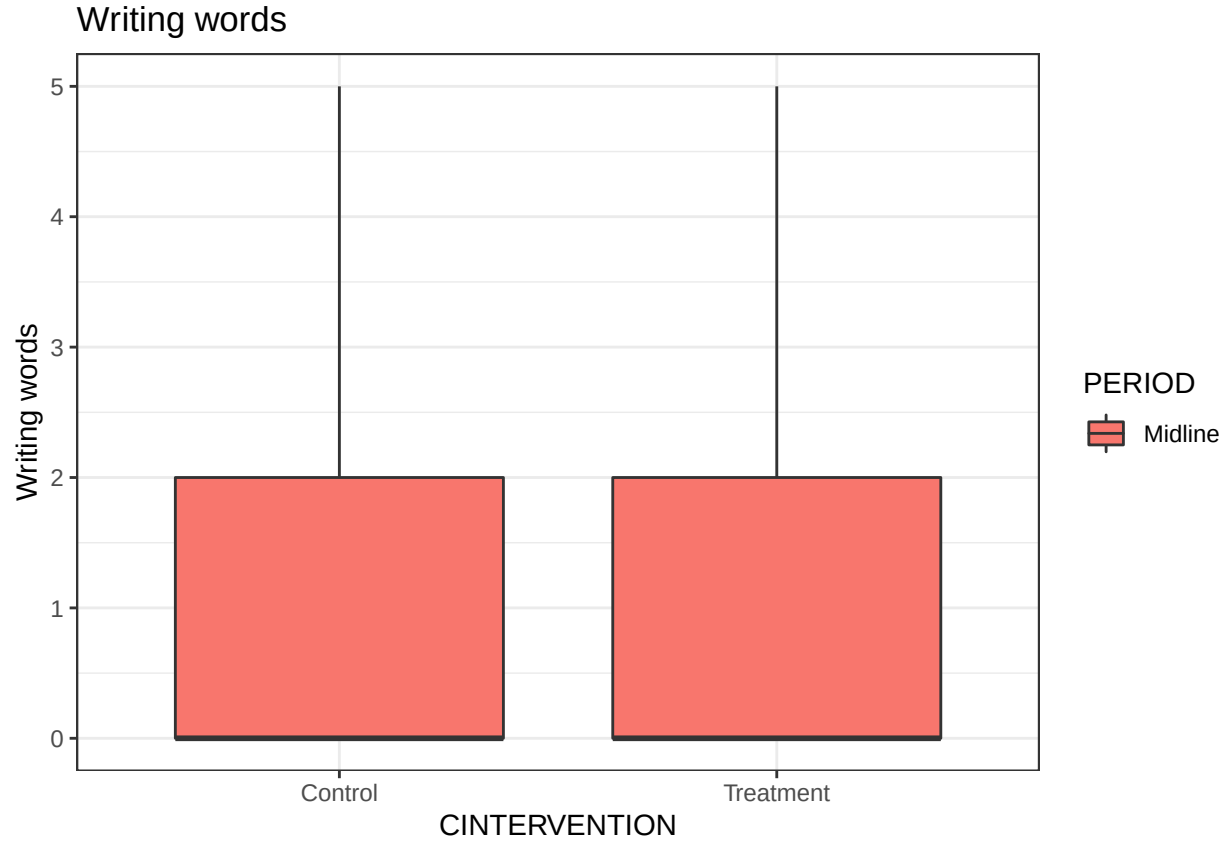
1.14 EGRA_ST9_2B: Writing words

1.14.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 66: Writing words

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	1.03	1.776	1081	0	5
Treatment	NA	NA	0	NA	NA	1.01	1.759	1808	0	5



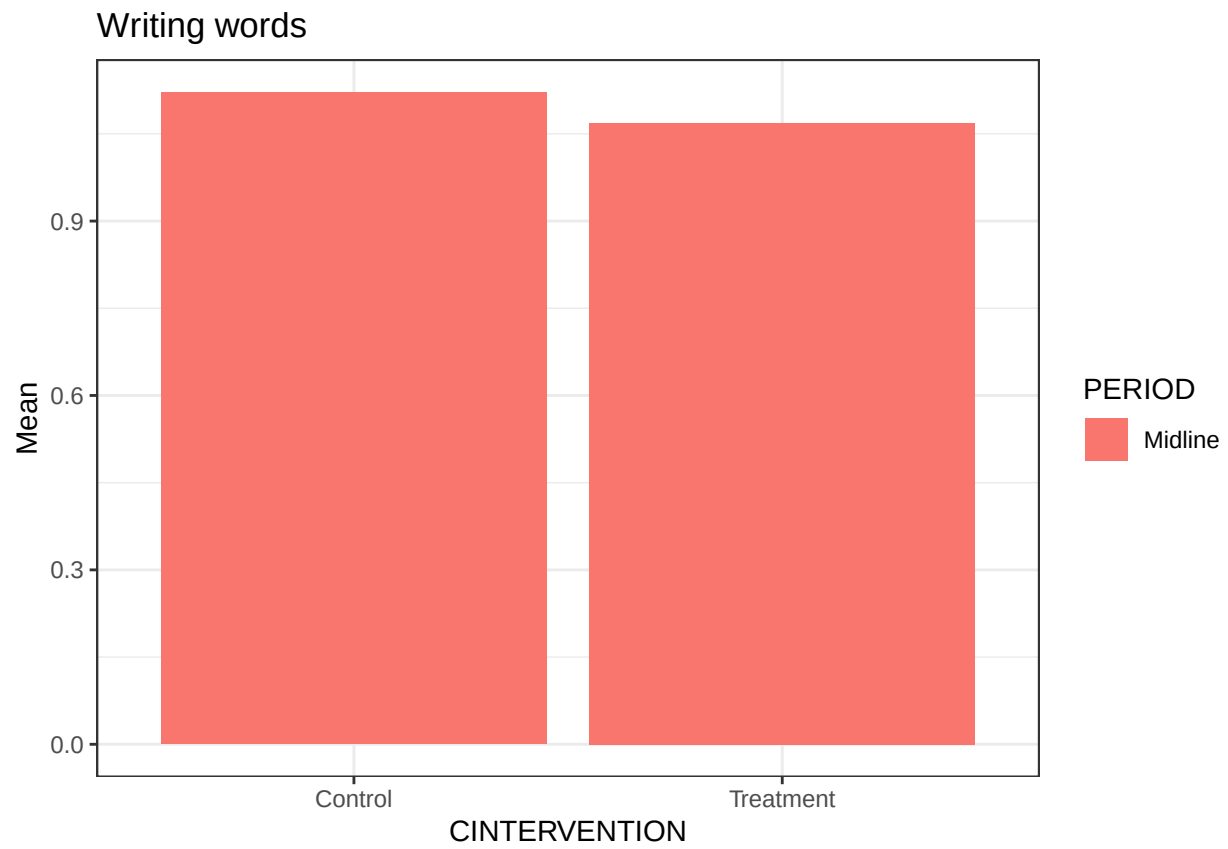


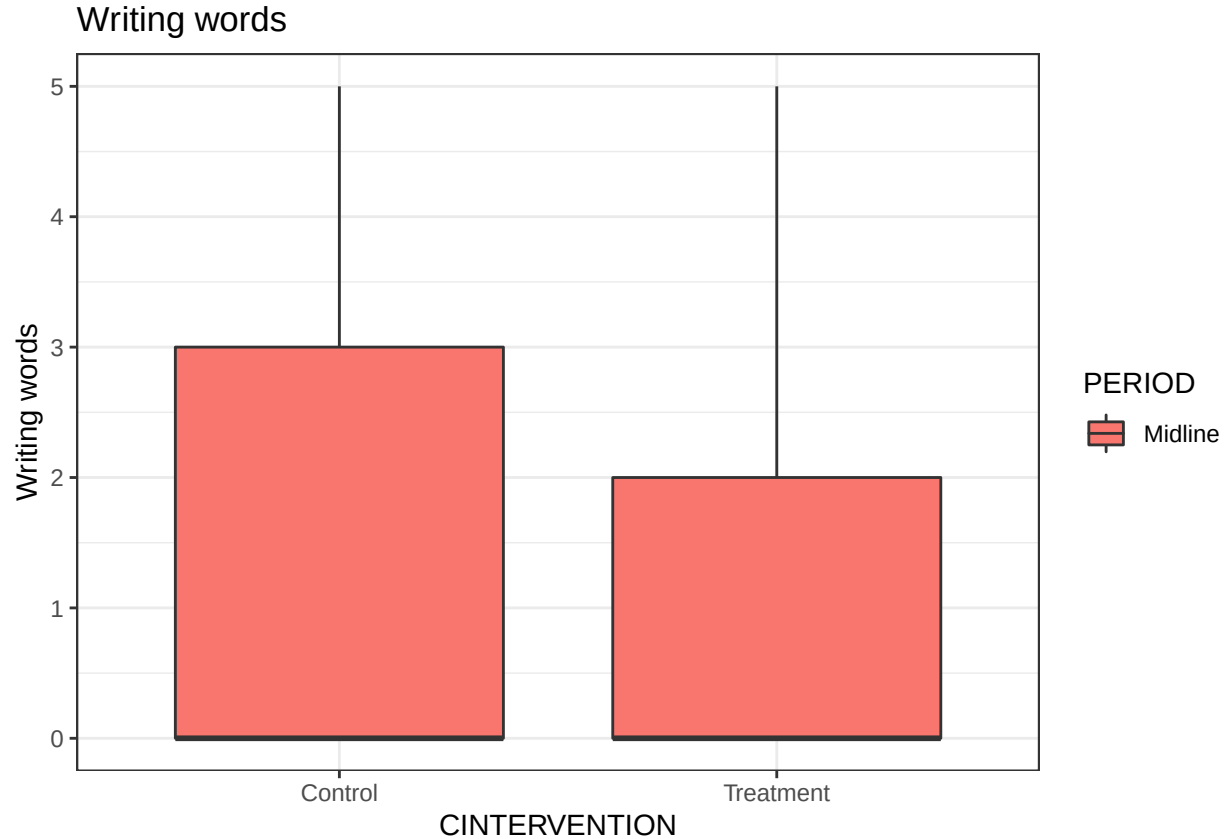
As shown in the table above, for the Writing words EGRA subtask, the mean for the Control (Comparison (all)) condition at midline was 1.029602 (SD = 1.77562) and the mean for the Treatment (FFE + lit (all)) condition at midline was 1.009956 (SD = 1.759443). The difference for the Writing words across the Control (Comparison (all)) and Treatment (FFE + lit (all)) conditions at midline was thus -0.01964647 points (there was no baseline measurement for this variable). The p-value for this difference was 0.878366. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the levels of Writing words between the Control (Comparison (all)) and Treatment (FFE + lit (all)) conditions at midline beyond that which could result from sampling error.

1.14.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 67: Writing words

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	1.121	1.829	972	0	5
Treatment	NA	NA	0	NA	NA	1.069	1.776	1047	0	5



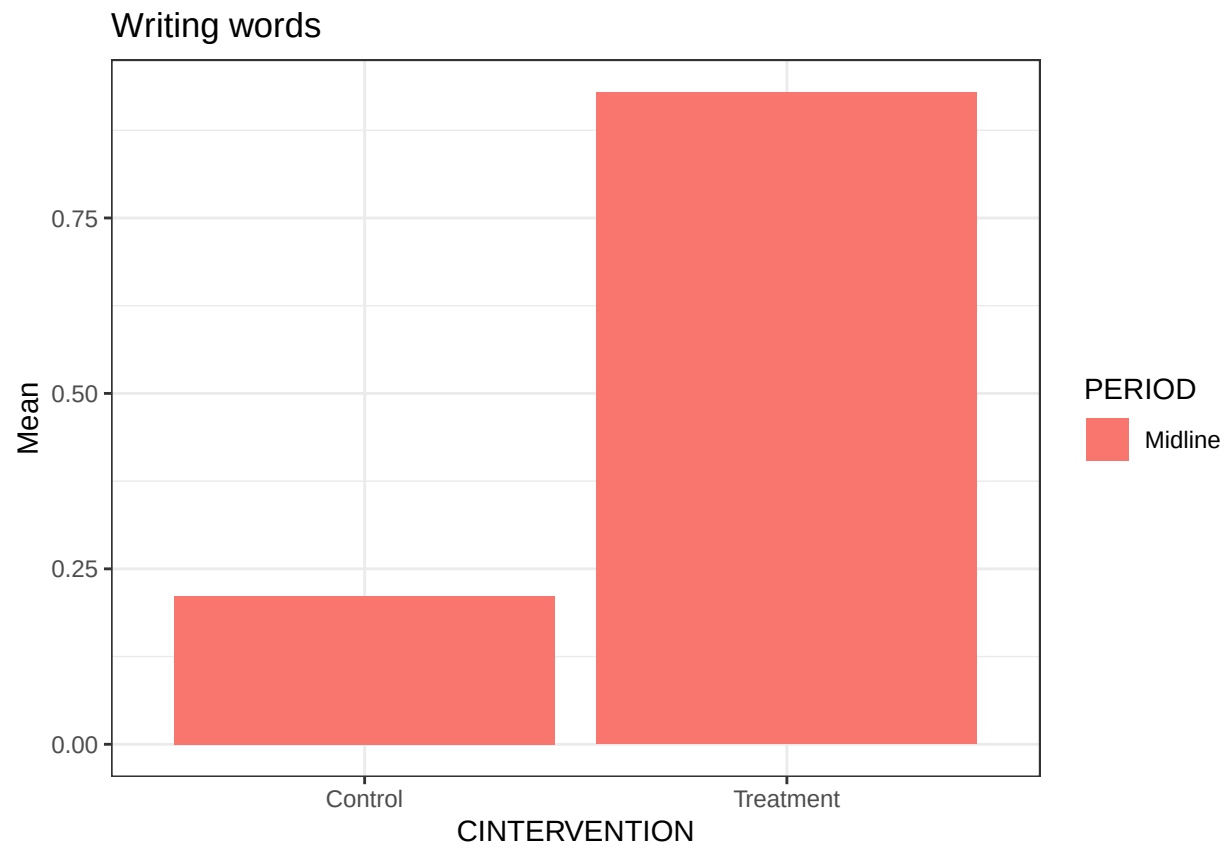


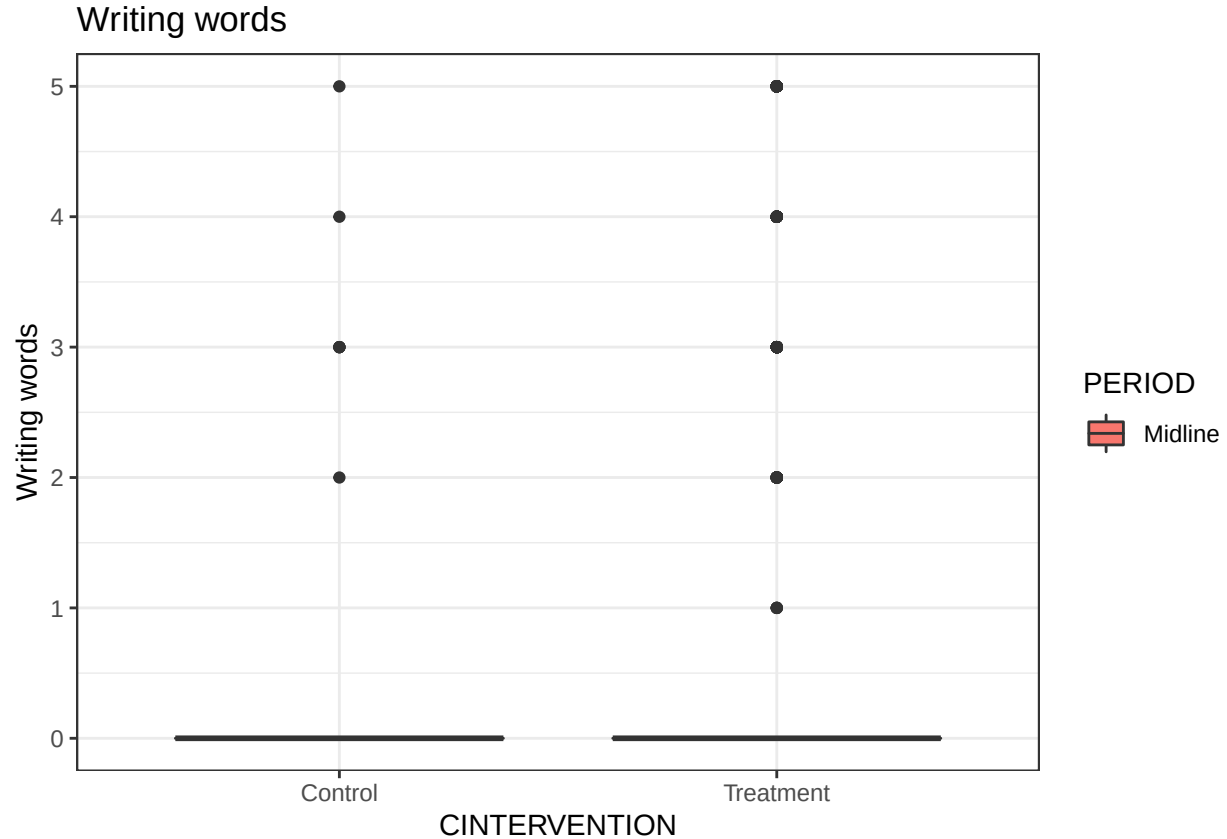
As shown in the table above, for the Writing words EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at midline was 1.121399 (SD = 1.828844) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 1.068768 (SD = 1.775941). The difference for the Writing words across the Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline was thus -0.05263127 points (there was no baseline measurement for this variable). The p-value for this difference was 0.7195677. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the levels of Writing words between the Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline beyond that which could result from sampling error.

1.14.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 68: Writing words

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	0.211	0.840	109	0	5
Treatment	NA	NA	0	NA	NA	0.929	1.734	761	0	5



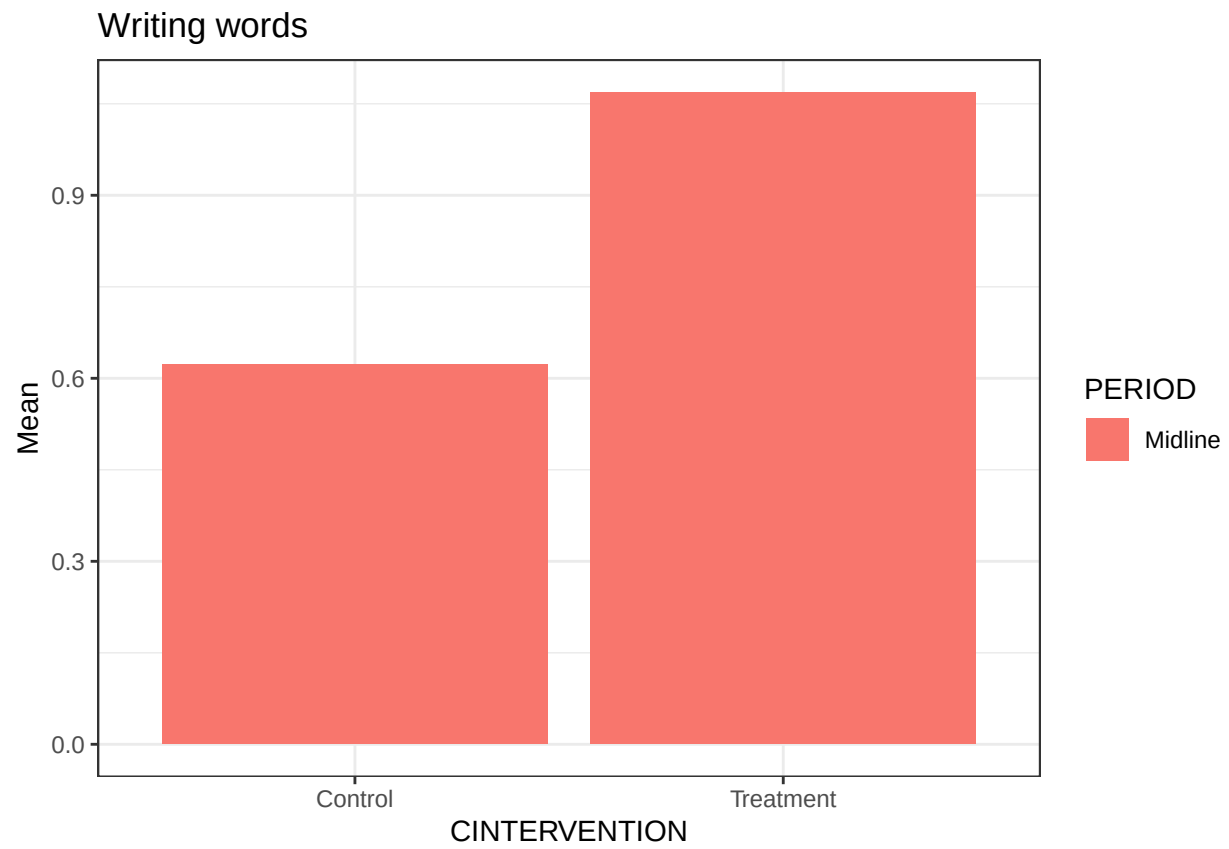


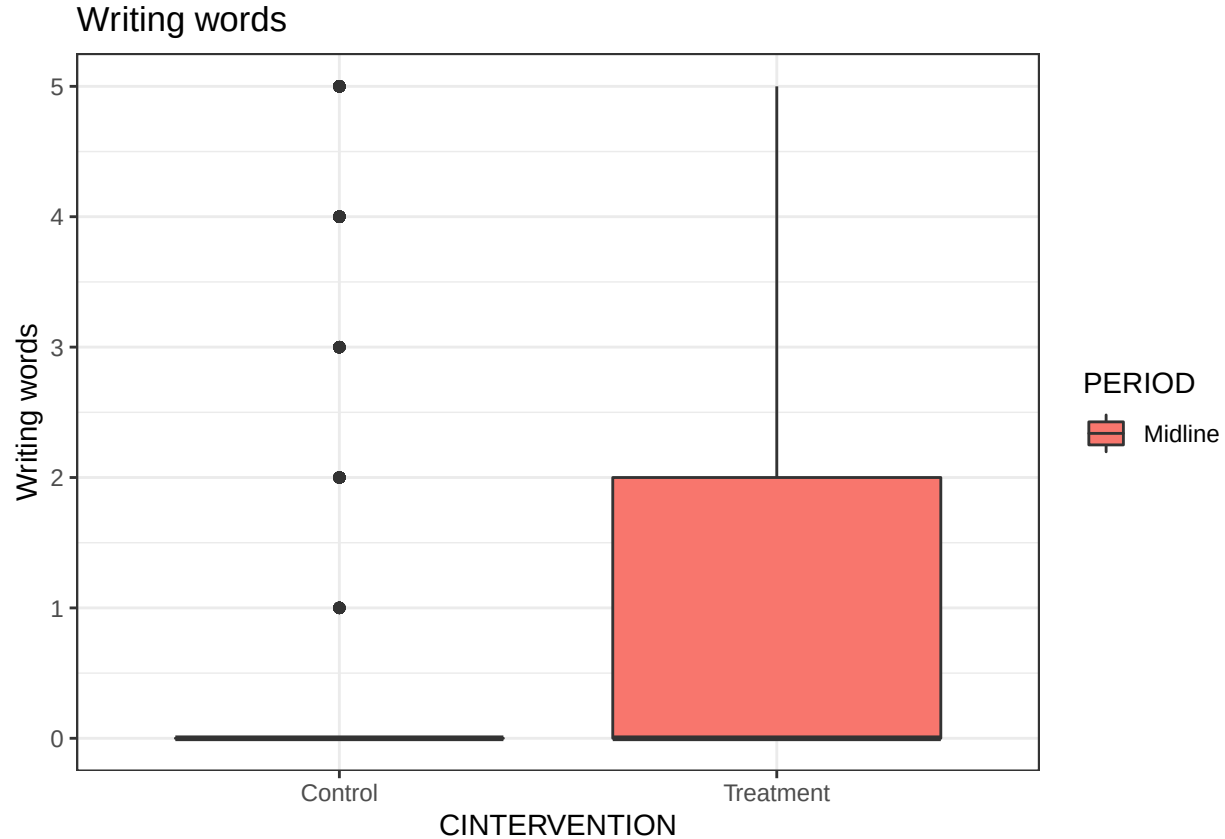
As shown in the table above, for the Writing words EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at midline was 0.2110092 (SD = 0.8396802) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 0.9290407 (SD = 1.734392). The difference for the Writing words across the Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) conditions at midline was thus 0.7180316 points (there was no baseline measurement for this variable). The p-value for this difference was 1.588162e-06. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts). Nevertheless, due to the absence of baseline values and lack of random allocation to the treatment and control conditions, a causal attribution of this effect to the treatment is not possible

1.14.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 69: Writing words

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	0.623	1.424	952	0	5
Treatment	NA	NA	0	NA	NA	1.069	1.776	1047	0	5



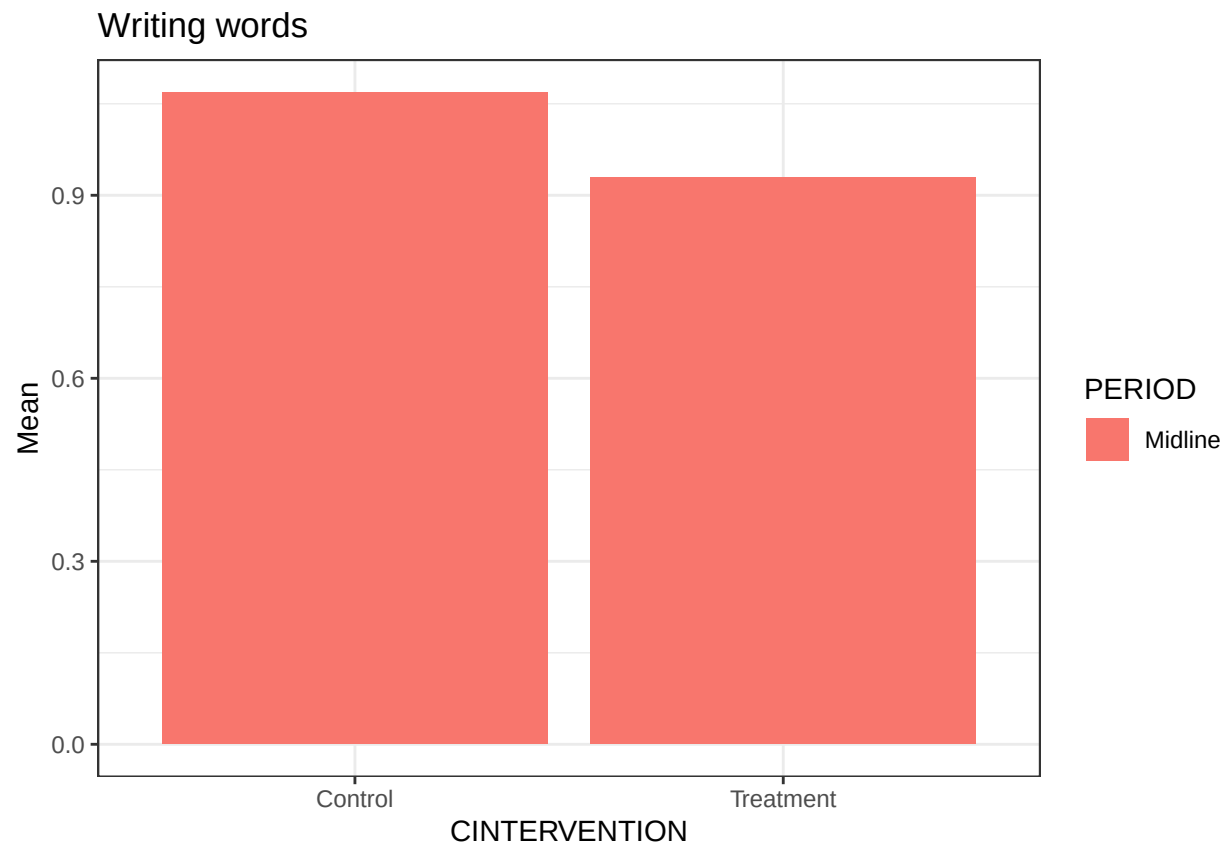


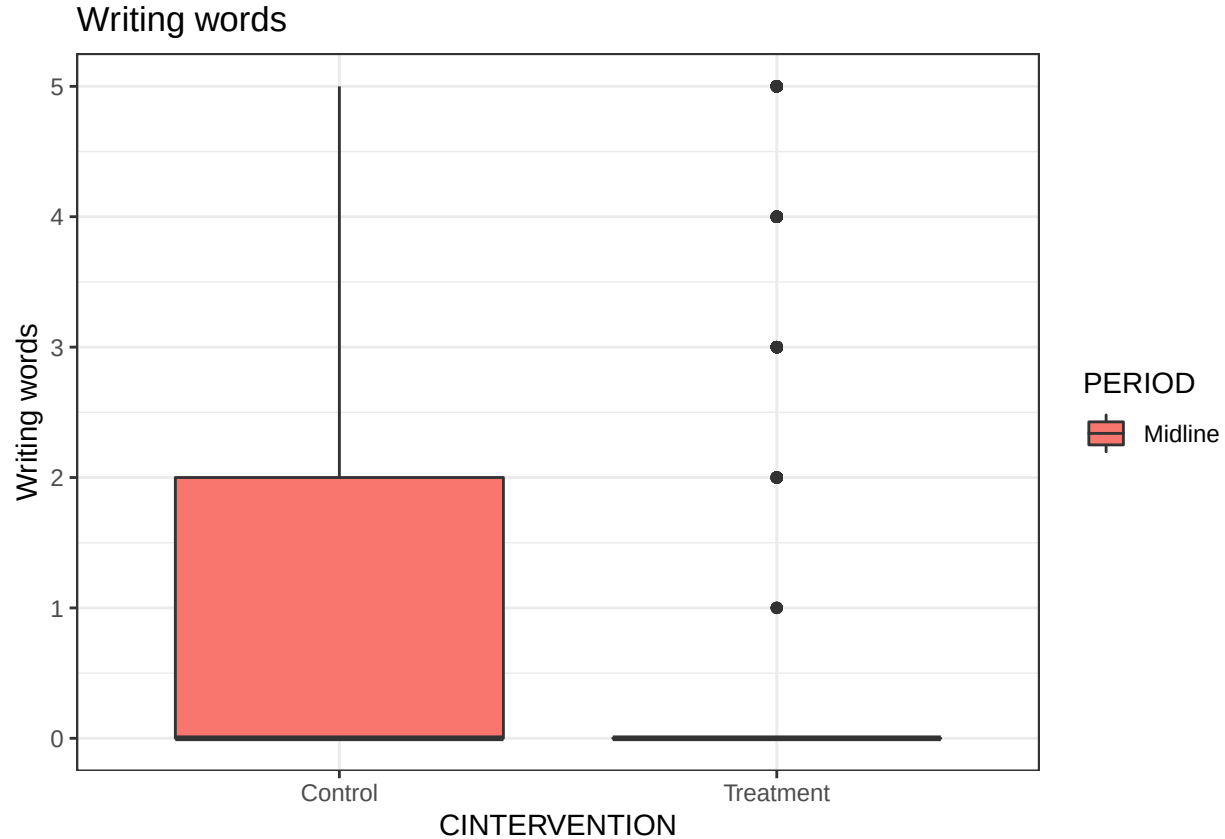
As shown in the table above, for the Writing words EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at midline was 0.6228992 (SD = 1.424445) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 1.068768 (SD = 1.775941). The difference for the Writing words across the Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline was thus 0.4458687 points (there was no baseline measurement for this variable). The p-value for this difference was 0.001099579. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts). Nevertheless, due to the absence of baseline values and lack of random allocation to the treatment and control conditions, a causal attribution of this effect to the treatment is not possible

1.14.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 70: Writing words

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	1.069	1.776	1047	0	5
Treatment	NA	NA	0	NA	NA	0.929	1.734	761	0	5





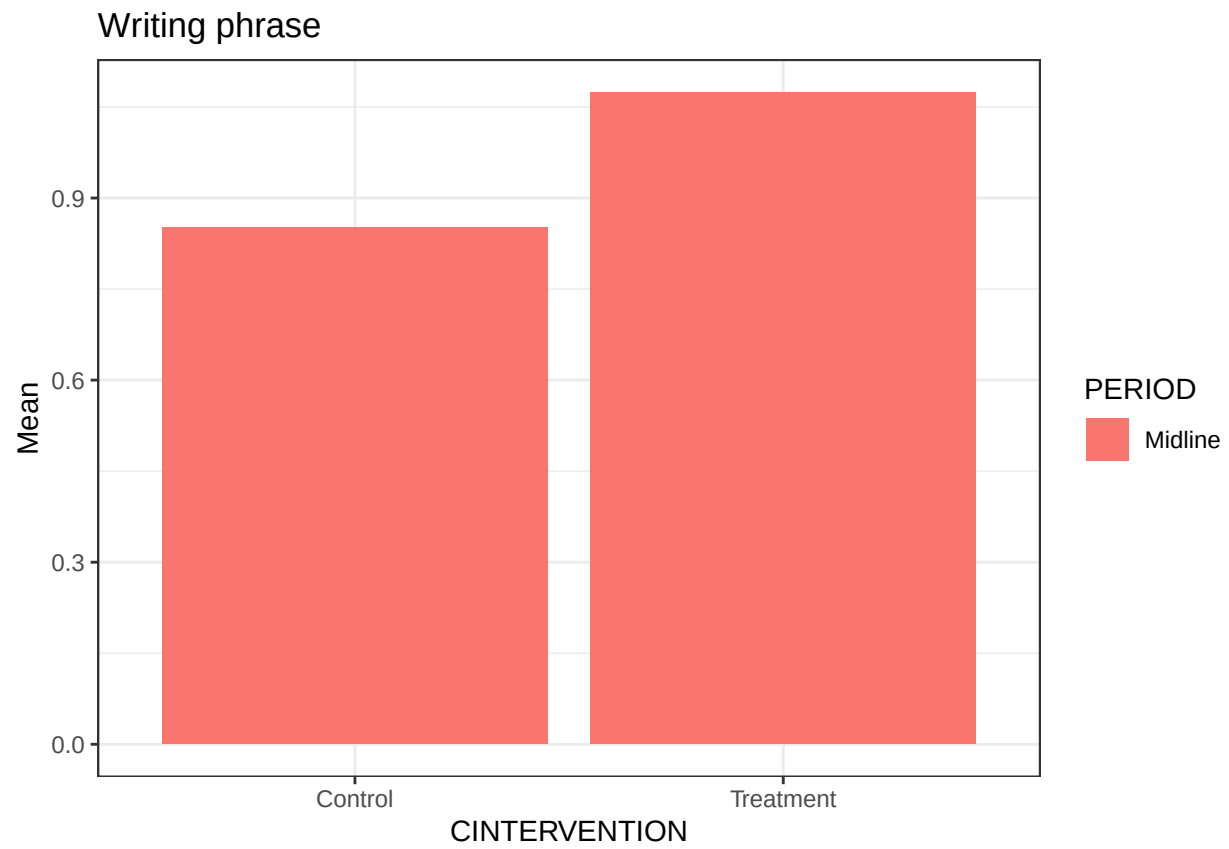
As shown in the table above, for the Writing words EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at midline was 1.068768 (SD = 1.775941) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 0.9290407 (SD = 1.734392). The difference for the Writing words across the Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) conditions at midline was thus -0.1397272 points (there was no baseline measurement for this variable). The p-value for this difference was 0.3384802. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the levels of Writing words between the Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) conditions at midline beyond that which could result from sampling error.

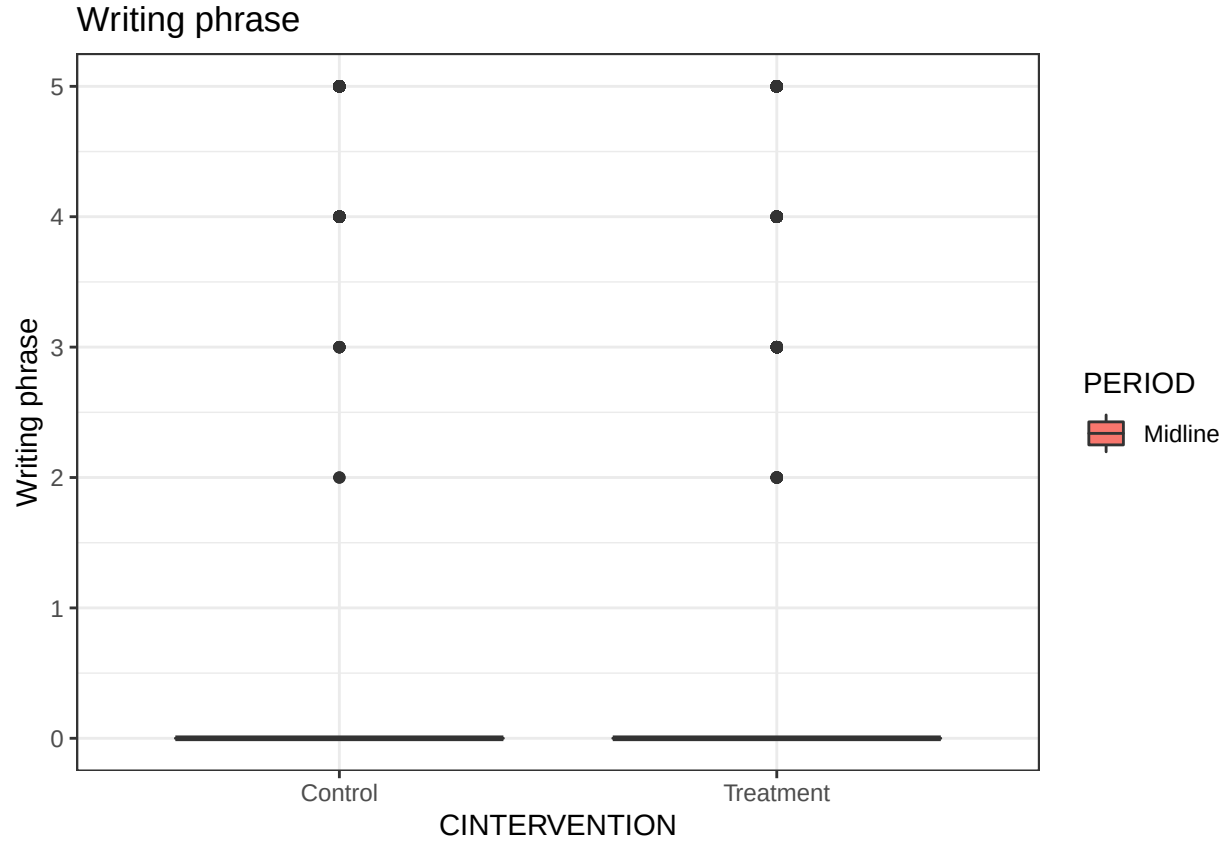
1.15 EGRA_ST9_2C: Writing phrase

1.15.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 71: Writing phrase

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	0.851	1.806	1081	0	5
Treatment	NA	NA	0	NA	NA	1.074	1.941	1808	0	5



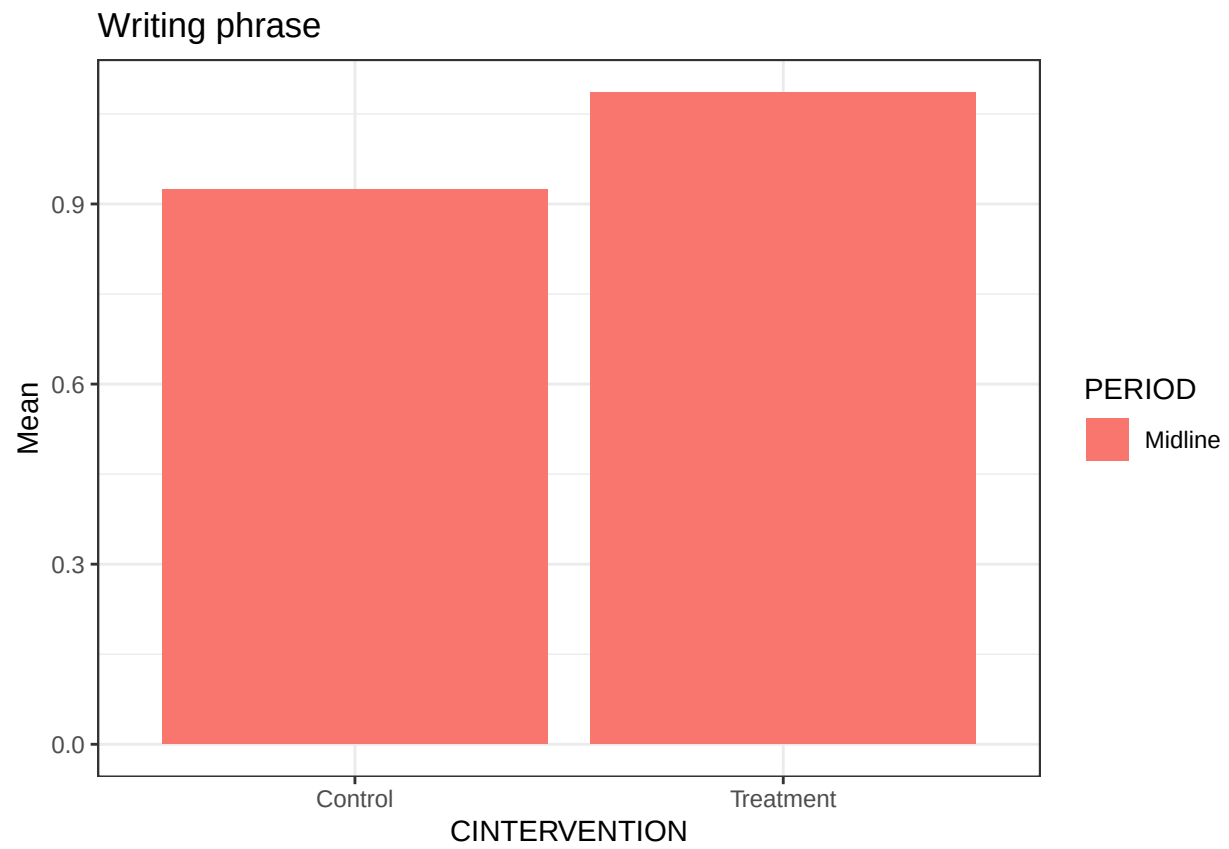


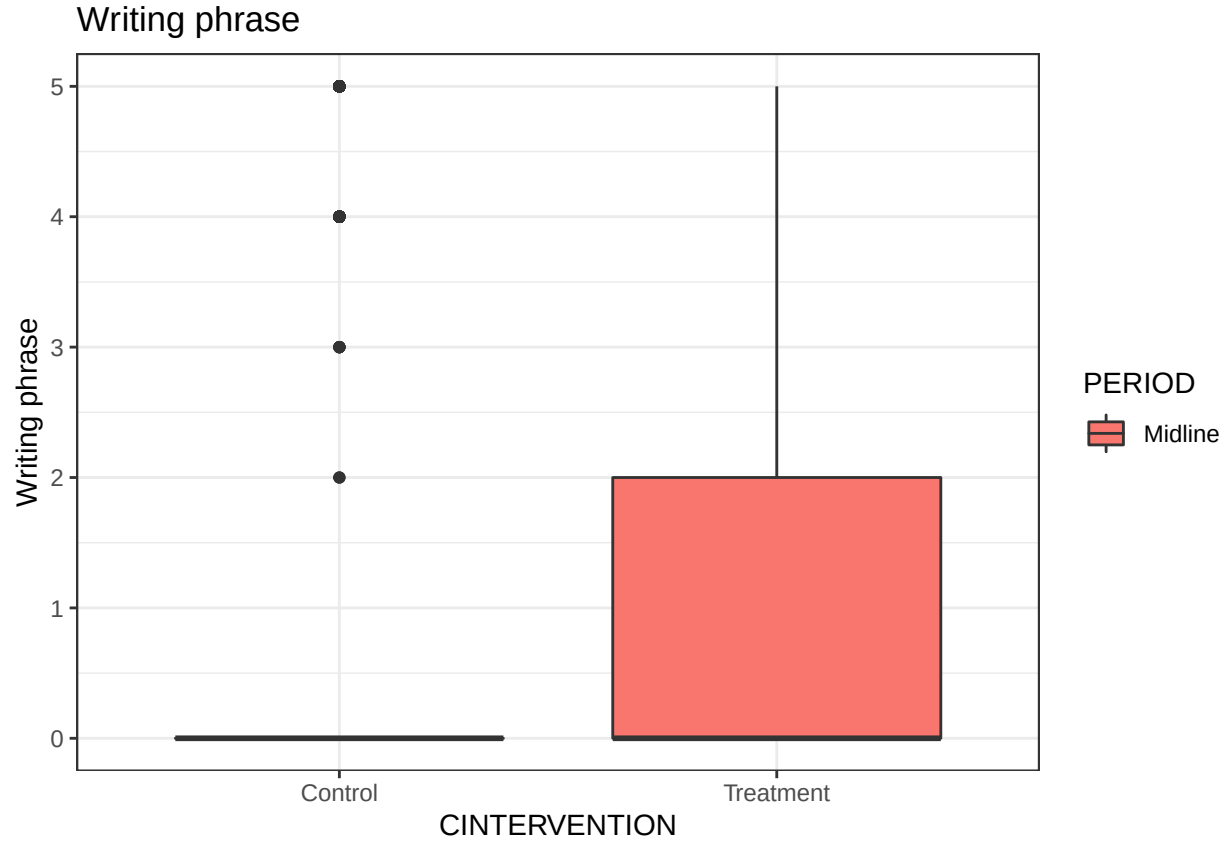
As shown in the table above, for the Writing phrase EGRA subtask, the mean for the Control (Comparison (all)) condition at midline was 0.8510638 (SD = 1.805604) and the mean for the Treatment (FFE + lit (all)) condition at midline was 1.074115 (SD = 1.940748). The difference for the Writing phrase across the Control (Comparison (all)) and Treatment (FFE + lit (all)) conditions at midline was thus 0.2230512 points (there was no baseline measurement for this variable). The p-value for this difference was 0.06842061. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the levels of Writing phrase between the Control (Comparison (all)) and Treatment (FFE + lit (all)) conditions at midline beyond that which could result from sampling error.

1.15.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 72: Writing phrase

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	0.924	1.864	972	0	5
Treatment	NA	NA	0	NA	NA	1.086	1.923	1047	0	5



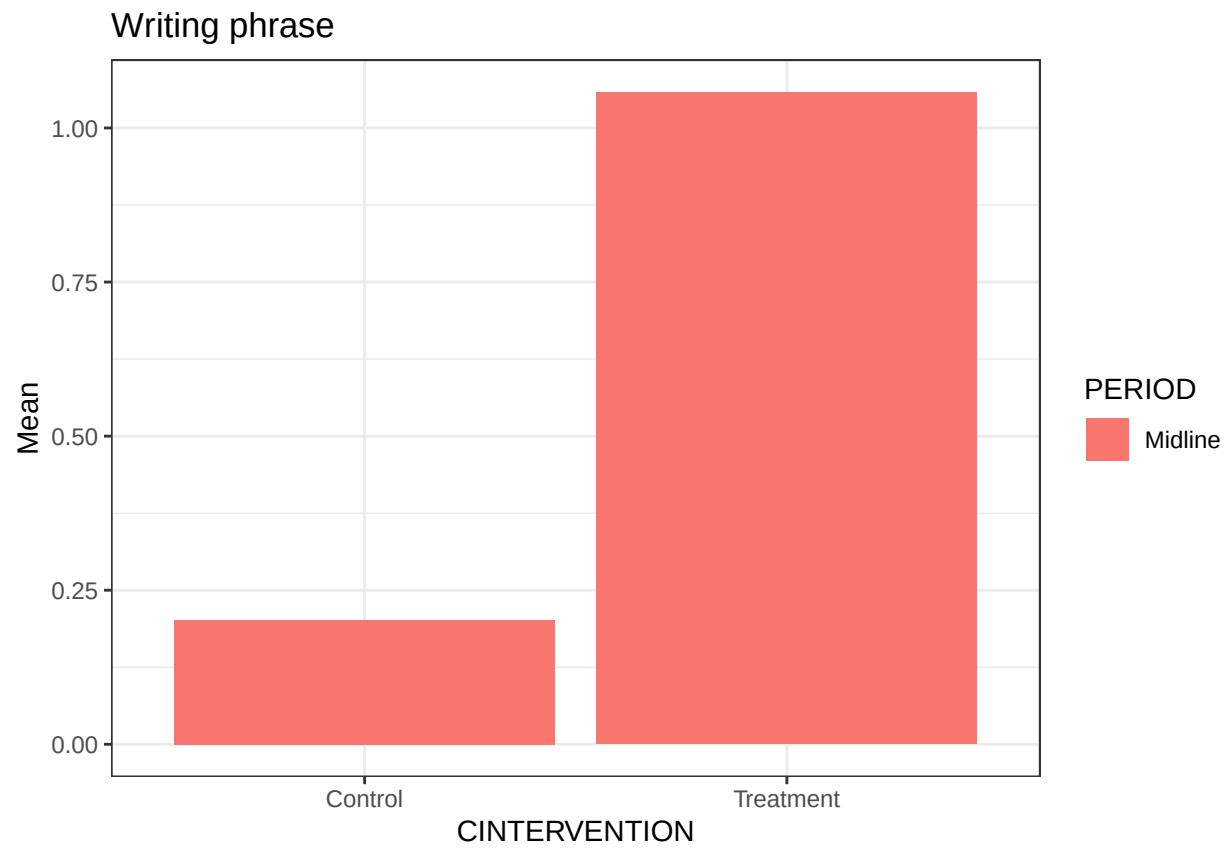


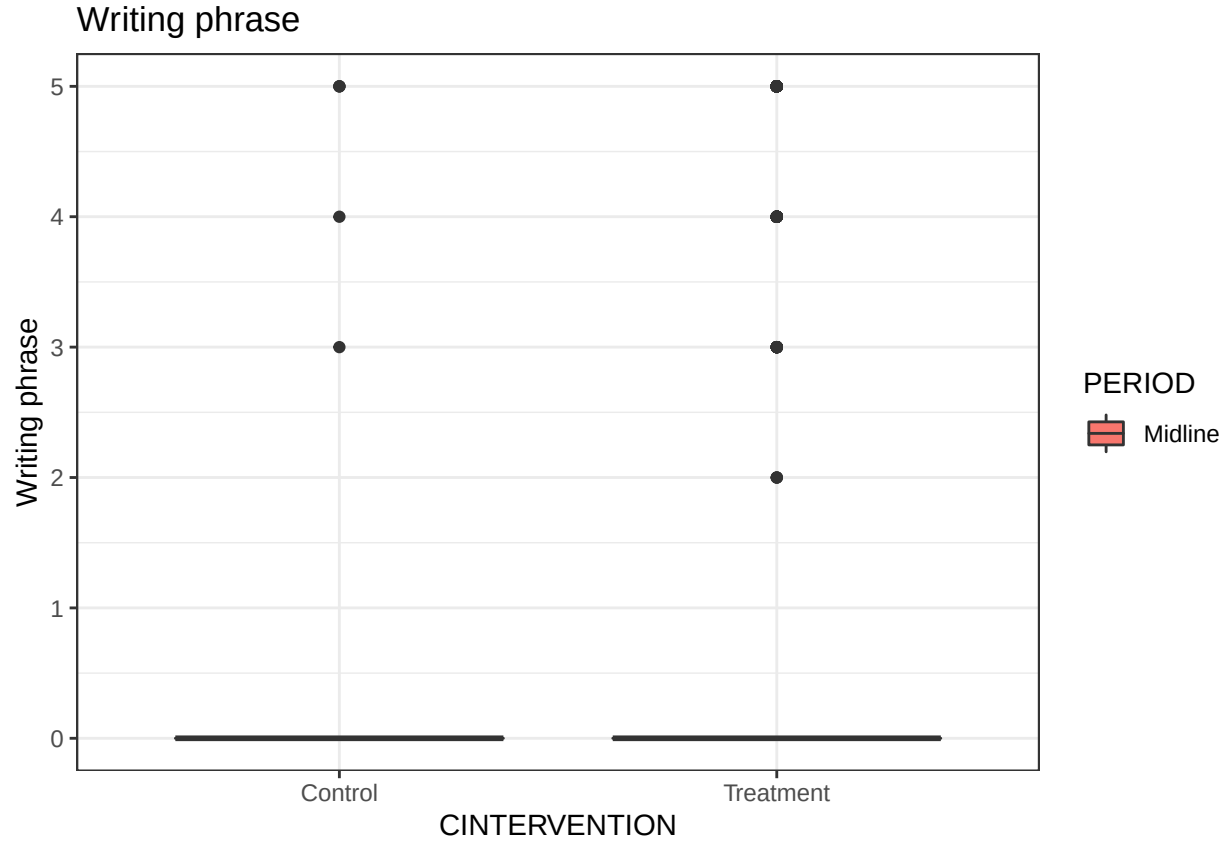
As shown in the table above, for the Writing phrase EGRA subtask, the mean for the Control (Comparison (Portuguese)) condition at midline was 0.9238683 (SD = 1.864174) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 1.08596 (SD = 1.923057). The difference for the Writing phrase across the Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline was thus 0.1620916 points (there was no baseline measurement for this variable). The p-value for this difference was 0.2546944. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the levels of Writing phrase between the Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline beyond that which could result from sampling error.

1.15.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 73: Writing phrase

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	0.202	0.941	109	0	5
Treatment	NA	NA	0	NA	NA	1.058	1.966	761	0	5



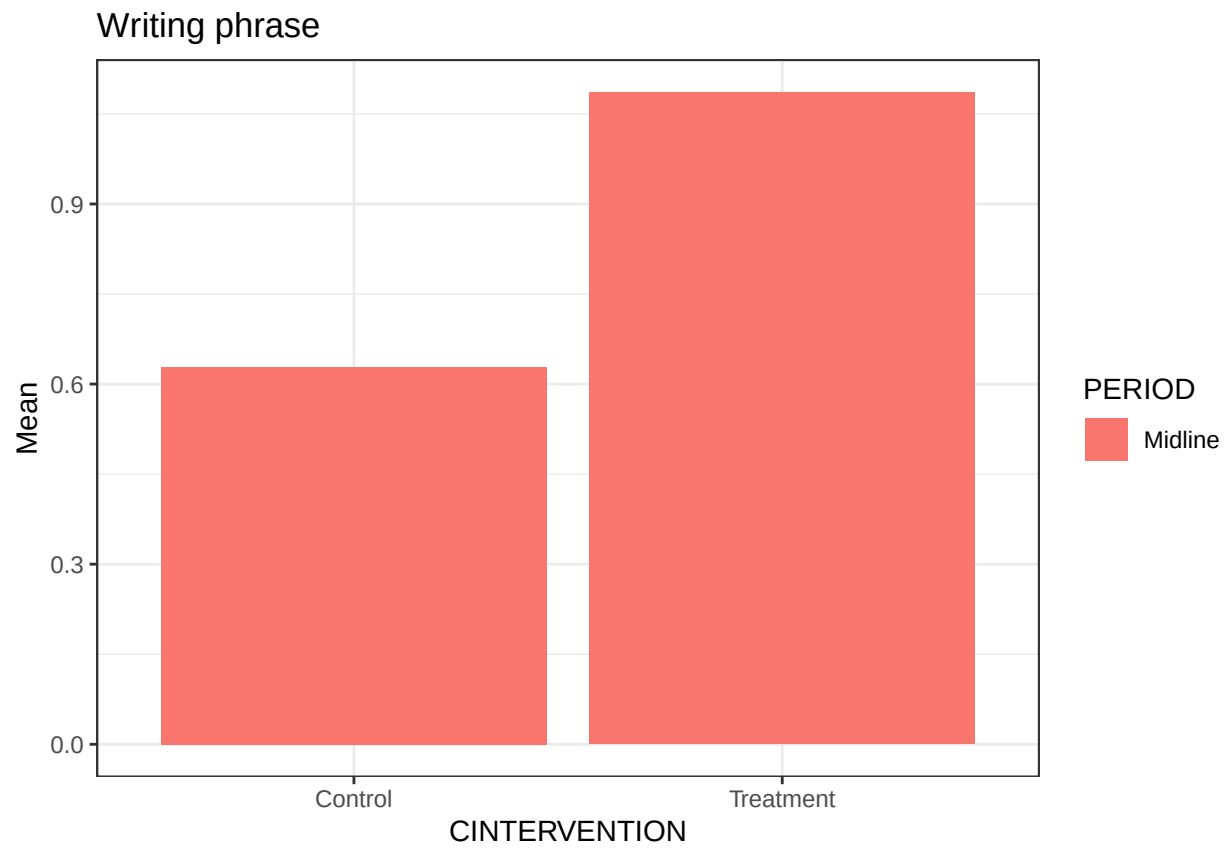


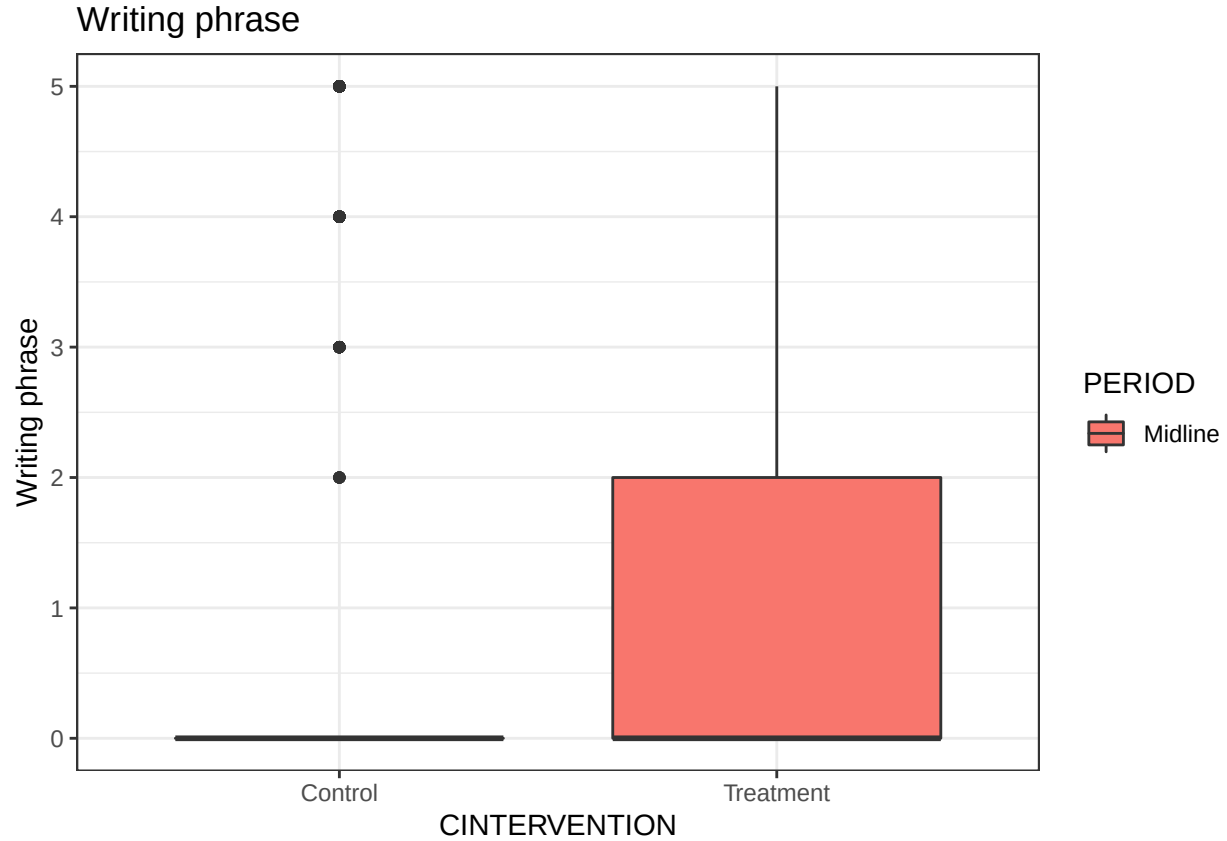
As shown in the table above, for the Writing phrase EGRA subtask, the mean for the Control (Comparison (Bilingual)) condition at midline was 0.2018349 (SD = 0.9406441) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 1.057819 (SD = 1.965979). The difference for the Writing phrase across the Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) conditions at midline was thus 0.8559838 points (there was no baseline measurement for this variable). The p-value for this difference was 1.358622e-05. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts). Nevertheless, due to the absence of baseline values and lack of random allocation to the treatment and control conditions, a causal attribution of this effect to the treatment is not possible

1.15.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 74: Writing phrase

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	0.628	1.533	952	0	5
Treatment	NA	NA	0	NA	NA	1.086	1.923	1047	0	5



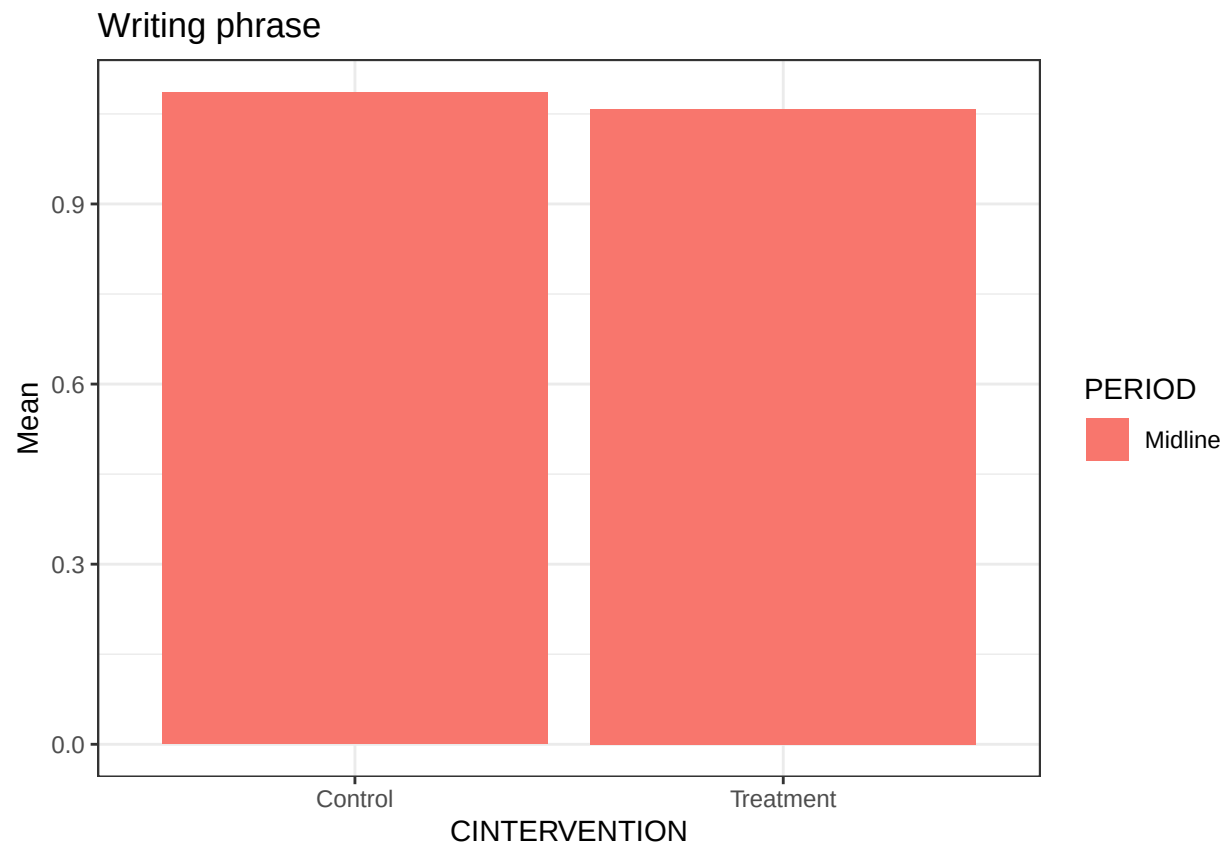


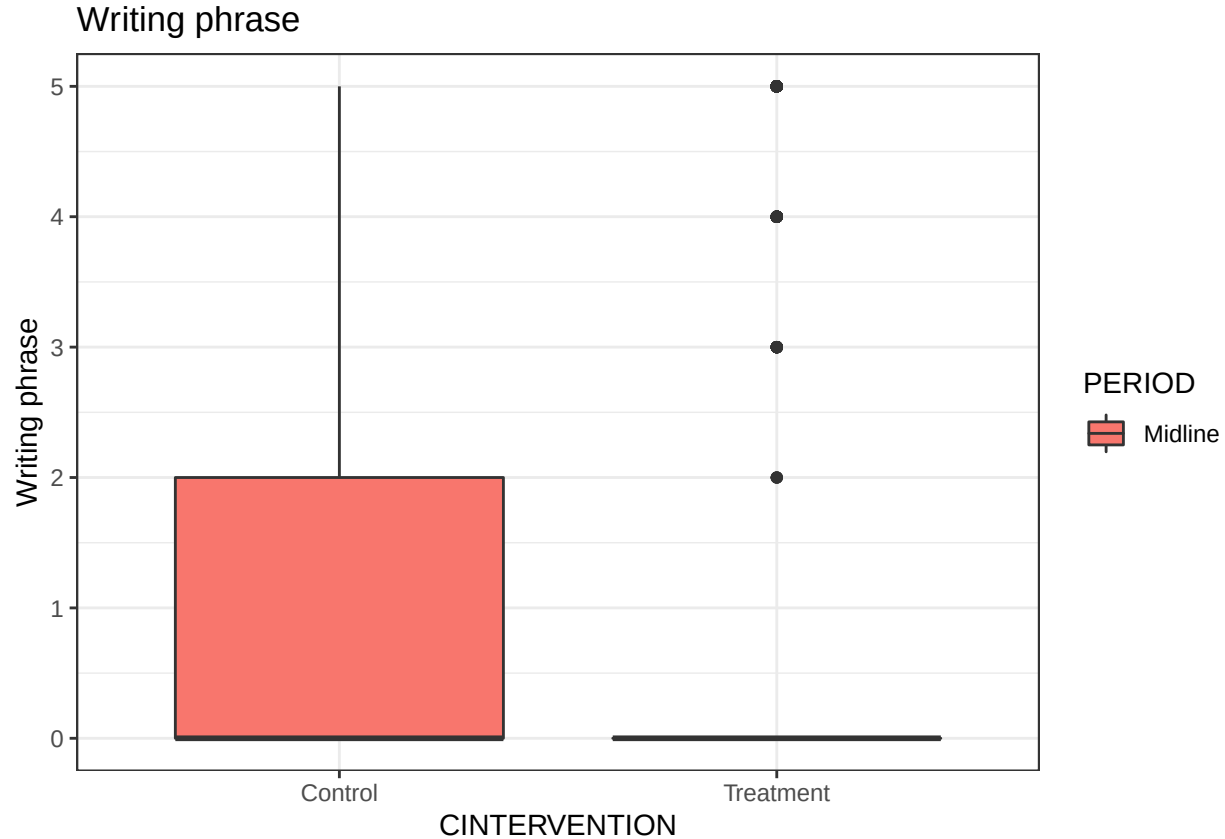
As shown in the table above, for the Writing phrase EGRA subtask, the mean for the Control (FFE only (Portuguese)) condition at midline was 0.6281513 (SD = 1.532805) and the mean for the Treatment (FFE + lit (Portuguese)) condition at midline was 1.08596 (SD = 1.923057). The difference for the Writing phrase across the Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline was thus 0.4578086 points (there was no baseline measurement for this variable). The p-value for this difference was 0.001876018. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts). Nevertheless, due to the absence of baseline values and lack of random allocation to the treatment and control conditions, a causal attribution of this effect to the treatment is not possible

1.15.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 75: Writing phrase

Intervention	Baseline					Midline				
	Mean	SD	n	min	max	Mean	SD	n	min	max
Control	NA	NA	0	NA	NA	1.086	1.923	1047	0	5
Treatment	NA	NA	0	NA	NA	1.058	1.966	761	0	5





As shown in the table above, for the Writing phrase EGRA subtask, the mean for the Control (FFE + lit (Portuguese)) condition at midline was 1.08596 (SD = 1.923057) and the mean for the Treatment (FFE + lit (Bilingual)) condition at midline was 1.057819 (SD = 1.965979). The difference for the Writing phrase across the Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) conditions at midline was thus -0.02814123 points (there was no baseline measurement for this variable). The p-value for this difference was 0.8598707. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the levels of Writing phrase between the Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) conditions at midline beyond that which could result from sampling error.

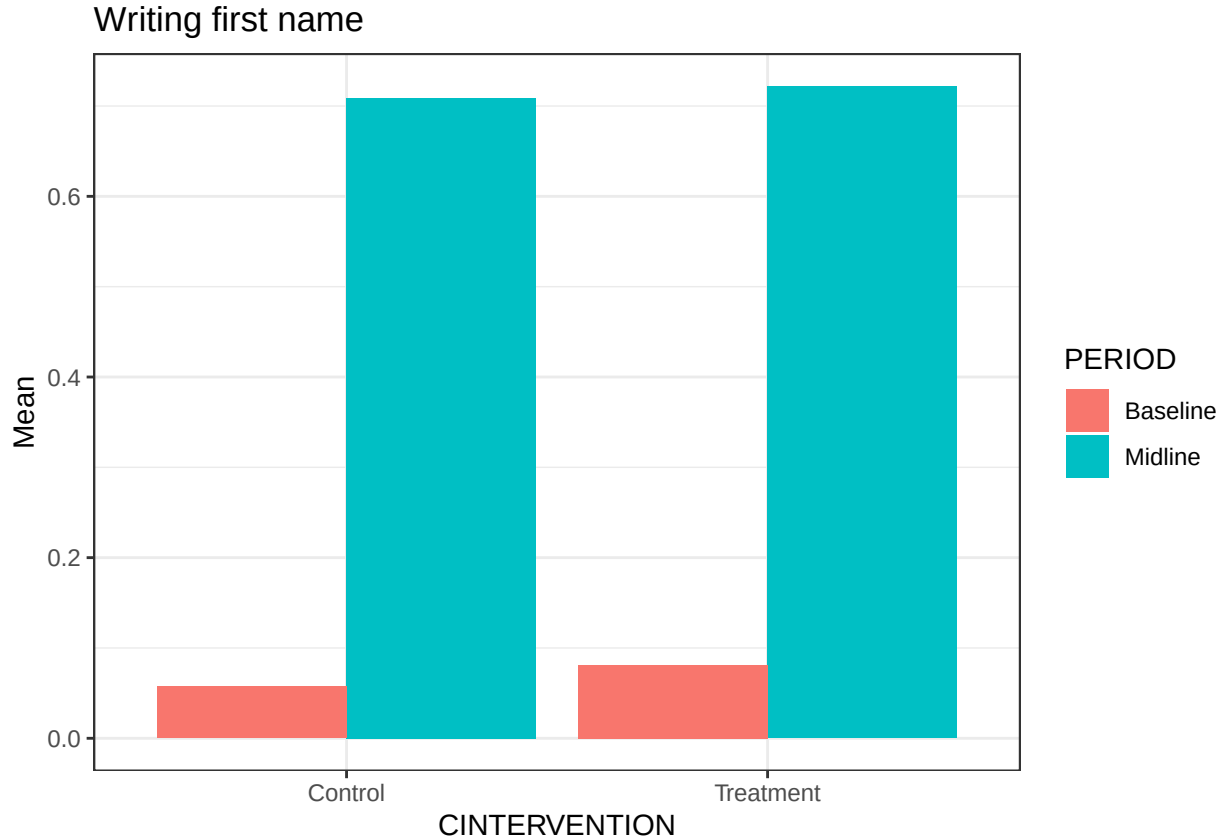
2 EGRA Categorical Variables

2.1 EGRA_ST9_1A: Writing first name

2.1.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 76: Writing first name

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.057	1136	0.709	1081
Treatment	0.081	1578	0.722	1808

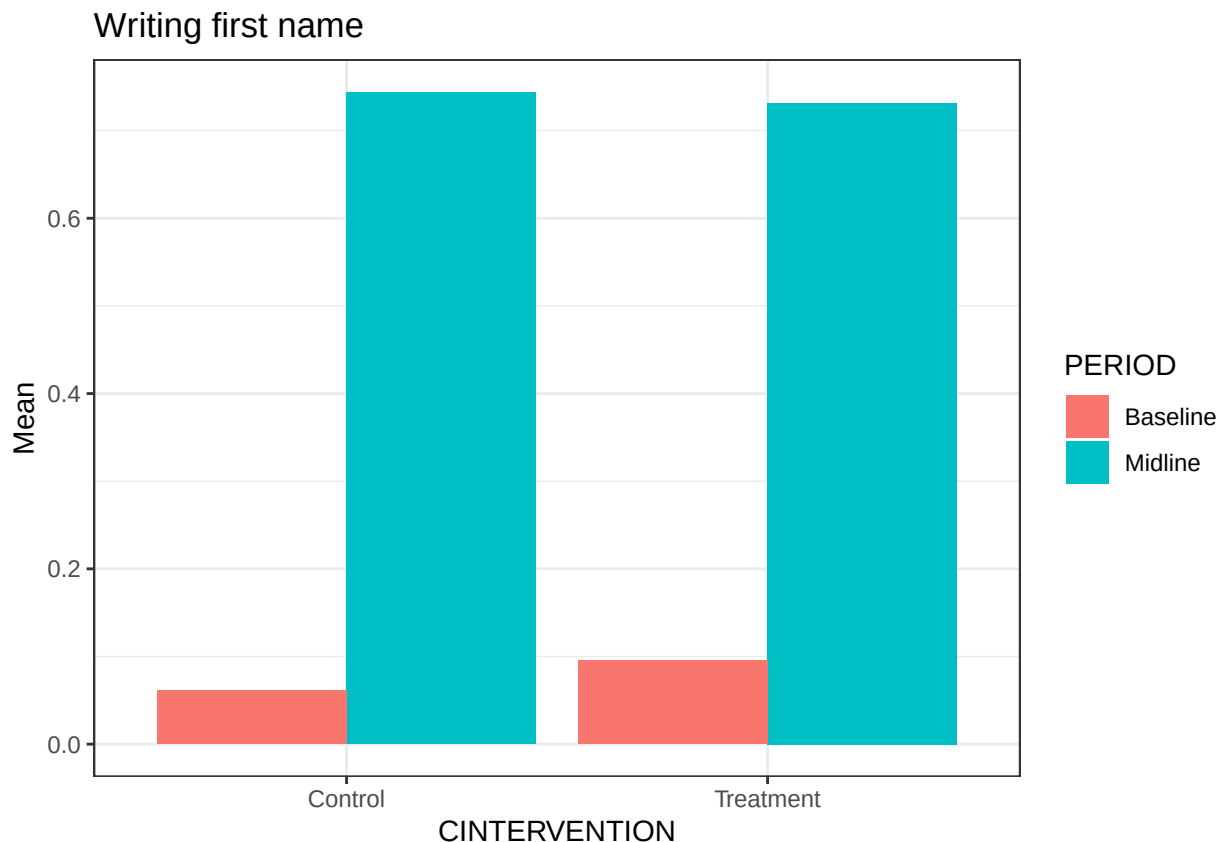


As shown in the table above, for the the Writing first name EGRA subtask, the percentage correct for the Control (Comparison (all)) condition at baseline was 0.05721831 and the percentage correct for the Treatment (FFE + lit (all)) condition at baseline was 0.08111534. The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at baseline was thus of 0.02389703 percentage points. The p-value for this difference was 0.1636942. The percentage correct for the Control (Comparison (all)) condition at midline was 0.7086031 and the percentage correct for the Treatment (FFE + lit (all)) condition at midline was 0.721792. The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at midline was thus of 0.01318889 points. The p-value for this difference was 0.6214181. The change from the baseline to the midline of 0.6513848 points for the Control (Comparison (all)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (all)) can be assessed against. The change for the Treatment (FFE + lit (all)) from the baseline to the midline was of 0.6406767 points. Consequently, the change for the Control (Comparison (all)) relative to the change for the Treatment (FFE + lit (all)) condition was of -0.01070814 points. The p-value for this difference was 0.2805932. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Writing first name EGRA subtask across the Control (Comparison (all)) and the Treatment (FFE + lit (all)) conditions beyond that expected from sampling error.

2.1.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 77: Writing first name

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.061	1047	0.744	972
Treatment	0.095	1040	0.732	1047



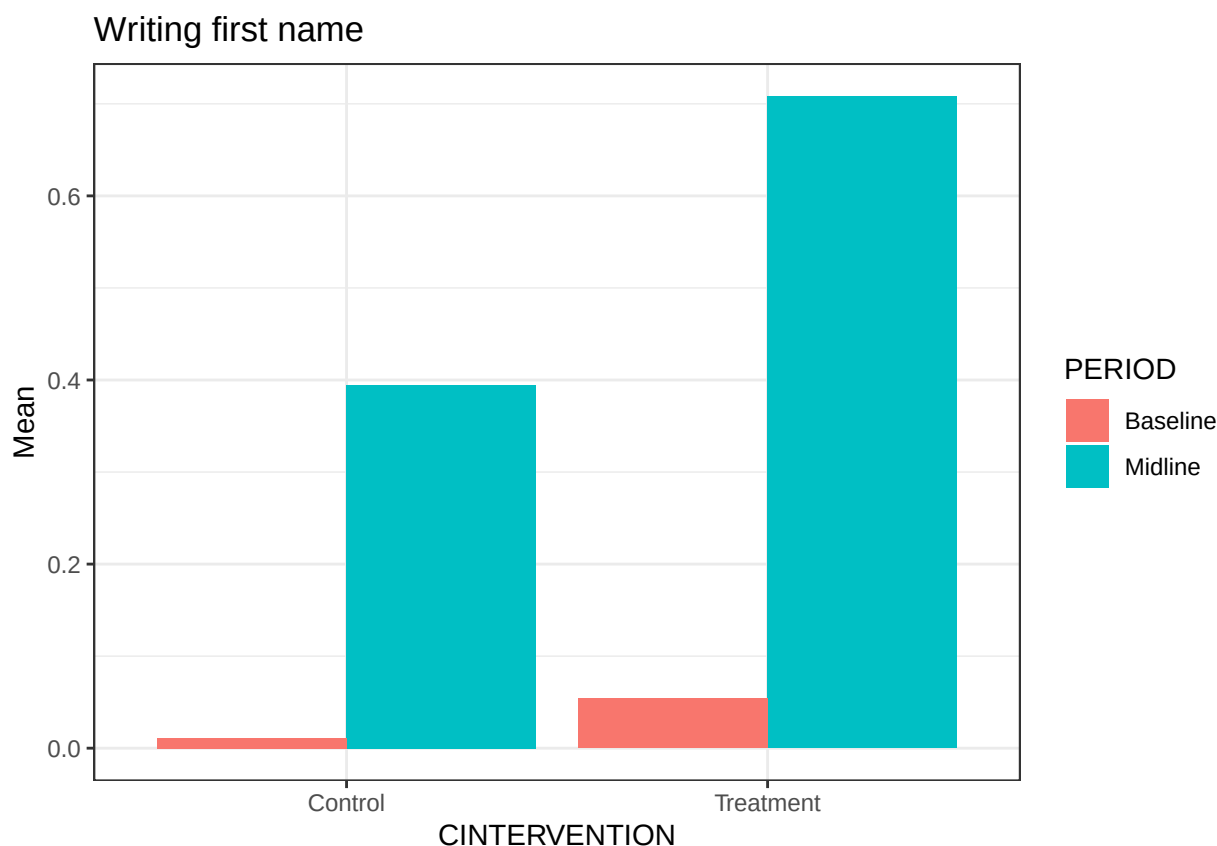
As shown in the table above, for the the Writing first name EGRA subtask, the percentage correct for the Control (Comparison (Portuguese)) condition at baseline was 0.06112703 and the percentage correct for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.09519231. The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.03406528 percentage points. The p-value for this difference was 0.09135377. The percentage correct for the Control (Comparison (Portuguese)) condition at midline was 0.7438272 and the percentage correct for the Treatment (FFE + lit (Portuguese)) condition at midline was 0.7316141. The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.01221302 points. The p-value for this difference was 0.6539434. The change from the baseline to the midline of 0.6827001 points for the Control (Comparison (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 0.6364218 points. Consequently, the change for the Control (Comparison (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of -0.0462783 points. The p-value for this difference was 0.07974952. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between

the baseline and the midline for the Writing first name EGRA subtask across the Control (Comparison (Portuguese)) and the Treatment (FFE + lit (Portuguese)) conditions beyond that expected from sampling error.

2.1.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 78: Writing first name

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.011	89	0.394	109
Treatment	0.054	538	0.708	761



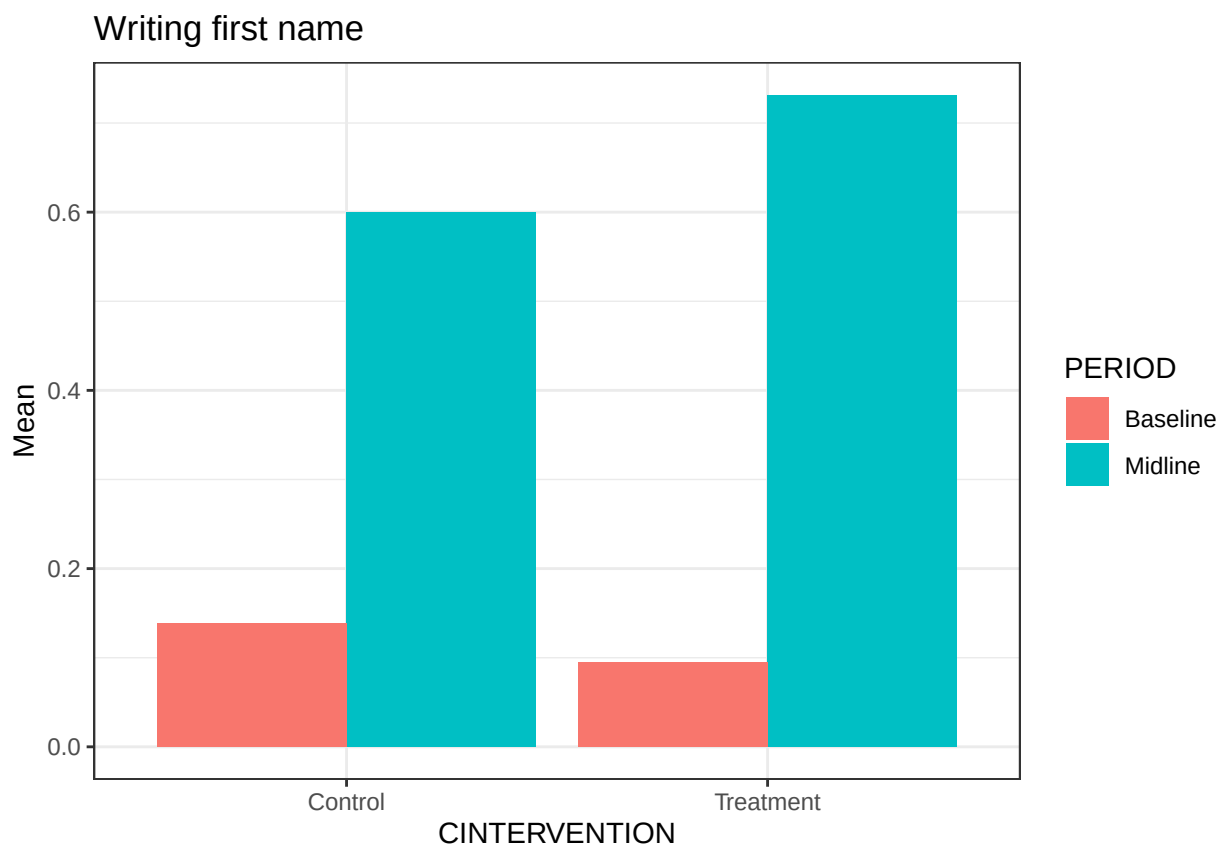
As shown in the table above, for the the Writing first name EGRA subtask, the percentage correct for the Control (Comparison (Bilingual)) condition at baseline was 0.01123596 and the percentage correct for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.05390335. The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.04266739 percentage points. The p-value for this difference was 0.1000596. The percentage correct for the Control (Comparison (Bilingual)) condition at midline was 0.3944954 and the percentage correct for the Treatment (FFE + lit (Bilingual)) condition at midline was 0.7082786. The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.3137832 points. The p-value for this difference was 4.158437e-06. The change from the baseline to the midline of 0.3832595 points for the Control (Comparison (Bilingual)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 0.6543752 points. Consequently, the change for the

Control (Comparison (Bilingual)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 0.2711158 points. The p-value for this difference was 0.7380733. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Writing first name EGRA subtask across the Control (Comparison (Bilingual)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

2.1.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 79: Writing first name

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.139	1007	0.600	952
Treatment	0.095	1040	0.732	1047



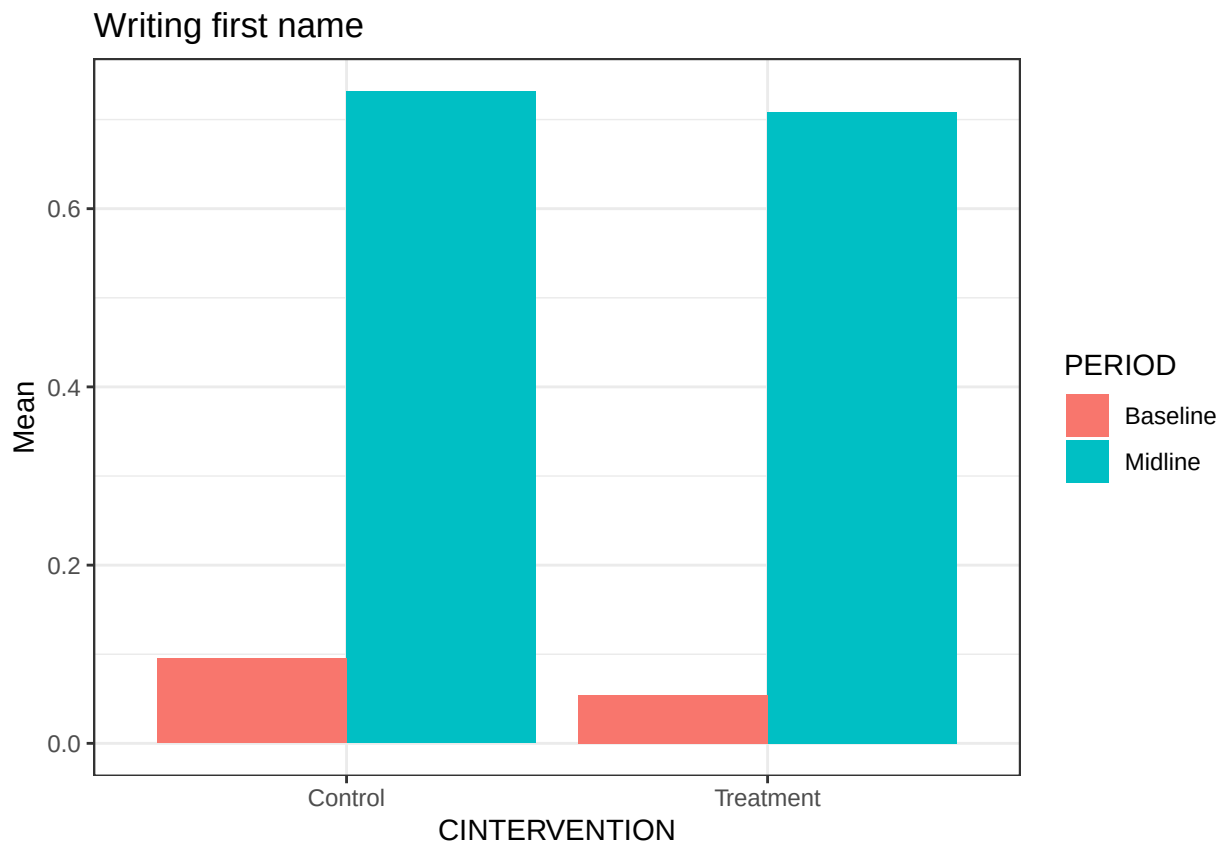
As shown in the table above, for the the Writing first name EGRA subtask, the percentage correct for the Control (FFE only (Portuguese)) condition at baseline was 0.1390268 and the percentage correct for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.09519231. The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.0438345 percentage points. The p-value for this difference was 0.05972446. The percentage correct for the Control (FFE only (Portuguese)) condition at midline was 0.5997899 and the percentage correct for the Treatment (FFE + lit (Portuguese)) condition at midline was 0.7316141. The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.1318242 points. The p-value for this difference was 1.907784e-05. The change from the baseline to the midline of 0.4607631 points for

the Control (FFE only (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 0.6364218 points. Consequently, the change for the Control (FFE only (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 0.1756587 points. The p-value for this difference was 3.997563e-05. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Portuguese)) when compared to the Control (FFE only (Portuguese)) condition. This provides evidence that the Treatment (FFE + lit (Portuguese)) (or some other unobserved process) impacted on the Writing first name EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

2.1.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 80: Writing first name

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.095	1040	0.732	1047
Treatment	0.054	538	0.708	761



As shown in the table above, for the the Writing first name EGRA subtask, the percentage correct for the Control (FFE + lit (Portuguese)) condition at baseline was 0.09519231 and the percentage correct for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.05390335. The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.04128896 percentage points. The p-value for this difference was 0.08149245. The percentage correct for the Control

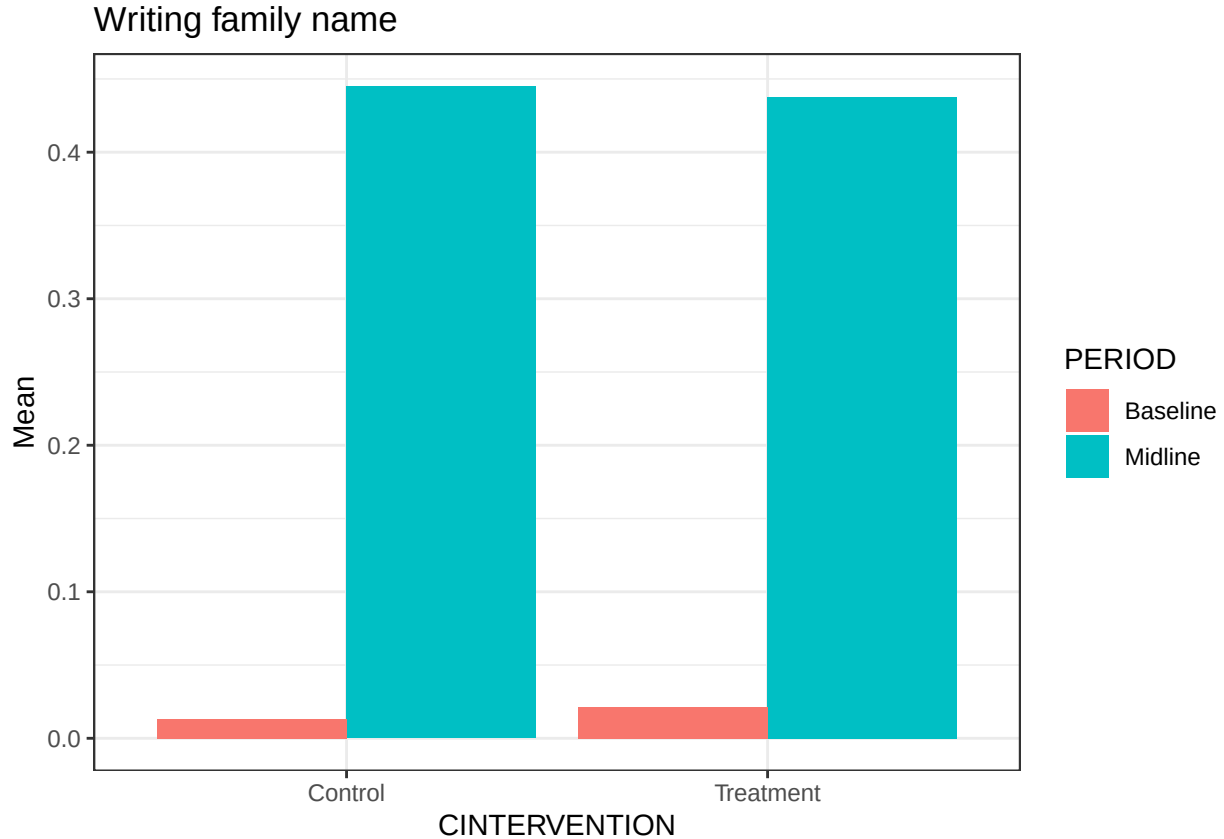
(FFE + lit (Portuguese)) condition at midline was 0.7316141 and the percentage correct for the Treatment (FFE + lit (Bilingual)) condition at midline was 0.7082786. The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.02333555 points. The p-value for this difference was 0.4414178. The change from the baseline to the midline of 0.6364218 points for the Control (FFE + lit (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 0.6543752 points. Consequently, the change for the Control (FFE + lit (Portuguese)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 0.01795341 points. The p-value for this difference was 0.1673764. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Writing first name EGRA subtask across the Control (FFE + lit (Portuguese)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

2.2 EGRA_ST9_1B: Writing family name

2.2.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 81: Writing family name

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.013	1136	0.445	1081
Treatment	0.022	1578	0.438	1808

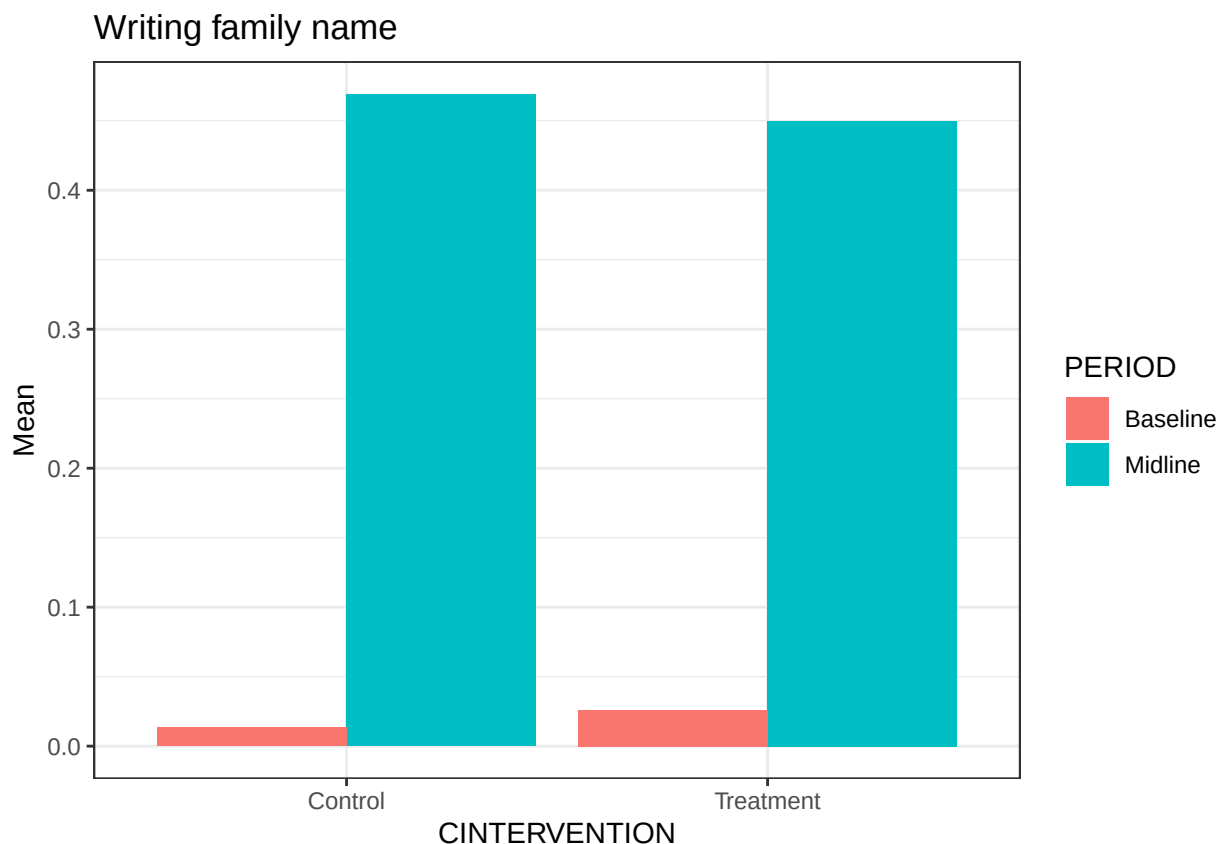


As shown in the table above, for the the Writing family name EGRA subtask, the percentage correct for the Control (Comparison (all)) condition at baseline was 0.01320423 and the percentage correct for the Treatment (FFE + lit (all)) condition at baseline was 0.02154626. The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at baseline was thus of 0.008342036 percentage points. The p-value for this difference was 0.2252622. The percentage correct for the Control (Comparison (all)) condition at midline was 0.4449584 and the percentage correct for the Treatment (FFE + lit (all)) condition at midline was 0.4375. The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at midline was thus of 0.007458372 points. The p-value for this difference was 0.8121453. The change from the baseline to the midline of 0.4317541 points for the Control (Comparison (all)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (all)) can be assessed against. The change for the Treatment (FFE + lit (all)) from the baseline to the midline was of 0.4159537 points. Consequently, the change for the Control (Comparison (all)) relative to the change for the Treatment (FFE + lit (all)) condition was of -0.01580041 points. The p-value for this difference was 0.2086559. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Writing family name EGRA subtask across the Control (Comparison (all)) and the Treatment (FFE + lit (all)) conditions beyond that expected from sampling error.

2.2.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 82: Writing family name

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.013	1047	0.469	972
Treatment	0.026	1040	0.450	1047



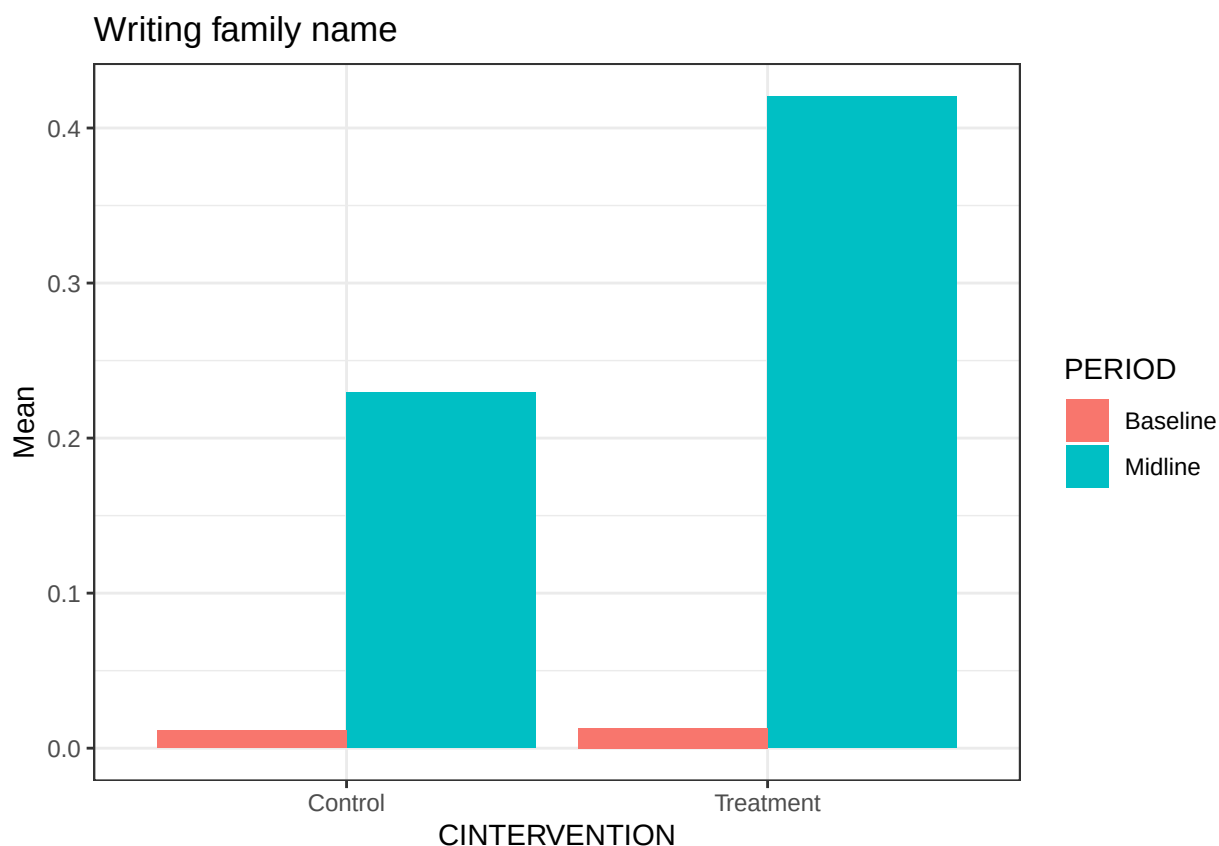
As shown in the table above, for the the Writing family name EGRA subtask, the percentage correct for the Control (Comparison (Portuguese)) condition at baseline was 0.01337154 and the percentage correct for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.02596154. The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.01259 percentage points. The p-value for this difference was 0.1339904. The percentage correct for the Control (Comparison (Portuguese)) condition at midline was 0.4691358 and the percentage correct for the Treatment (FFE + lit (Portuguese)) condition at midline was 0.4498567. The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.01927907 points. The p-value for this difference was 0.5487912. The change from the baseline to the midline of 0.4557643 points for the Control (Comparison (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 0.4238952 points. Consequently, the change for the Control (Comparison (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of -0.03186907 points. The p-value for this difference was 0.09599616. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of

change between the baseline and the midline for the Writing family name EGRA subtask across the Control (Comparison (Portuguese)) and the Treatment (FFE + lit (Portuguese)) conditions beyond that expected from sampling error.

2.2.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 83: Writing family name

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.011	89	0.229	109
Treatment	0.013	538	0.420	761



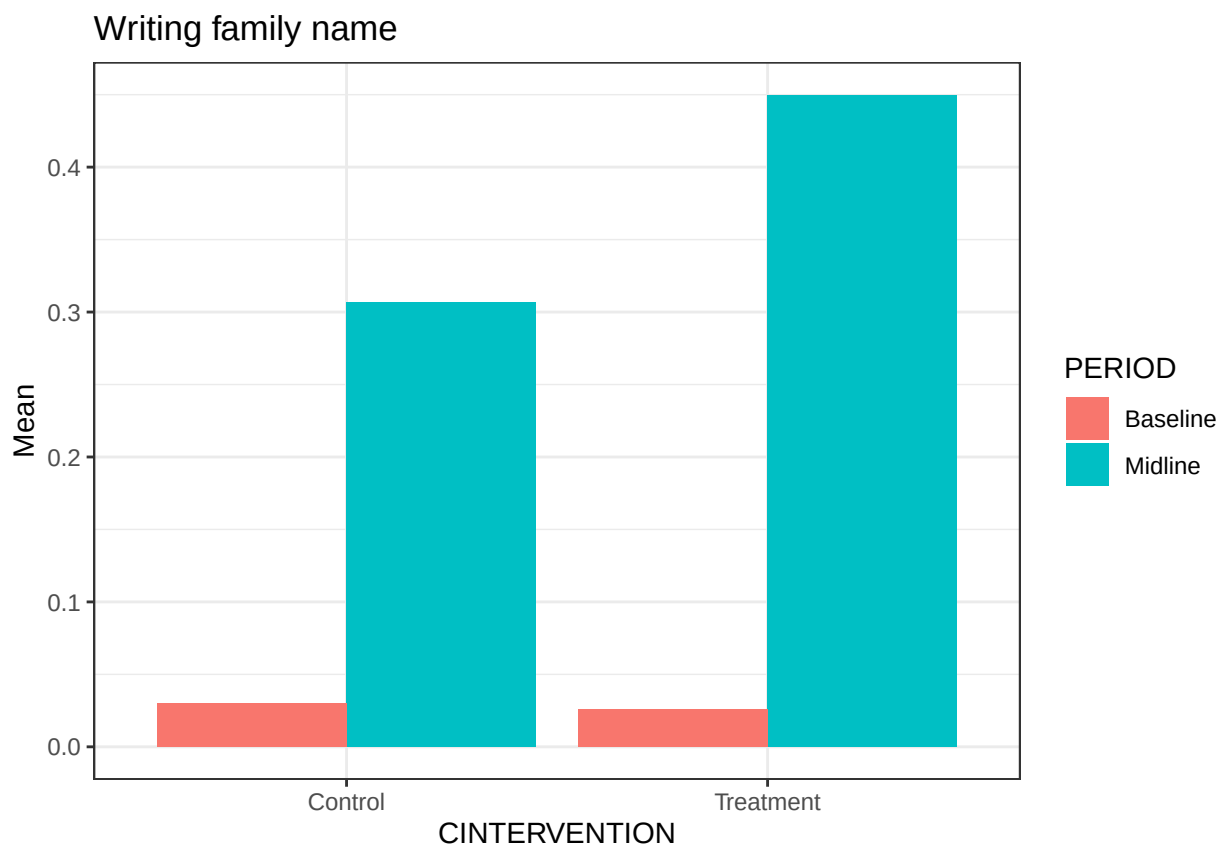
As shown in the table above, for the the Writing family name EGRA subtask, the percentage correct for the Control (Comparison (Bilingual)) condition at baseline was 0.01123596 and the percentage correct for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.01301115. The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.001775197 percentage points. The p-value for this difference was 0.8855256. The percentage correct for the Control (Comparison (Bilingual)) condition at midline was 0.2293578 and the percentage correct for the Treatment (FFE + lit (Bilingual)) condition at midline was 0.4204993. The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.1911415 points. The p-value for this difference was 0.03141115. The change from the baseline to the midline of 0.2181218 points for the Control (Comparison (Bilingual)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 0.4074882 points. Consequently, the change for

the Control (Comparison (Bilingual)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 0.1893663 points. The p-value for this difference was 0.5372766. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Writing family name EGRA subtask across the Control (Comparison (Bilingual)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

2.2.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 84: Writing family name

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.030	1007	0.307	952
Treatment	0.026	1040	0.450	1047



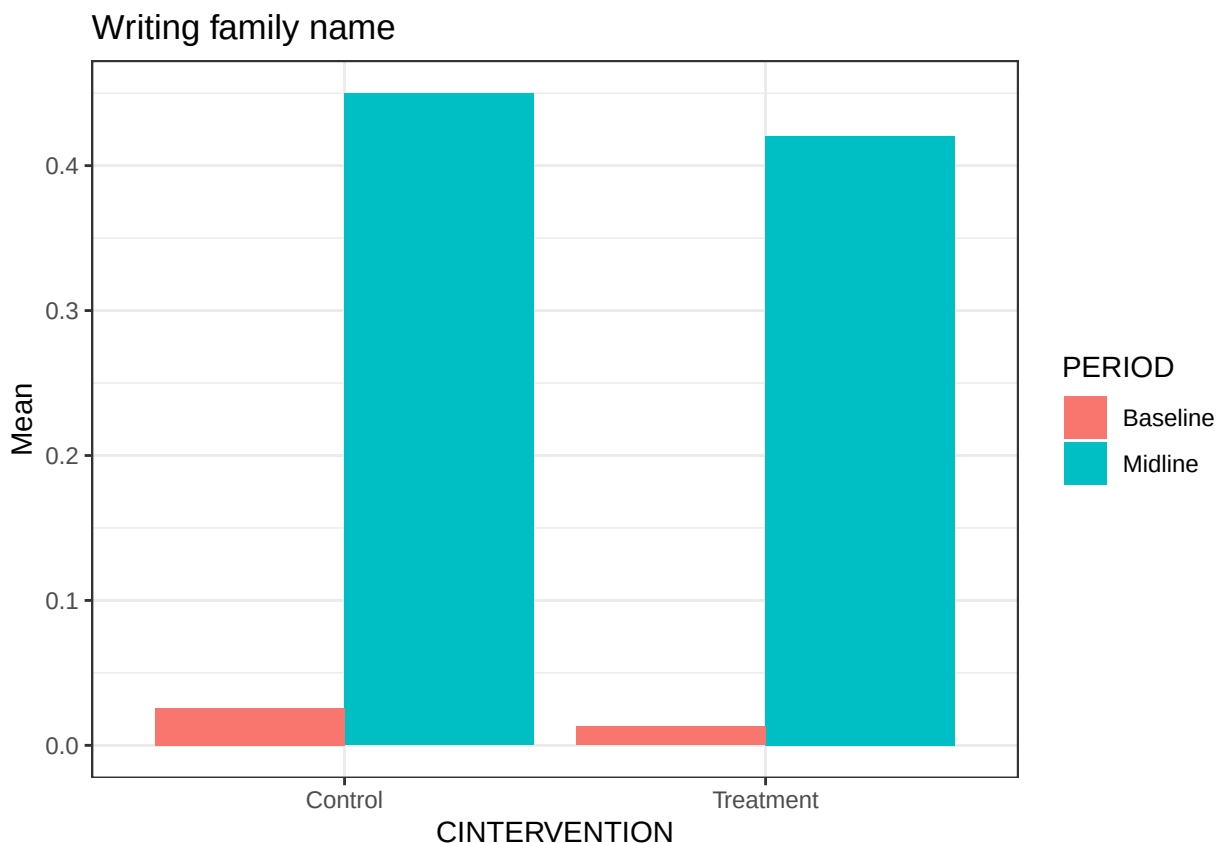
As shown in the table above, for the the Writing family name EGRA subtask, the percentage correct for the Control (FFE only (Portuguese)) condition at baseline was 0.02979146 and the percentage correct for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.02596154. The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.003829921 percentage points. The p-value for this difference was 0.6835565. The percentage correct for the Control (FFE only (Portuguese)) condition at midline was 0.3067227 and the percentage correct for the Treatment (FFE + lit (Portuguese)) condition at midline was 0.4498567. The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.143134 points. The p-value for this difference was 0.0001087363. The change from the baseline to the midline of 0.2769312 points for

the Control (FFE only (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 0.4238952 points. Consequently, the change for the Control (FFE only (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 0.146964 points. The p-value for this difference was 0.03874017. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Writing family name EGRA subtask across the Control (FFE only (Portuguese)) and the Treatment (FFE + lit (Portuguese)) conditions beyond that expected from sampling error.

2.2.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 85: Writing family name

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.026	1040	0.45	1047
Treatment	0.013	538	0.42	761



As shown in the table above, for the the Writing family name EGRA subtask, the percentage correct for the Control (FFE + lit (Portuguese)) condition at baseline was 0.02596154 and the percentage correct for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.01301115. The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.01295039 percentage points. The p-value for this difference was 0.1866927. The percentage correct for the Control

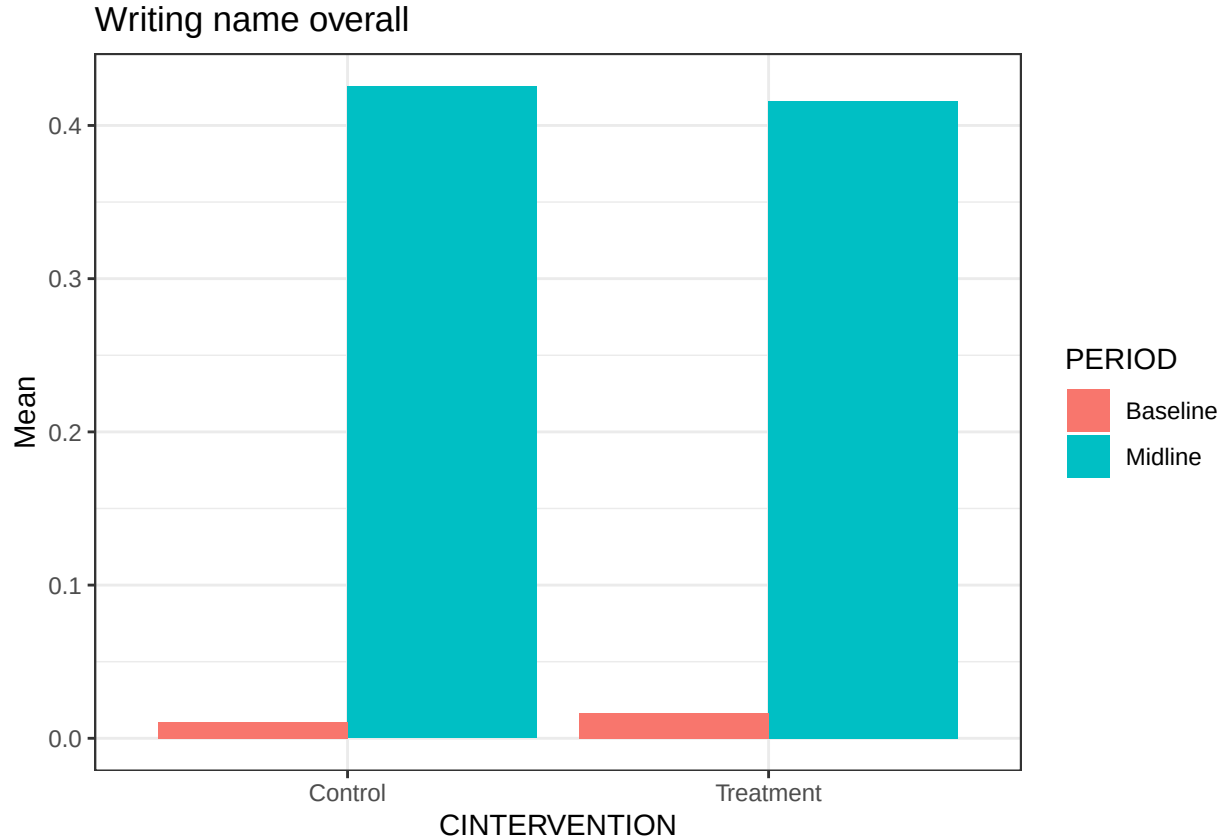
(FFE + lit (Portuguese)) condition at midline was 0.4498567 and the percentage correct for the Treatment (FFE + lit (Bilingual)) condition at midline was 0.4204993. The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.02935739 points. The p-value for this difference was 0.4254192. The change from the baseline to the midline of 0.4238952 points for the Control (FFE + lit (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 0.4074882 points. Consequently, the change for the Control (FFE + lit (Portuguese)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of -0.016407 points. The p-value for this difference was 0.2840003. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Writing family name EGRA subtask across the Control (FFE + lit (Portuguese)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

2.3 EGRA_ST9_1C: Writing name overall

2.3.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 86: Writing name overall

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.011	1136	0.426	1081
Treatment	0.016	1578	0.416	1808

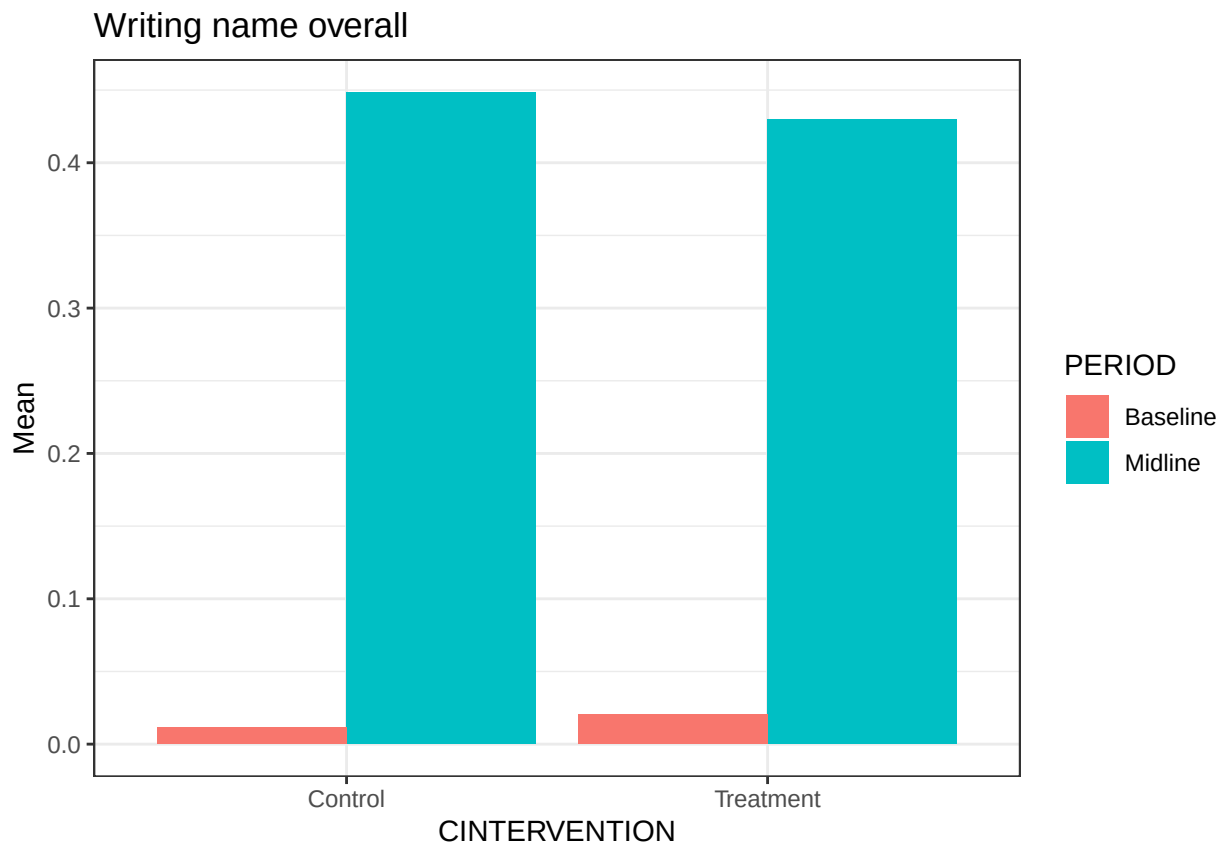


As shown in the table above, for the the Writing name overall EGRA subtask, the percentage correct for the Control (Comparison (all)) condition at baseline was 0.01056338 and the percentage correct for the Treatment (FFE + lit (all)) condition at baseline was 0.01647655. The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at baseline was thus of 0.005913172 percentage points. The p-value for this difference was 0.3554128. The percentage correct for the Control (Comparison (all)) condition at midline was 0.4255319 and the percentage correct for the Treatment (FFE + lit (all)) condition at midline was 0.4159292. The difference between Control (Comparison (all)) and Treatment (FFE + lit (all)) at midline was thus of 0.009602711 points. The p-value for this difference was 0.7573427. The change from the baseline to the midline of 0.4149685 points for the Control (Comparison (all)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (all)) can be assessed against. The change for the Treatment (FFE + lit (all)) from the baseline to the midline was of 0.3994527 points. Consequently, the change for the Control (Comparison (all)) relative to the change for the Treatment (FFE + lit (all)) condition was of -0.01551588 points. The p-value for this difference was 0.3152659. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Writing name overall EGRA subtask across the Control (Comparison (all)) and the Treatment (FFE + lit (all)) conditions beyond that expected from sampling error.

2.3.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 87: Writing name overall

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.011	1047	0.449	972
Treatment	0.020	1040	0.430	1047



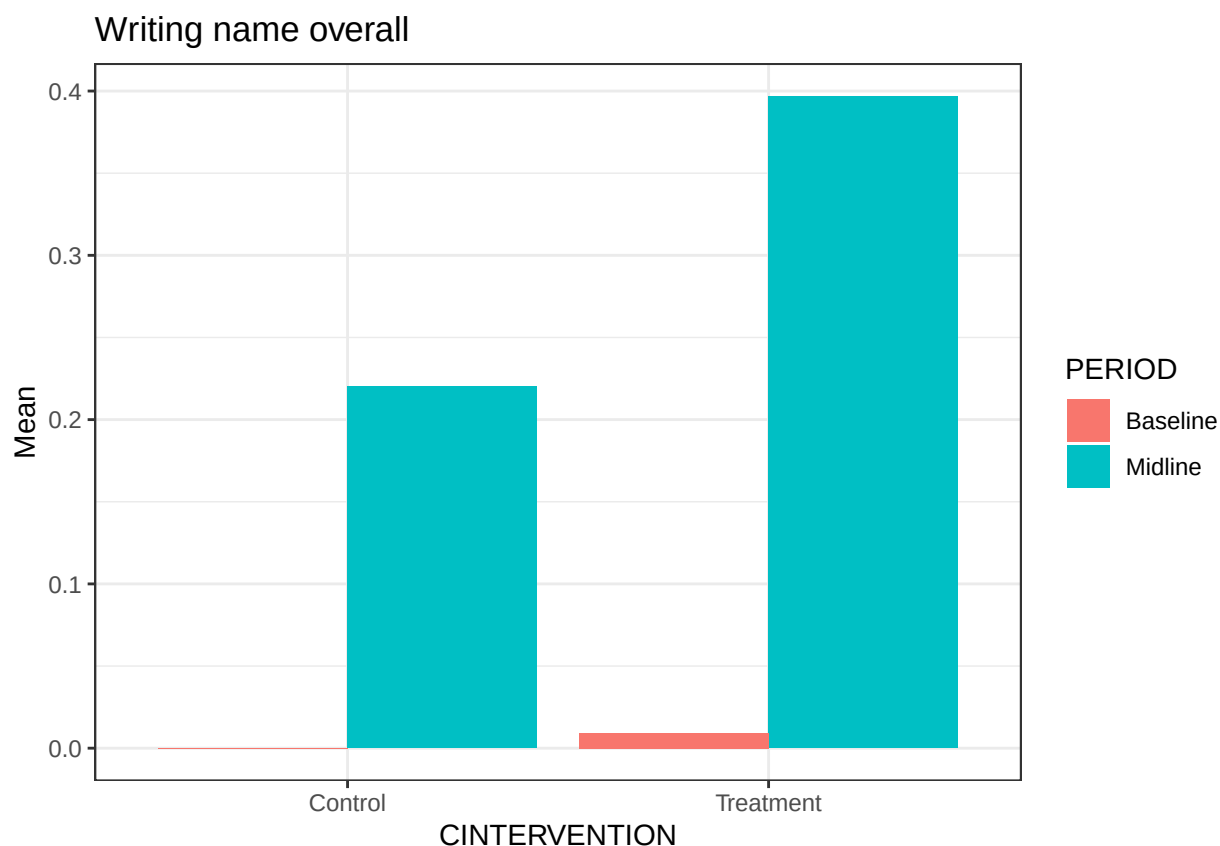
As shown in the table above, for the the Writing name overall EGRA subtask, the percentage correct for the Control (Comparison (Portuguese)) condition at baseline was 0.01146132 and the percentage correct for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.02019231. The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.00873099 percentage points. The p-value for this difference was 0.2690187. The percentage correct for the Control (Comparison (Portuguese)) condition at midline was 0.4485597 and the percentage correct for the Treatment (FFE + lit (Portuguese)) condition at midline was 0.4297994. The difference between Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.01876024 points. The p-value for this difference was 0.5658451. The change from the baseline to the midline of 0.4370984 points for the Control (Comparison (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 0.4096071 points. Consequently, the change for the Control (Comparison (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of -0.02749123 points. The p-value for this difference was 0.2054075. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between

the baseline and the midline for the Writing name overall EGRA subtask across the Control (Comparison (Portuguese)) and the Treatment (FFE + lit (Portuguese)) conditions beyond that expected from sampling error.

2.3.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 88: Writing name overall

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.000	89	0.220	109
Treatment	0.009	538	0.397	761



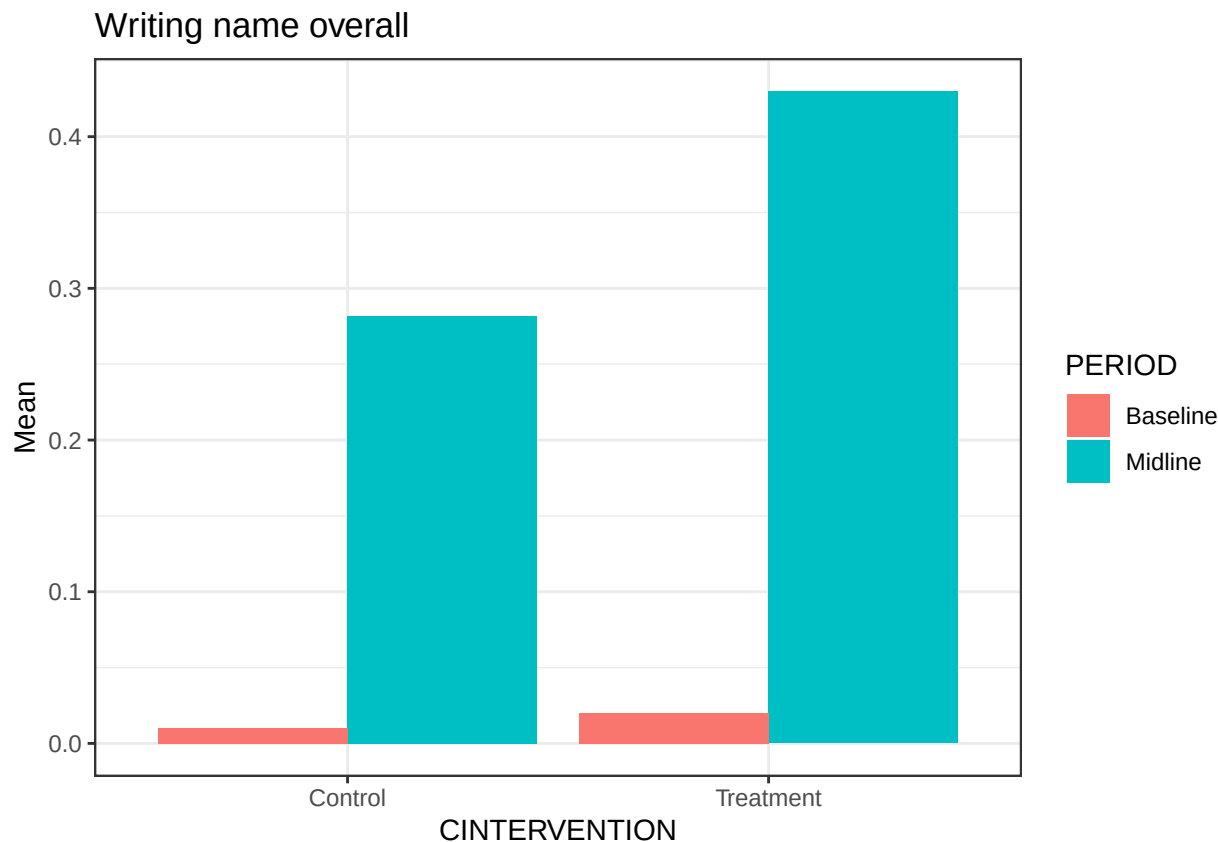
As shown in the table above, for the the Writing name overall EGRA subtask, the percentage correct for the Control (Comparison (Bilingual)) condition at baseline was 0 and the percentage correct for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.00929368. The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.00929368 percentage points. The p-value for this difference was 3.017859e-138. The percentage correct for the Control (Comparison (Bilingual)) condition at midline was 0.2201835 and the percentage correct for the Treatment (FFE + lit (Bilingual)) condition at midline was 0.3968463. The difference between Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.1766628 points. The p-value for this difference was 0.03840806. The change from the baseline to the midline of 0.2201835 points for the Control (Comparison (Bilingual)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 0.3875526 points. Consequently, the change for the Control (Comparison (Bilingual))

relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of 0.1673691 points. The p-value for this difference was 9.558092e-53. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts) and thus there is evidence of a different degree of change from baseline to midline in the Treatment (FFE + lit (Bilingual)) when compared to the Control (Comparison (Bilingual)) condition. This provides evidence that the Treatment (FFE + lit (Bilingual)) (or some other unobserved process) impacted on the Writing name overall EGRA subtask and thus the rate of change of the two conditions relative to their baselines differed.

2.3.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 89: Writing name overall

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.01	1007	0.282	952
Treatment	0.02	1040	0.430	1047



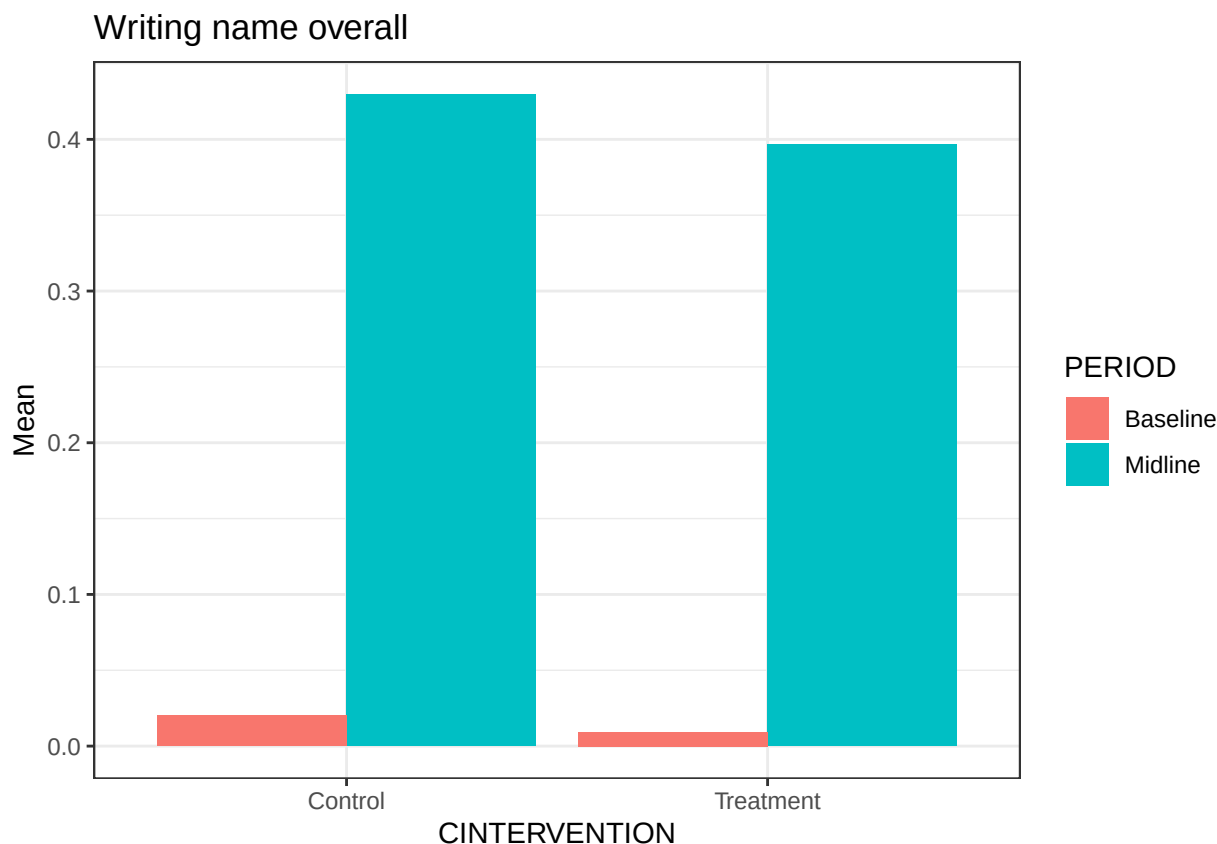
As shown in the table above, for the the Writing name overall EGRA subtask, the percentage correct for the Control (FFE only (Portuguese)) condition at baseline was 0.009930487 and the percentage correct for the Treatment (FFE + lit (Portuguese)) condition at baseline was 0.02019231. The difference between Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) at baseline was thus of 0.01026182 percentage points. The p-value for this difference was 0.1579548. The percentage correct for the Control (FFE only (Portuguese)) condition at midline was 0.2815126 and the percentage correct for the Treatment (FFE + lit (Portuguese)) condition at midline was 0.4297994. The difference between Control (FFE only

(Portuguese)) and Treatment (FFE + lit (Portuguese)) at midline was thus of 0.1482868 points. The p-value for this difference was 3.8535e-05. The change from the baseline to the midline of 0.2715821 points for the Control (FFE only (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Portuguese)) can be assessed against. The change for the Treatment (FFE + lit (Portuguese)) from the baseline to the midline was of 0.4096071 points. Consequently, the change for the Control (FFE only (Portuguese)) relative to the change for the Treatment (FFE + lit (Portuguese)) condition was of 0.138025 points. The p-value for this difference was 0.8948978. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Writing name overall EGRA subtask across the Control (FFE only (Portuguese)) and the Treatment (FFE + lit (Portuguese)) conditions beyond that expected from sampling error.

2.3.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 90: Writing name overall

Intervention	Baseline		Midline	
	%	n	%	n
Control	0.020	1040	0.430	1047
Treatment	0.009	538	0.397	761



As shown in the table above, for the the Writing name overall EGRA subtask, the percentage correct for the Control (FFE + lit (Portuguese)) condition at baseline was 0.02019231 and the percentage correct for the Treatment (FFE + lit (Bilingual)) condition at baseline was 0.00929368. The difference between

Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at baseline was thus of 0.01089863 percentage points. The p-value for this difference was 0.2044722. The percentage correct for the Control (FFE + lit (Portuguese)) condition at midline was 0.4297994 and the percentage correct for the Treatment (FFE + lit (Bilingual)) condition at midline was 0.3968463. The difference between Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) at midline was thus of 0.03295317 points. The p-value for this difference was 0.361577. The change from the baseline to the midline of 0.4096071 points for the Control (FFE + lit (Portuguese)) condition establishes the counterfactual against which the change for the Treatment (FFE + lit (Bilingual)) can be assessed against. The change for the Treatment (FFE + lit (Bilingual)) from the baseline to the midline was of 0.3875526 points. Consequently, the change for the Control (FFE + lit (Portuguese)) relative to the change for the Treatment (FFE + lit (Bilingual)) condition was of -0.02205454 points. The p-value for this difference was 0.3028833. Accordingly, the difference was not statistically significant ($p < 0.01$ corrected for 5 additional contrasts) and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Writing name overall EGRA subtask across the Control (FFE + lit (Portuguese)) and the Treatment (FFE + lit (Bilingual)) conditions beyond that expected from sampling error.

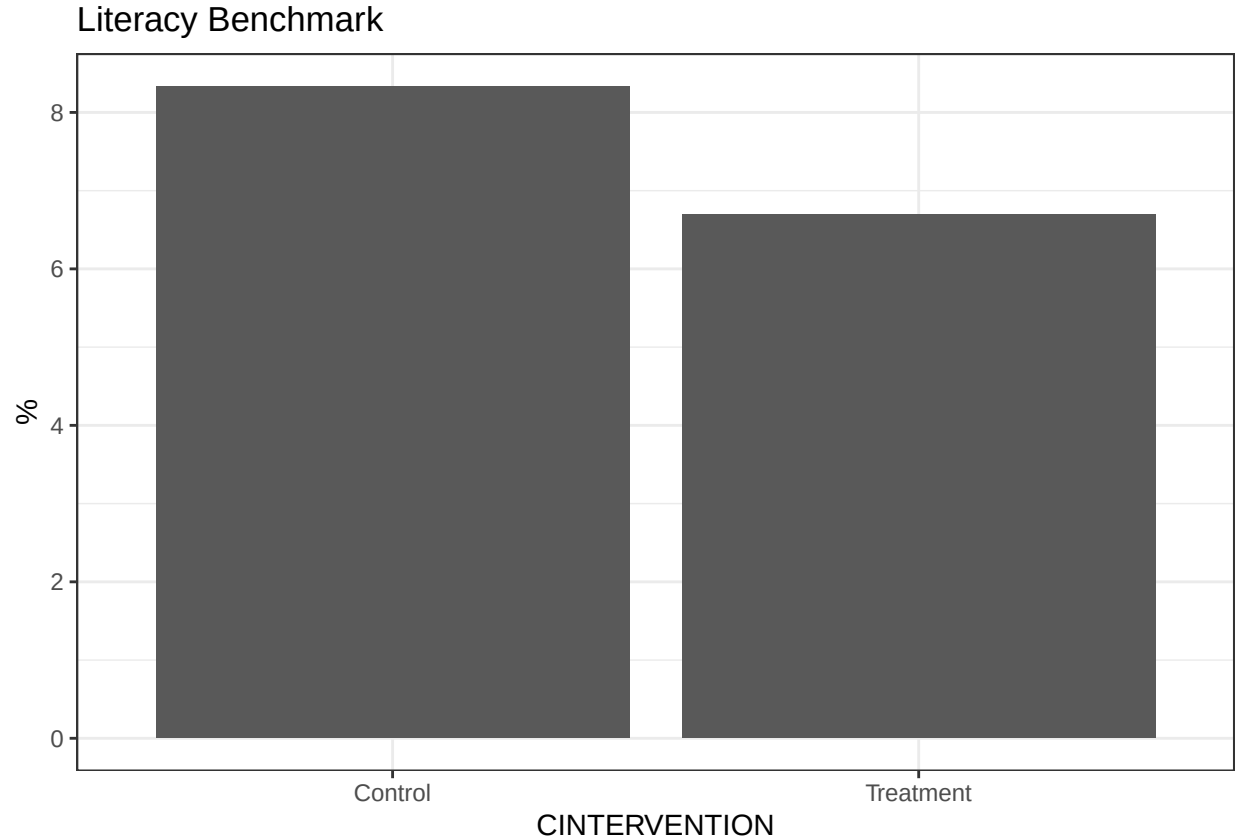
3 Literacy Benchmark

3.1 Literacy Benchmark: Boys and Girls

3.1.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 91: Literacy Benchmark

Intervention	Midline	
	%	n
Control	8.333	1080
Treatment	6.696	1792

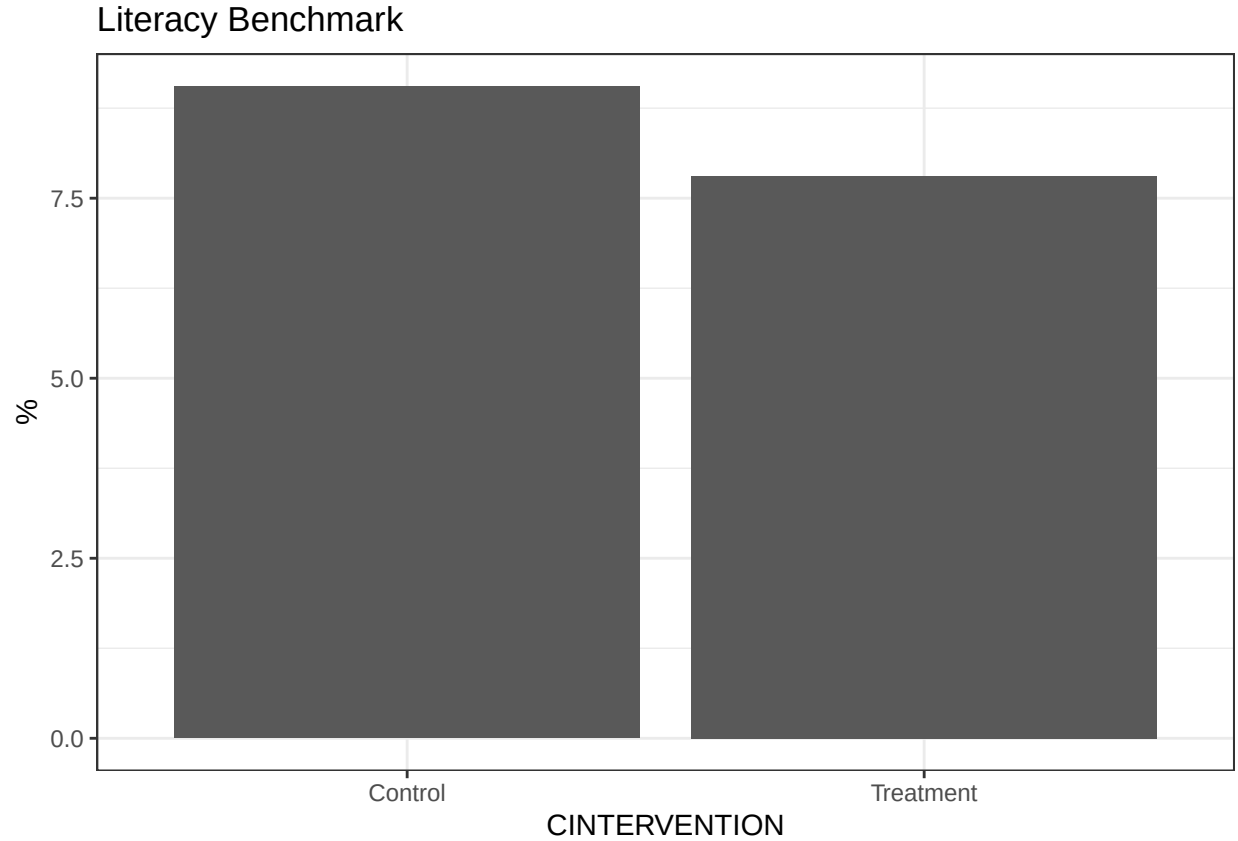


As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (Comparison (all)) condition at midline was 8.333333 and the percentage achieving it for the Treatment (FFE + lit (all)) condition at midline was 6.696429. The difference for the Literacy Benchmark across the Control (Comparison (all)) and Treatment (FFE + lit (all)) conditions at midline was thus -1.636905 points (there was no baseline measurement for this variable). The p-value for this difference was 0.2983881. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the percentage achieving the Literacy Benchmark between the Control (Comparison (all)) and Treatment (FFE + lit (all)) conditions at midline beyond that which could result from sampling error.

3.1.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 92: Literacy Benchmark

Intervention	Midline	
	%	n
Control	9.053	972
Treatment	7.811	1037

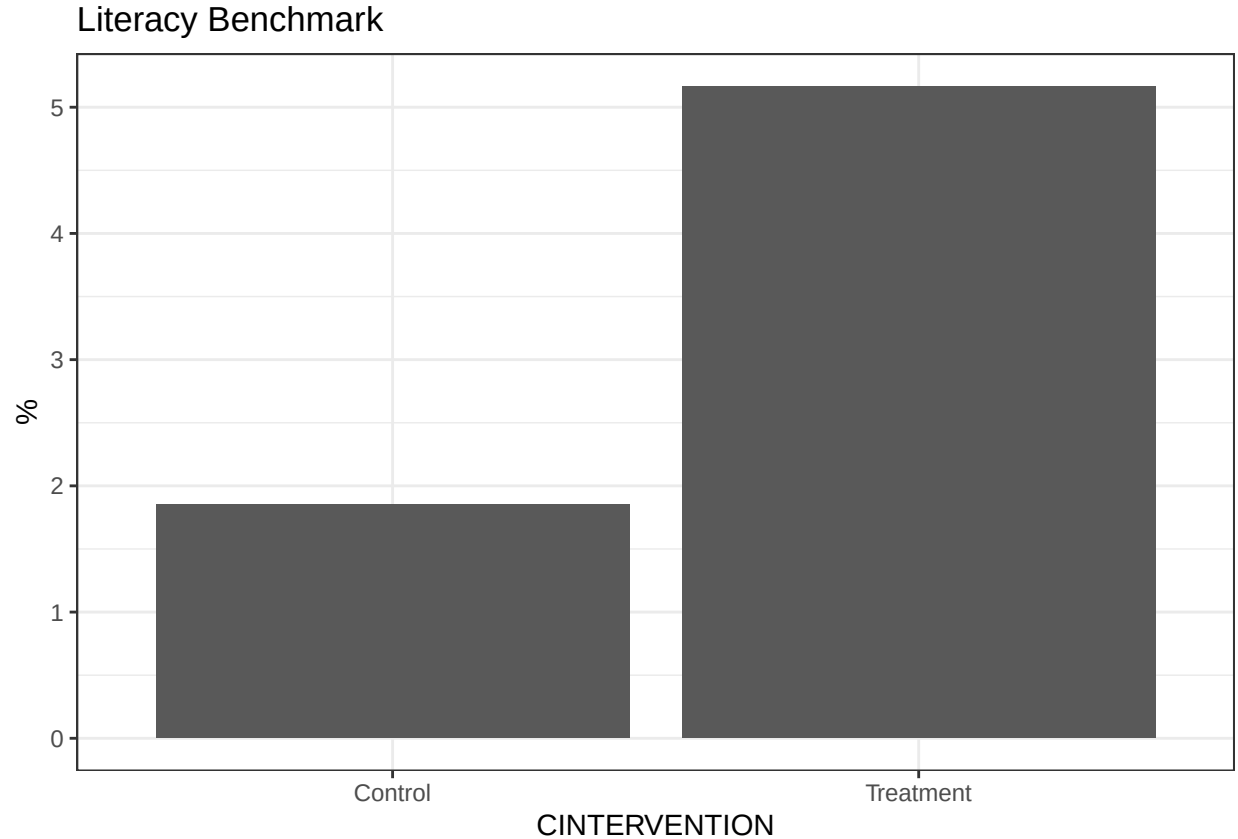


As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (Comparison (Portuguese)) condition at midline was 9.053498 and the percentage achieving it for the Treatment (FFE + lit (Portuguese)) condition at midline was 7.810993. The difference for the Literacy Benchmark across the Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline was thus -1.242505 points (there was no baseline measurement for this variable). The p-value for this difference was 0.511773. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the percentage achieving the Literacy Benchmark between the Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline beyond that which could result from sampling error.

3.1.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 93: Literacy Benchmark

Intervention	Midline	
	%	n
Control	1.852	108
Treatment	5.166	755

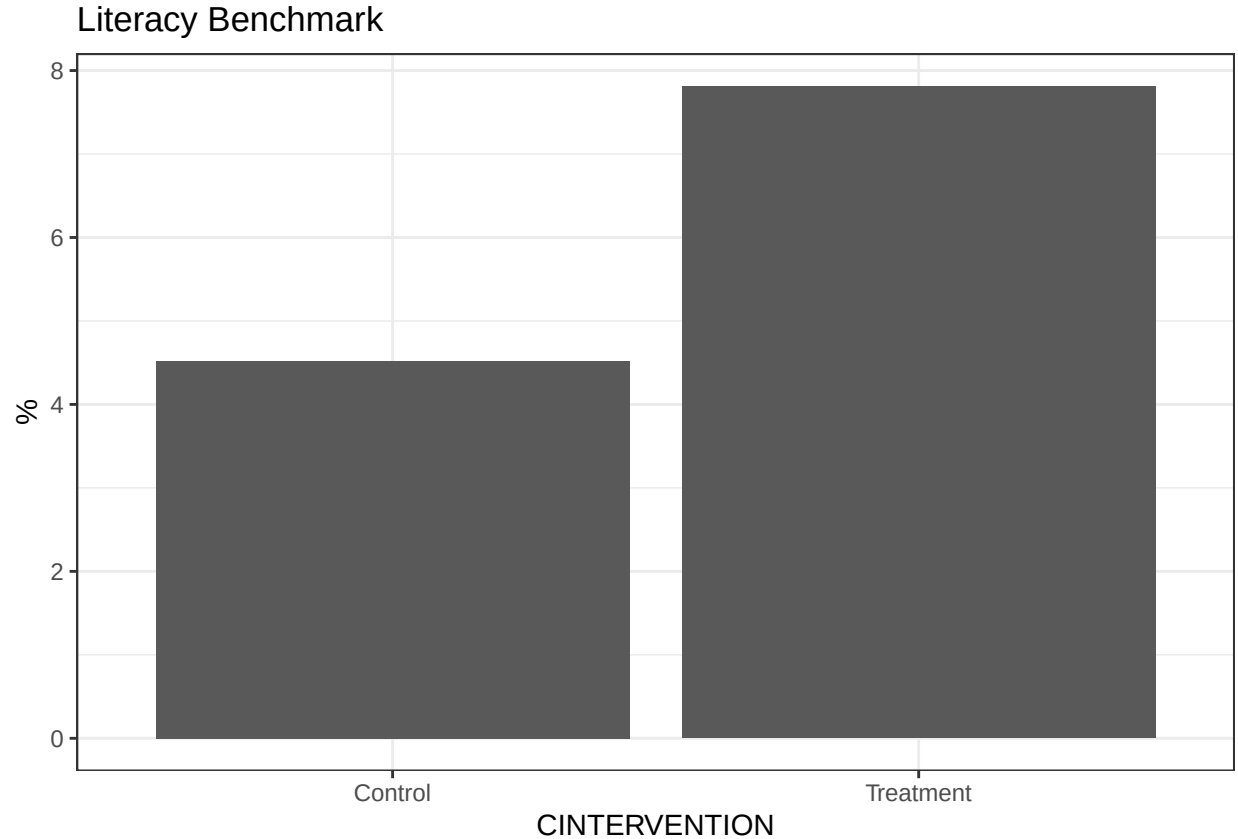


As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (Comparison (Bilingual)) condition at midline was 1.851852 and the percentage achieving it for the Treatment (FFE + lit (Bilingual)) condition at midline was 5.165563. The difference for the Literacy Benchmark across the Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) conditions at midline was thus 3.313711 points (there was no baseline measurement for this variable). The p-value for this difference was 0.2811655. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the percentage achieving the Literacy Benchmark between the Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) conditions at midline beyond that which could result from sampling error.

3.1.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 94: Literacy Benchmark

Intervention	Midline	
	%	n
Control	4.522	951
Treatment	7.811	1037

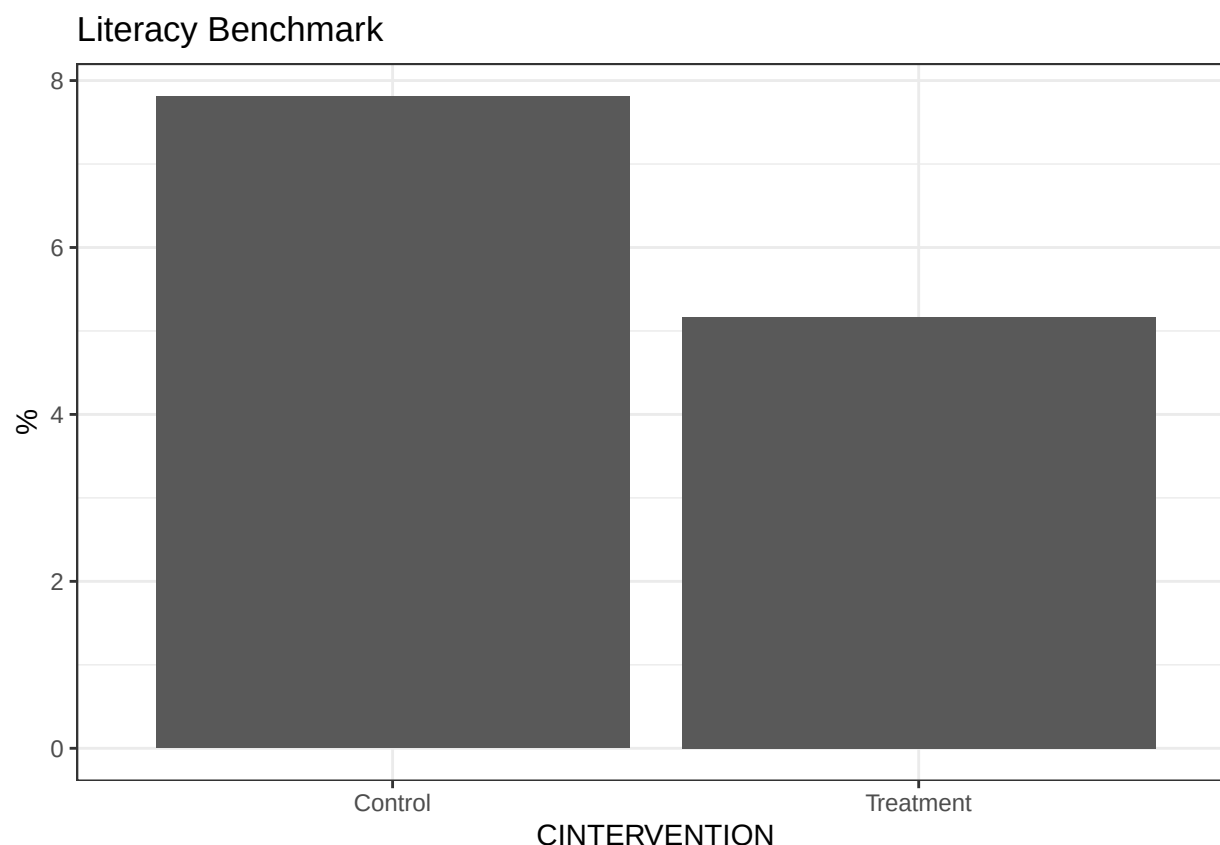


As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (FFE only (Portuguese)) condition at midline was 4.521556 and the percentage achieving it for the Treatment (FFE + lit (Portuguese)) condition at midline was 7.810993. The difference for the Literacy Benchmark across the Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline was thus 3.289437 points (there was no baseline measurement for this variable). The p-value for this difference was 0.06589293. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the percentage achieving the Literacy Benchmark between the Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline beyond that which could result from sampling error.

3.1.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 95: Literacy Benchmark

Intervention	Midline	
	%	n
Control	7.811	1037
Treatment	5.166	755



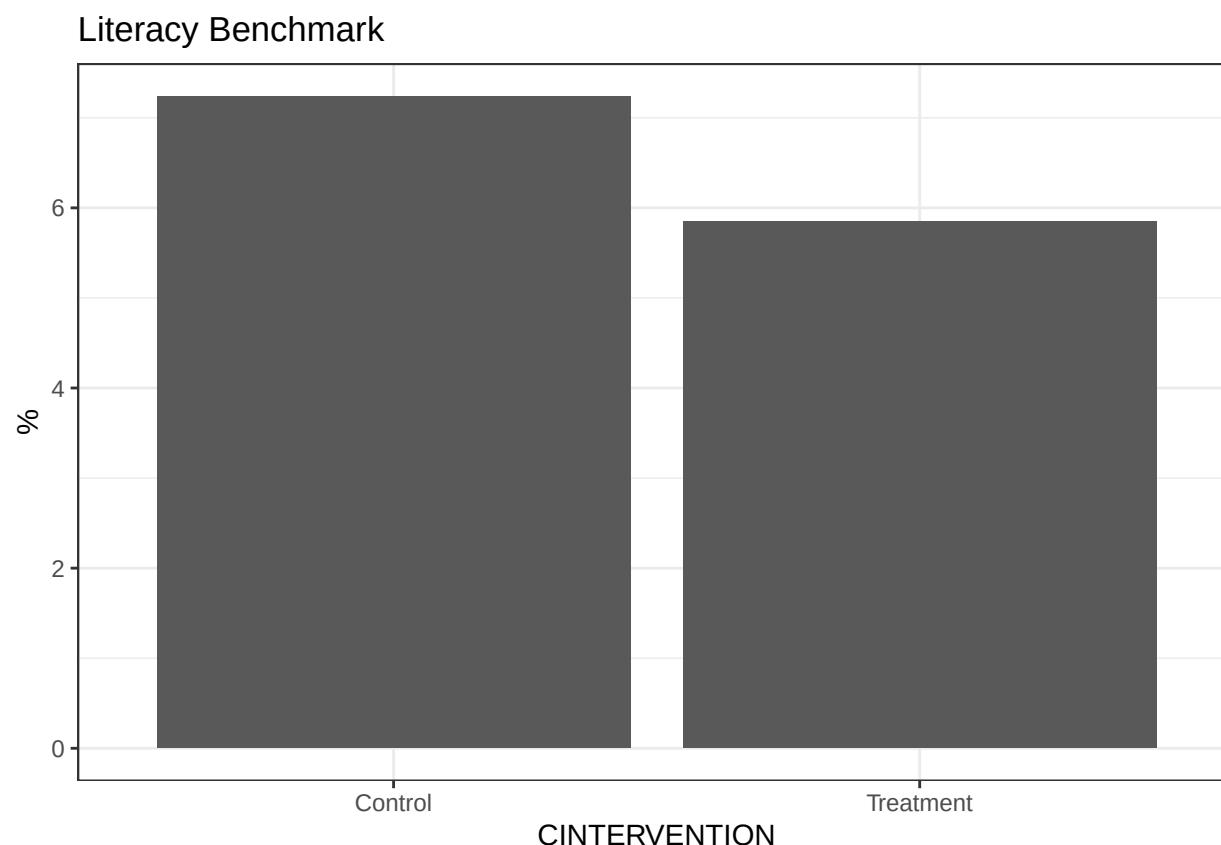
As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (FFE + lit (Portuguese)) condition at midline was 7.810993 and the percentage achieving it for the Treatment (FFE + lit (Bilingual)) condition at midline was 5.165563. The difference for the Literacy Benchmark across the Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) conditions at midline was thus -2.64543 points (there was no baseline measurement for this variable). The p-value for this difference was 0.1593429. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the percentage achieving the Literacy Benchmark between the Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) conditions at midline beyond that which could result from sampling error.

3.2 Literacy Benchmark: Boys only

3.2.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 96: Literacy Benchmark

Intervention	Midline	
	%	n
Control	7.238	525
Treatment	5.850	906

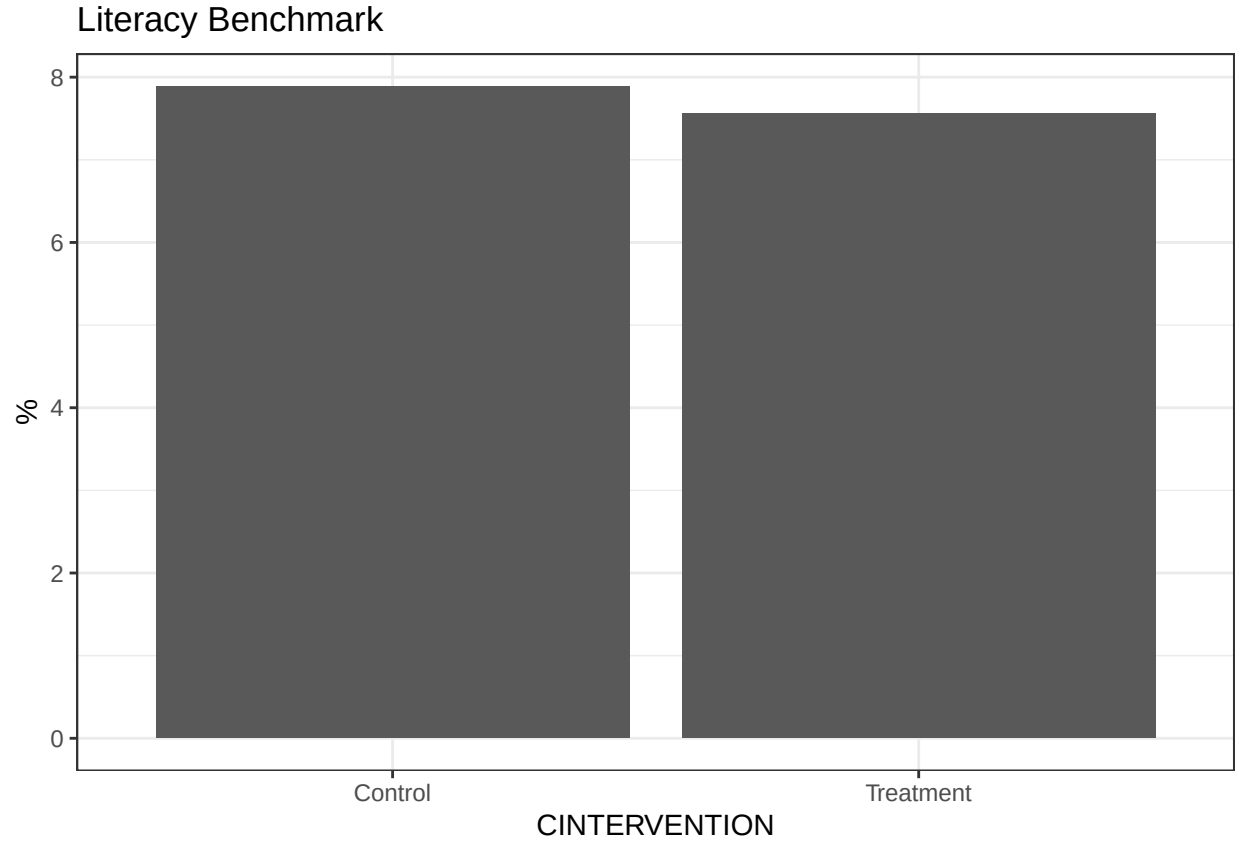


As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (Comparison (all)) condition at midline was 7.238095 and the percentage achieving it for the Treatment (FFE + lit (all)) condition at midline was 5.84989. The difference for the Literacy Benchmark across the Control (Comparison (all)) and Treatment (FFE + lit (all)) conditions at midline was thus -1.388206 points (there was no baseline measurement for this variable). The p-value for this difference was 0.4108179. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the percentage achieving the Literacy Benchmark between the Control (Comparison (all)) and Treatment (FFE + lit (all)) conditions at midline beyond that which could result from sampling error.

3.2.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 97: Literacy Benchmark

Intervention	Midline	
	%	n
Control	7.889	469
Treatment	7.561	529

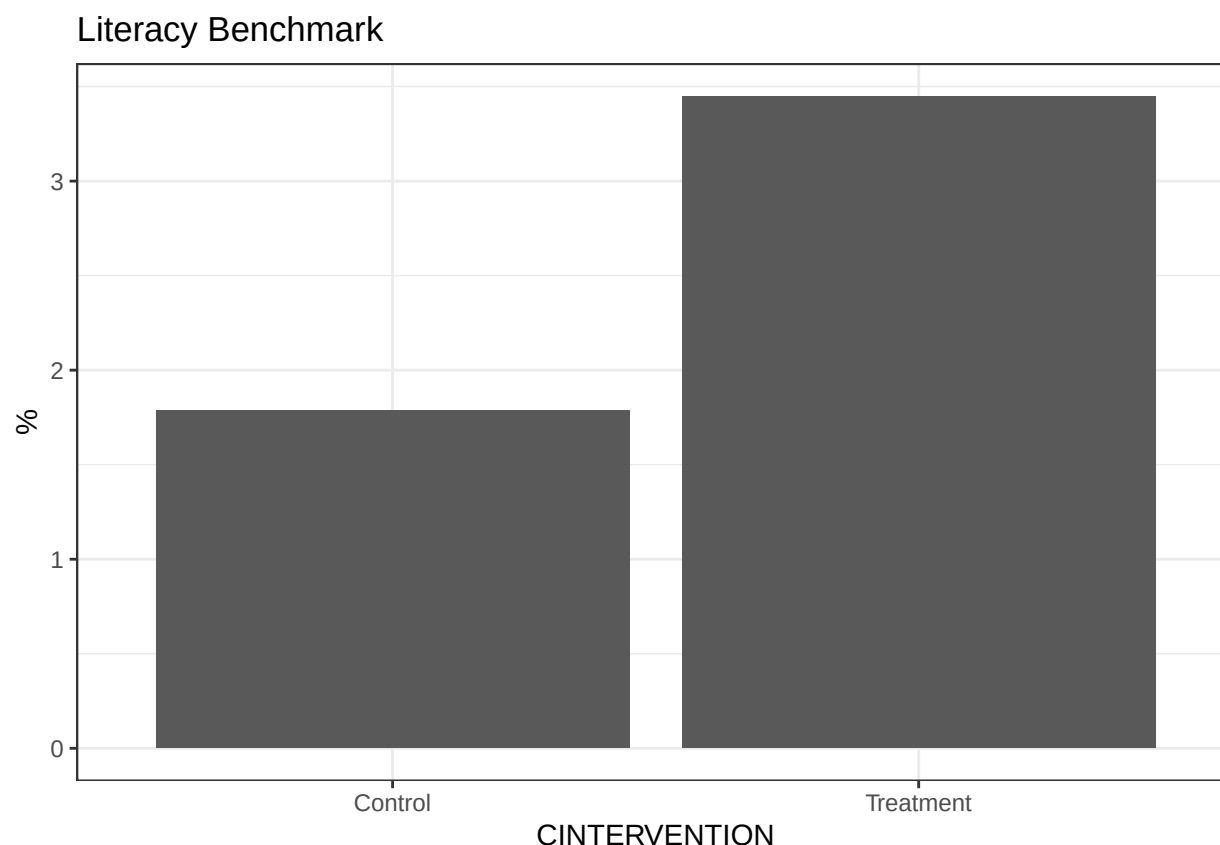


As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (Comparison (Portuguese)) condition at midline was 7.889126 and the percentage achieving it for the Treatment (FFE + lit (Portuguese)) condition at midline was 7.561437. The difference for the Literacy Benchmark across the Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline was thus -0.3276891 points (there was no baseline measurement for this variable). The p-value for this difference was 0.8773095. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the percentage achieving the Literacy Benchmark between the Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline beyond that which could result from sampling error.

3.2.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 98: Literacy Benchmark

Intervention	Midline	
	%	n
Control	1.786	56
Treatment	3.448	377

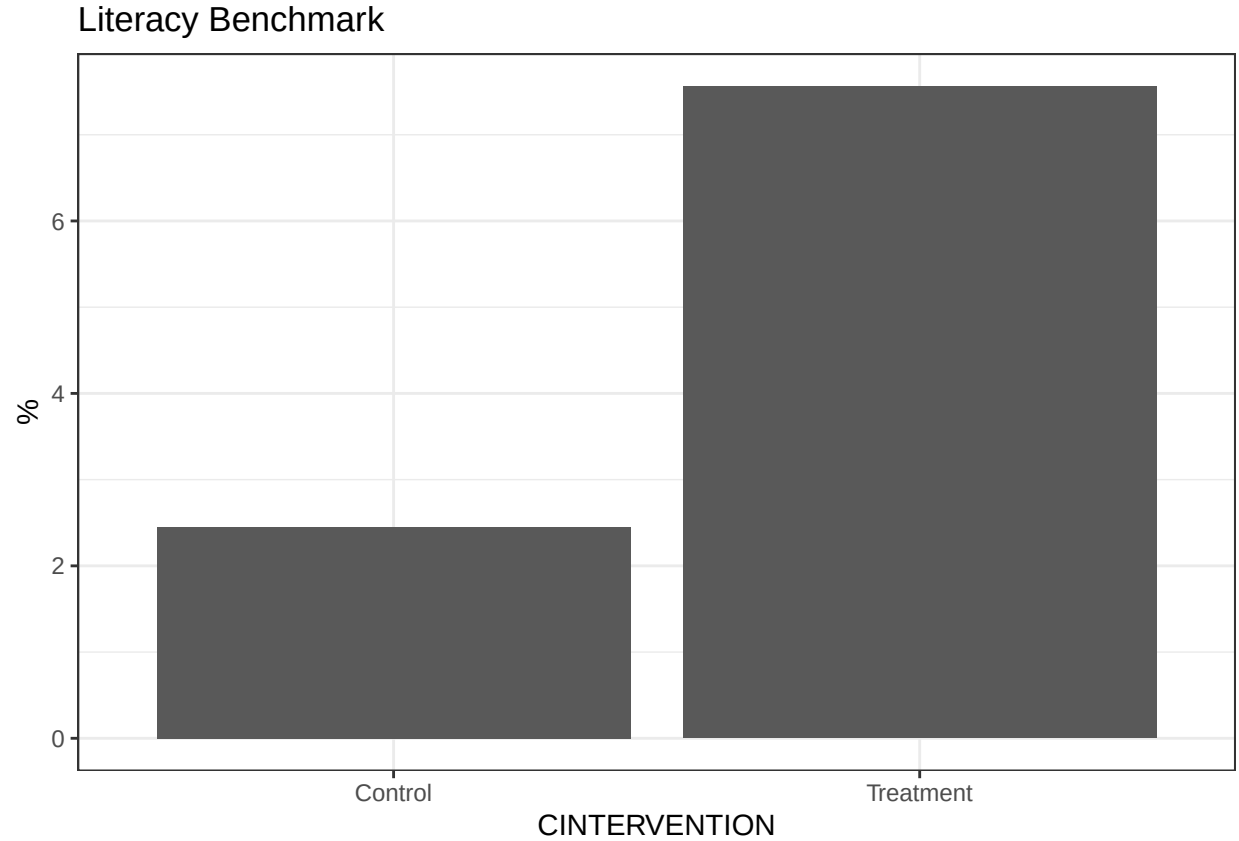


As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (Comparison (Bilingual)) condition at midline was 1.785714 and the percentage achieving it for the Treatment (FFE + lit (Bilingual)) condition at midline was 3.448276. The difference for the Literacy Benchmark across the Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) conditions at midline was thus 1.662562 points (there was no baseline measurement for this variable). The p-value for this difference was 0.5128991. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the percentage achieving the Literacy Benchmark between the Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) conditions at midline beyond that which could result from sampling error.

3.2.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 99: Literacy Benchmark

Intervention	Midline	
	%	n
Control	2.453	530
Treatment	7.561	529

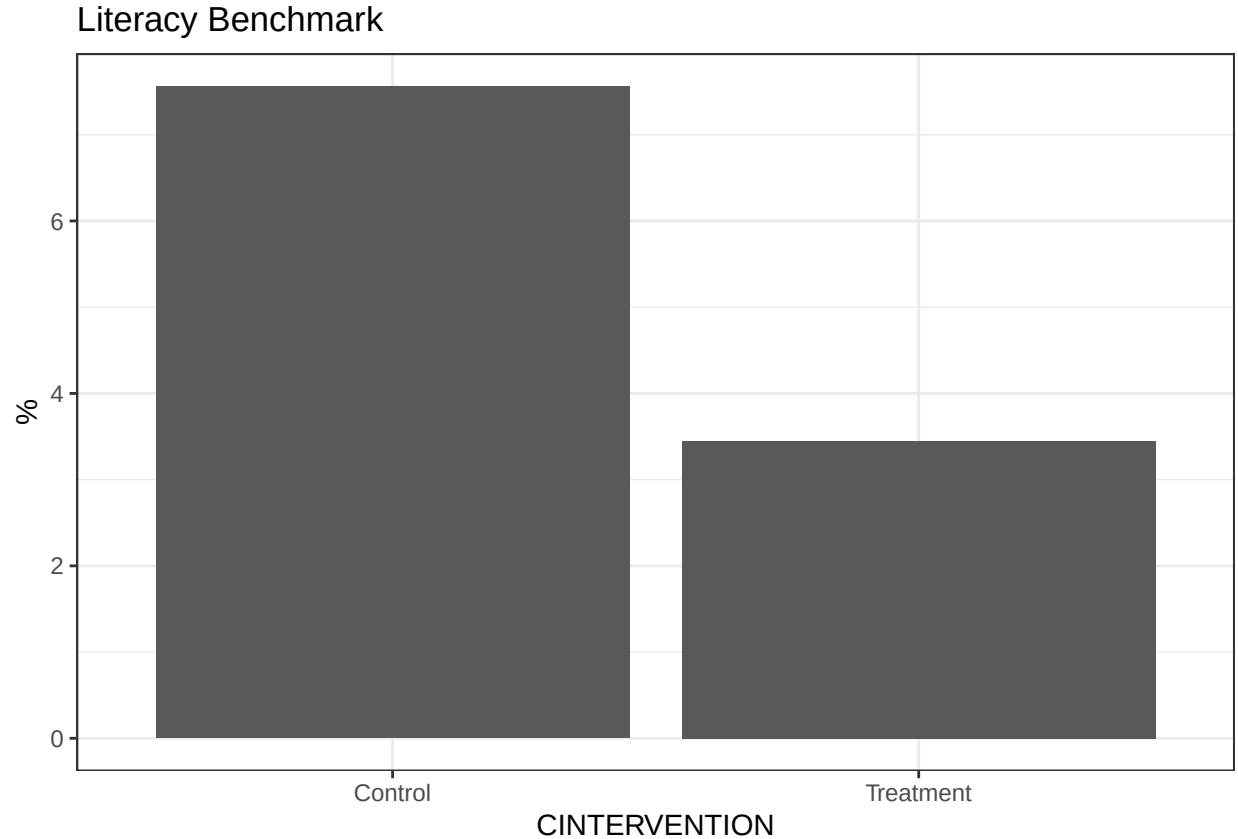


As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (FFE only (Portuguese)) condition at midline was 2.45283 and the percentage achieving it for the Treatment (FFE + lit (Portuguese)) condition at midline was 7.561437. The difference for the Literacy Benchmark across the Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline was thus 5.108606 points (there was no baseline measurement for this variable). The p-value for this difference was 0.004860343. Accordingly, the difference was statistically significant ($p < 0.01$ corrected for 4 additional contrasts). Nevertheless, due to the absence of baseline values and lack of random allocation to the treatment and control conditions, a causal attribution of this effect to the Treatment (FFE + lit (Portuguese)) is not possible

3.2.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 100: Literacy Benchmark

Intervention	Midline	
	%	n
Control	7.561	529
Treatment	3.448	377



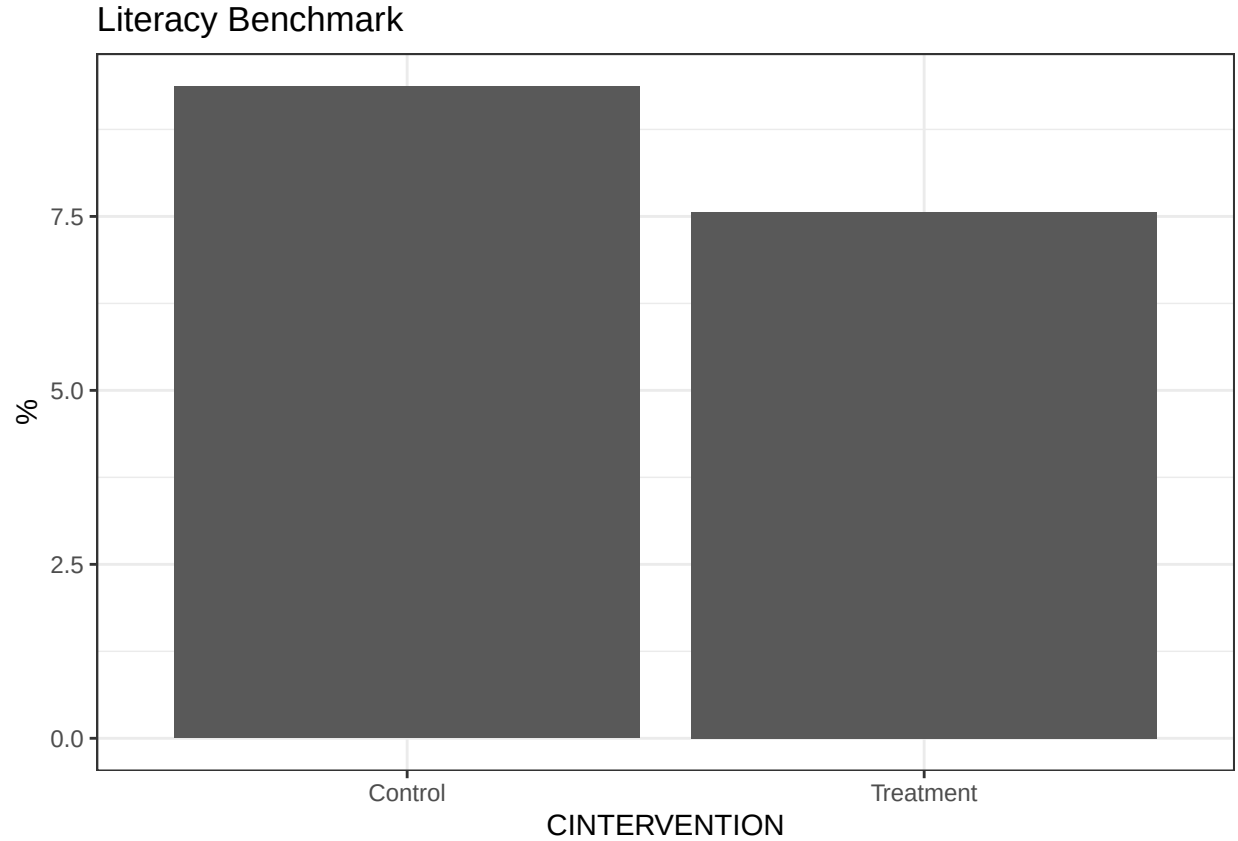
As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (FFE + lit (Portuguese)) condition at midline was 7.561437 and the percentage achieving it for the Treatment (FFE + lit (Bilingual)) condition at midline was 3.448276. The difference for the Literacy Benchmark across the Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) conditions at midline was thus -4.113161 points (there was no baseline measurement for this variable). The p-value for this difference was 0.04996474. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the percentage achieving the Literacy Benchmark between the Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) conditions at midline beyond that which could result from sampling error.

3.3 Literacy Benchmark: Girls only

3.3.1 Contrast 1: FFE + lit (all) vs Comparison (all)

Table 101: Literacy Benchmark

Intervention	Midline	
	%	n
Control	9.369	555
Treatment	7.562	886

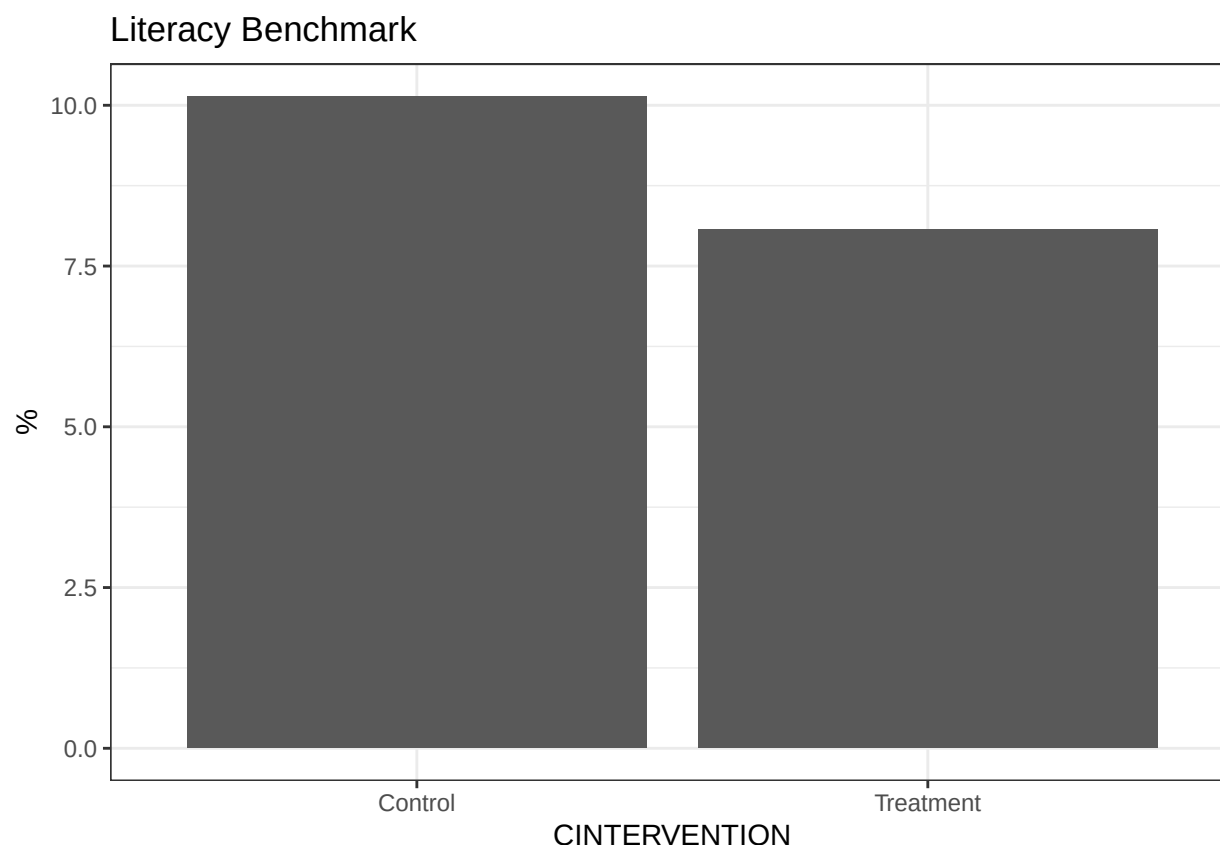


As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (Comparison (all)) condition at midline was 9.369369 and the percentage achieving it for the Treatment (FFE + lit (all)) condition at midline was 7.562077. The difference for the Literacy Benchmark across the Control (Comparison (all)) and Treatment (FFE + lit (all)) conditions at midline was thus -1.807293 points (there was no baseline measurement for this variable). The p-value for this difference was 0.3772846. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the percentage achieving the Literacy Benchmark between the Control (Comparison (all)) and Treatment (FFE + lit (all)) conditions at midline beyond that which could result from sampling error.

3.3.2 Contrast 2: FFE + lit (Portuguese) vs Comparison (Portuguese)

Table 102: Literacy Benchmark

Intervention	Midline	
	%	n
Control	10.139	503
Treatment	8.071	508

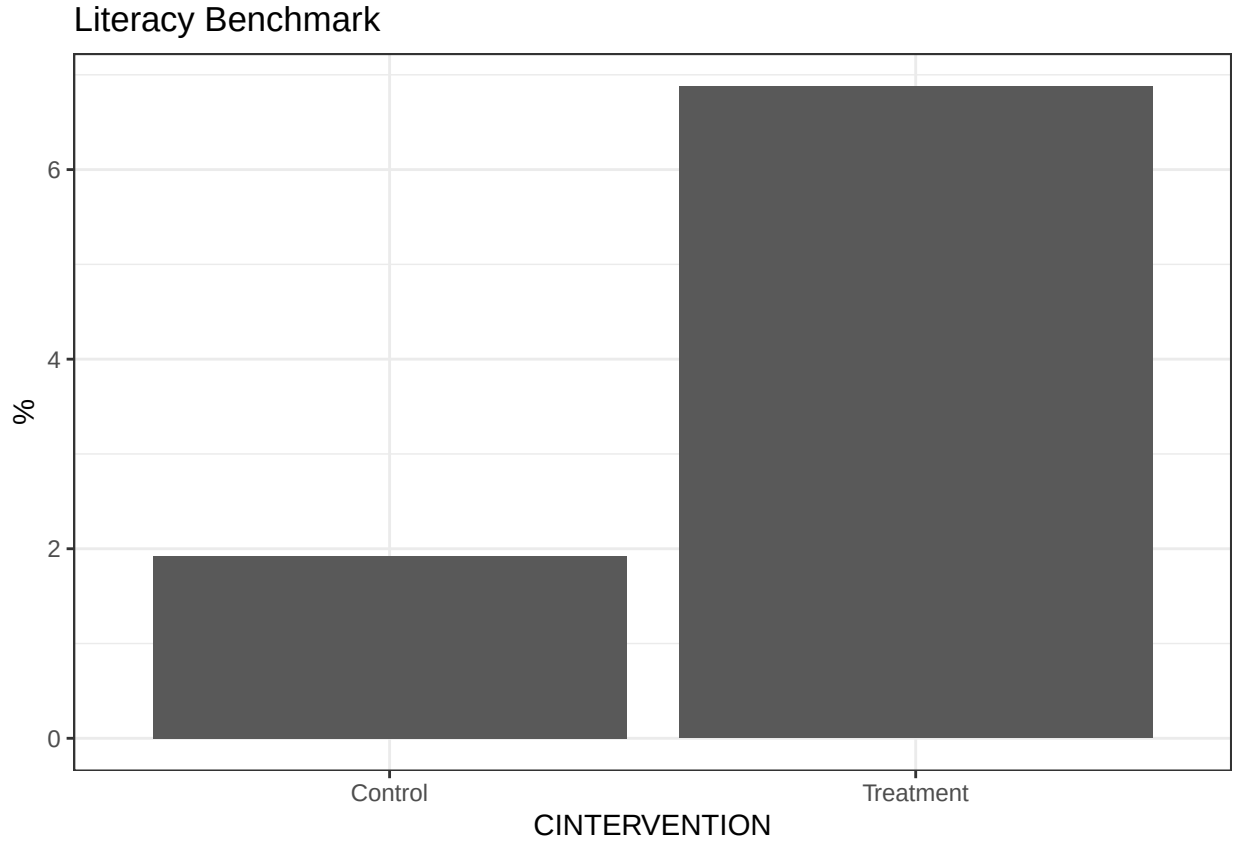


As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (Comparison (Portuguese)) condition at midline was 10.13917 and the percentage achieving it for the Treatment (FFE + lit (Portuguese)) condition at midline was 8.070866. The difference for the Literacy Benchmark across the Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline was thus -2.068299 points (there was no baseline measurement for this variable). The p-value for this difference was 0.3848931. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the percentage achieving the Literacy Benchmark between the Control (Comparison (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline beyond that which could result from sampling error.

3.3.3 Contrast 3: FFE + lit (Bilingual) vs Comparison (Bilingual)

Table 103: Literacy Benchmark

Intervention	Midline	
	%	n
Control	1.923	52
Treatment	6.878	378

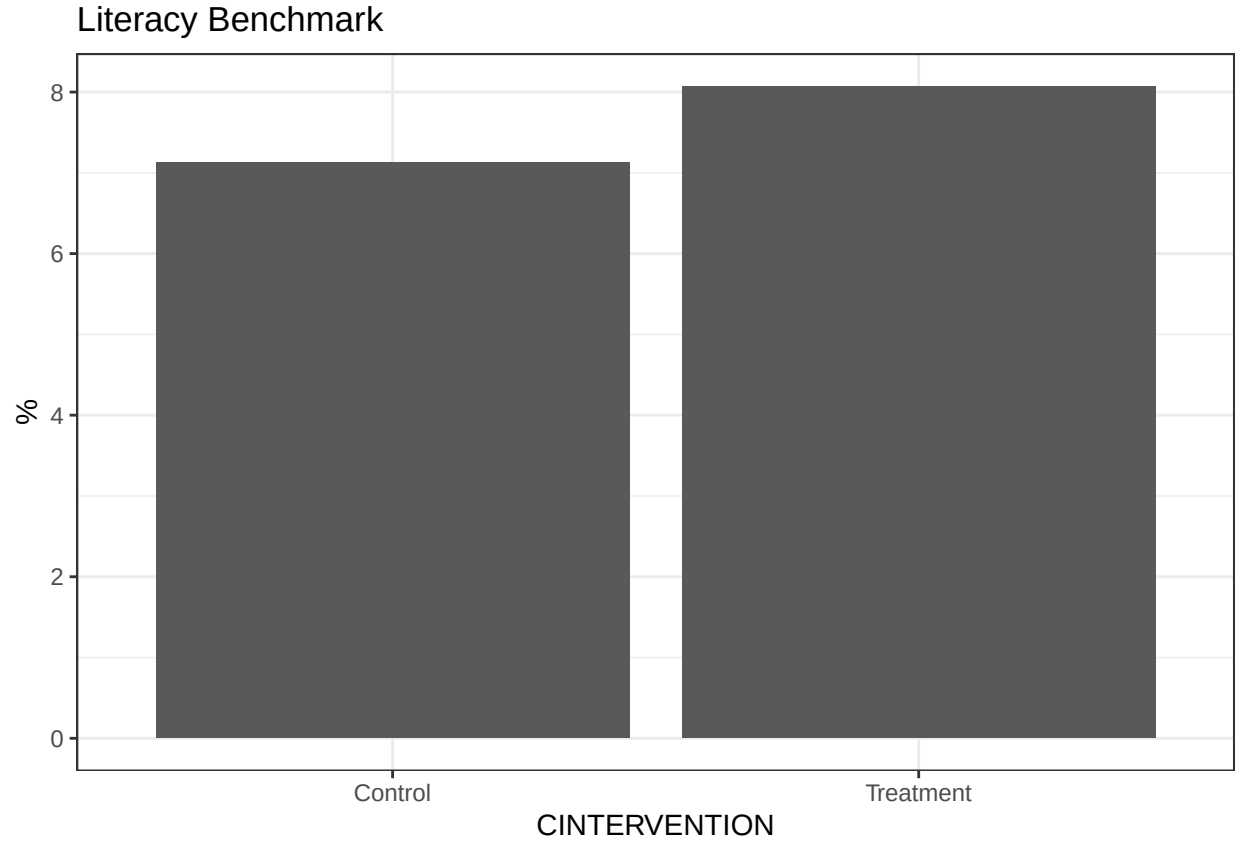


As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (Comparison (Bilingual)) condition at midline was 1.923077 and the percentage achieving it for the Treatment (FFE + lit (Bilingual)) condition at midline was 6.878307. The difference for the Literacy Benchmark across the Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) conditions at midline was thus 4.95523 points (there was no baseline measurement for this variable). The p-value for this difference was 0.1724789. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the percentage achieving the Literacy Benchmark between the Control (Comparison (Bilingual)) and Treatment (FFE + lit (Bilingual)) conditions at midline beyond that which could result from sampling error.

3.3.4 Contrast 4: FFE + lit (Portuguese) vs FFE only (Portuguese)

Table 104: Literacy Benchmark

Intervention	Midline	
	%	n
Control	7.126	421
Treatment	8.071	508

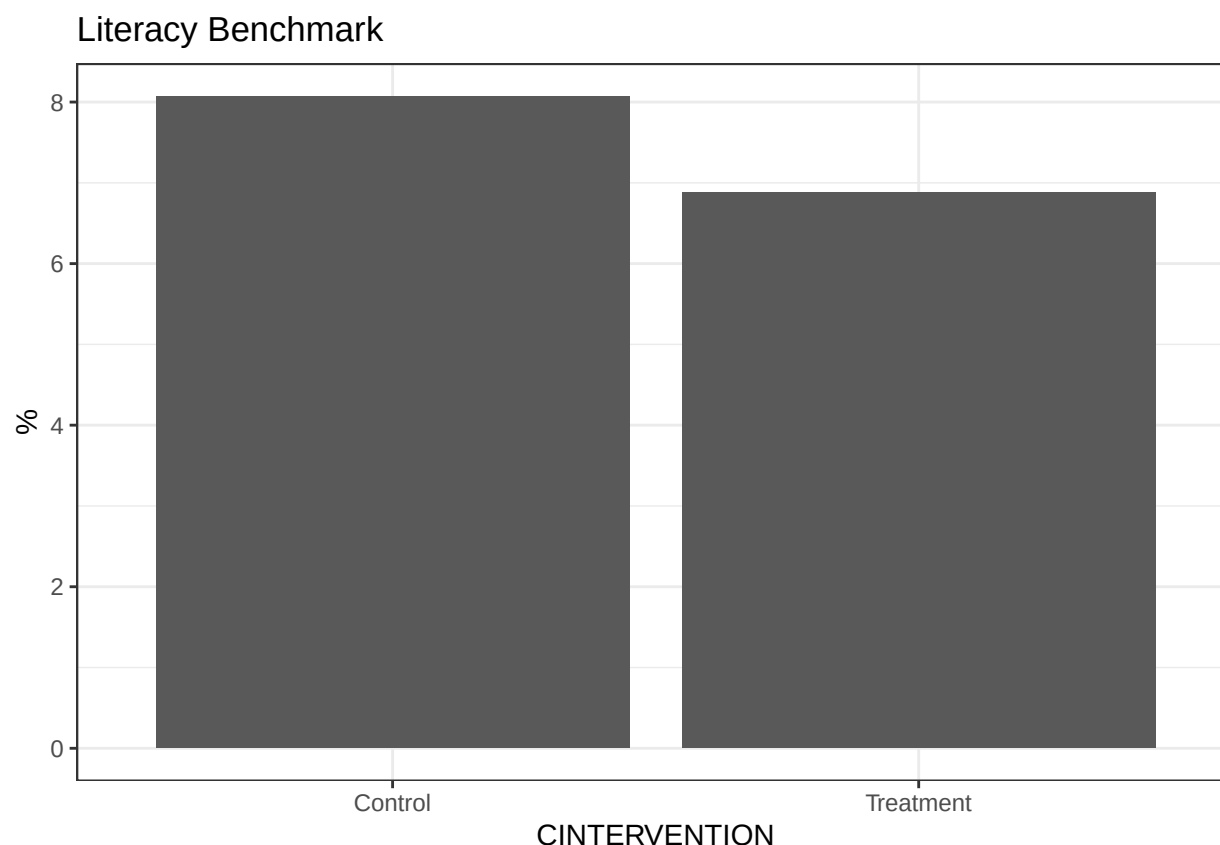


As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (FFE only (Portuguese)) condition at midline was 7.125891 and the percentage achieving it for the Treatment (FFE + lit (Portuguese)) condition at midline was 8.070866. The difference for the Literacy Benchmark across the Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline was thus 0.9449754 points (there was no baseline measurement for this variable). The p-value for this difference was 0.7122201. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the percentage achieving the Literacy Benchmark between the Control (FFE only (Portuguese)) and Treatment (FFE + lit (Portuguese)) conditions at midline beyond that which could result from sampling error.

3.3.5 Contrast 5: FFE + lit (Bilingual) vs FFE + lit (Portuguese)

Table 105: Literacy Benchmark

Intervention	Midline	
	%	n
Control	8.071	508
Treatment	6.878	378



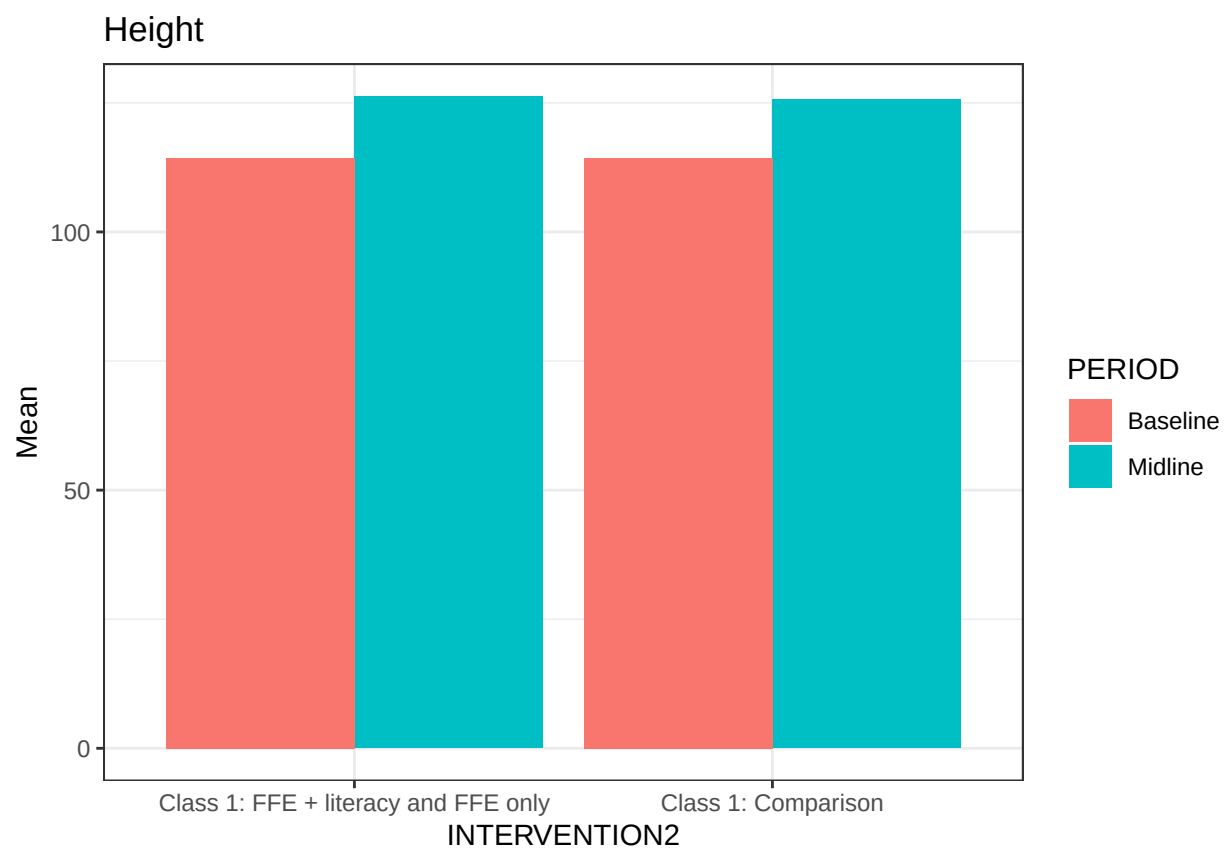
As shown in the table above, for the Literacy Benchmark, the percentage of students that achieved the benchmark for the Control (FFE + lit (Portuguese)) condition at midline was 8.070866 and the percentage achieving it for the Treatment (FFE + lit (Bilingual)) condition at midline was 6.878307. The difference for the Literacy Benchmark across the Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) conditions at midline was thus -1.192559 points (there was no baseline measurement for this variable). The p-value for this difference was 0.6024658. Accordingly, the difference was not statistically significant and therefore, the statistical test provides no conclusive evidence in regards to the there being a difference in the percentage achieving the Literacy Benchmark between the Control (FFE + lit (Portuguese)) and Treatment (FFE + lit (Bilingual)) conditions at midline beyond that which could result from sampling error.

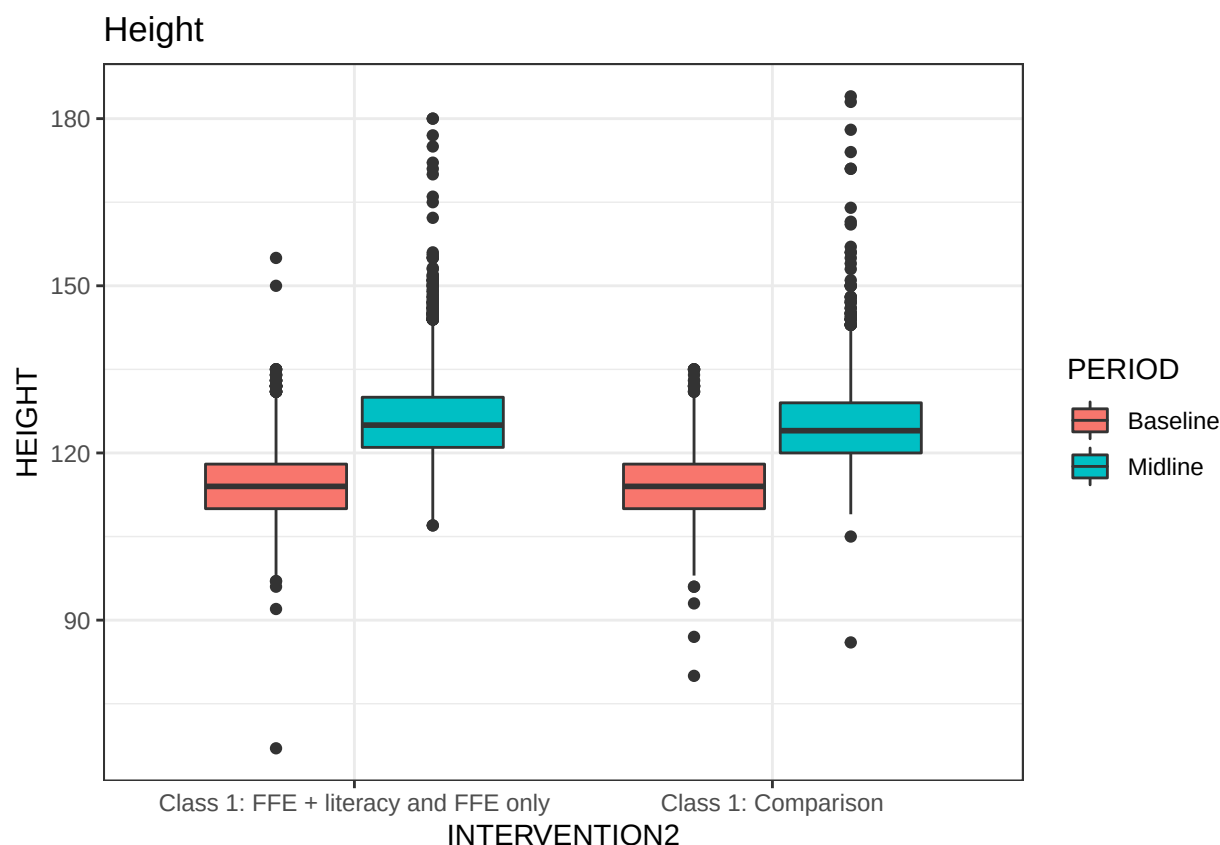
4 BMI

4.1 Height

Table 106: Height

Intervention	Baseline			Midline		
	Mean	SD	n	Mean	SD	n
Class 1: FFE + literacy and FFE only	114	6.62	2991	126	7.83	2873
Class 1: Comparison	114	6.29	1159	126	8.75	1081



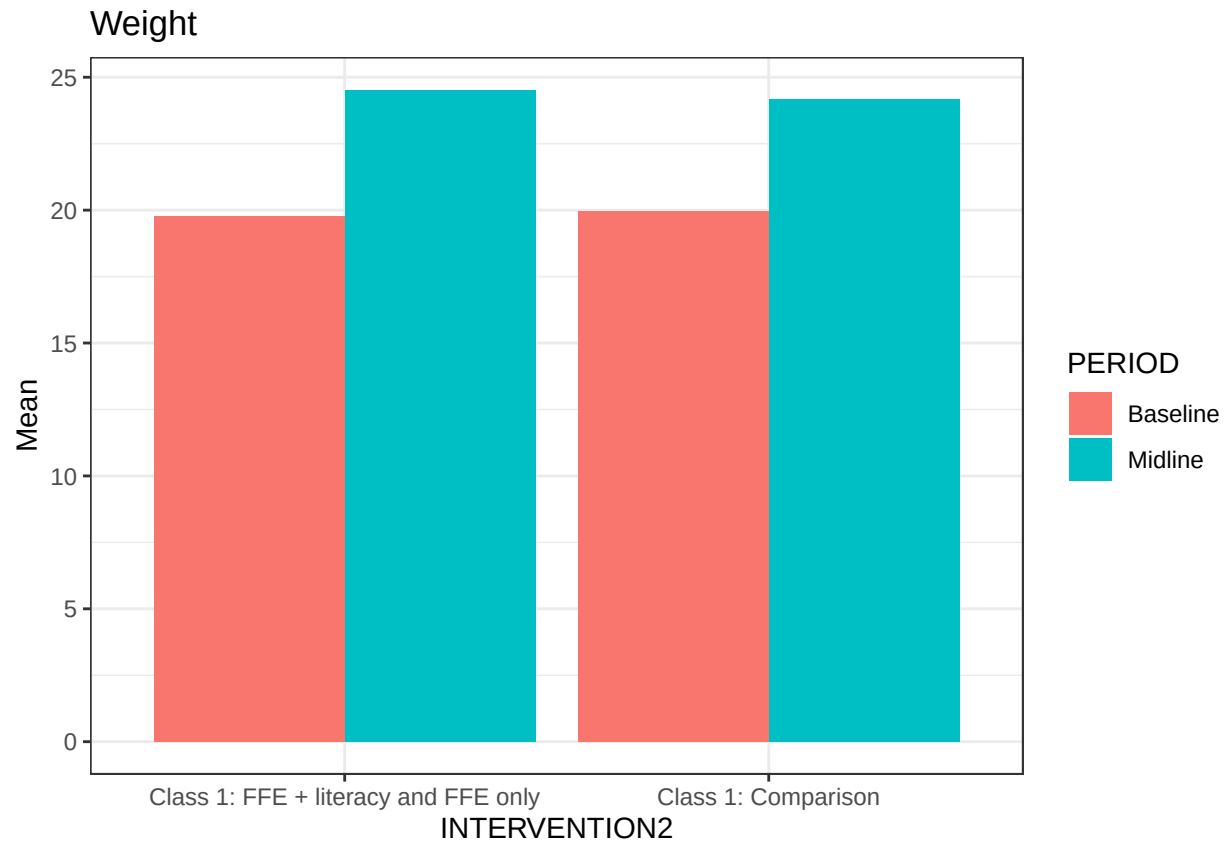


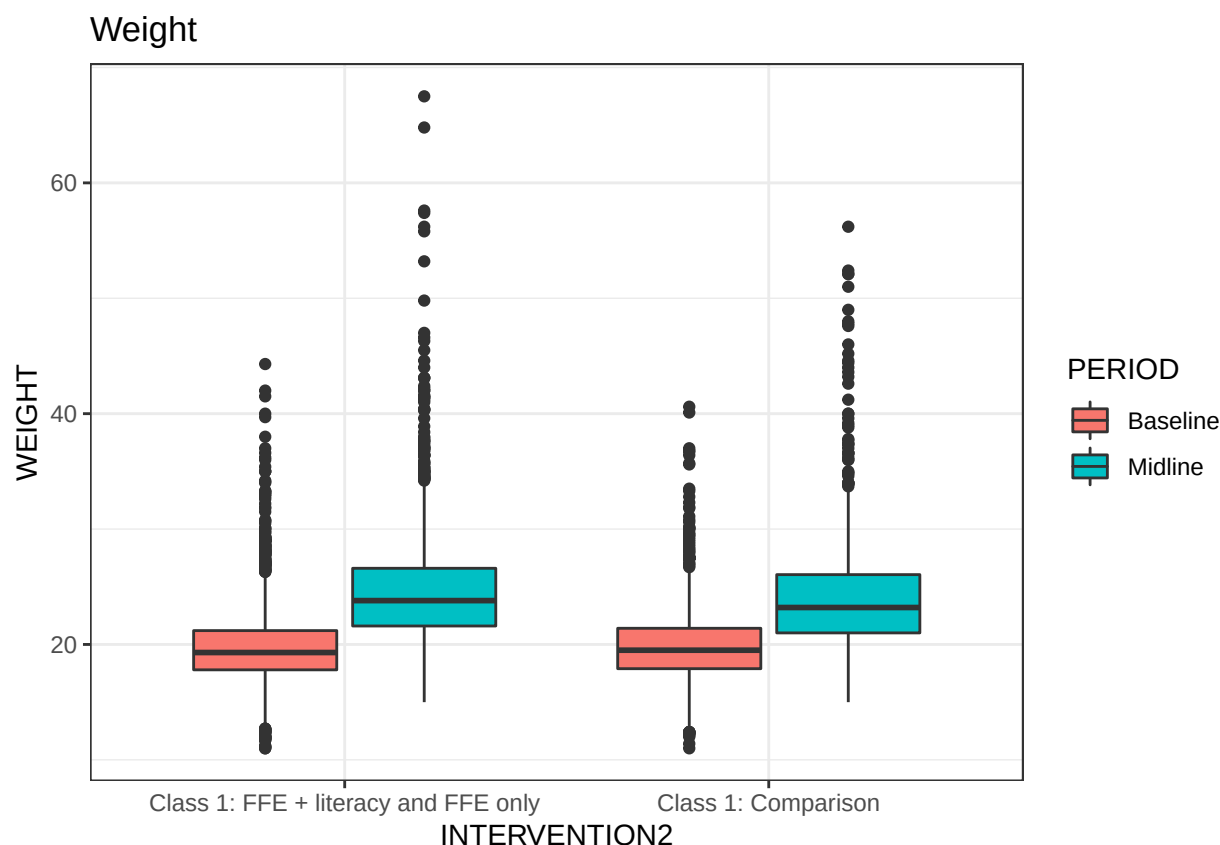
For the Height the FFE + Literacy and FFE only conditions have been collapsed into a single contrast to evaluate the impact of both of the intervention conditions with a nutritional component against the Comparison. As shown in the table above, for the the Height, the mean for the Comparison condition at baseline was 114 (SD = 6.29) and the mean for the FFE + Literacy and FFE only condition at baseline was 114 (SD = 6.62). The difference between the Comparison and FFE + Literacy and FFE only at baseline was thus of 0.0107 points. The p-value for this difference was 0.98. The mean for the Comparison condition at midline was 126 (SD = 8.75) and the mean for the FFE + Literacy condition and FFE only at midline was 126 (SD = 7.84). The difference between the Comparison and FFE + Literacy and FFE only at midline was thus of 0.586 points. The p-value for this difference was 0.187. The change from the baseline to the midline of 117 points for the Comparison condition establishes the counterfactual against which the change for the FFE + Literacy and FFE only can be assessed against. The change for the FFE + Literacy and FFE only from the baseline to the midline was of 11.9 points. Consequently, the change for the Comparison relative to the change for the FFE + Literacy and FFE only condition was of 0.575 points. The p-value for this difference was 0.288. Accordingly, the difference was not statistically significant ($p < 0.017$ corrected for 3 additional comparisons) and therefore, the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Height across the Comparison and the FFE + Literacy and FFE only conditions beyond that expected from sampling error. All statistical tests accounted for the schools' intraclass correlation effect on the outcome.

4.2 Weight

Table 107: Weight

Intervention	Baseline			Midline		
	Mean	SD	n	Mean	SD	n
Class 1: FFE + literacy and FFE only	19.8	3.39	2991	24.5	4.52	2873
Class 1: Comparison	20.0	3.62	1159	24.2	5.10	1083



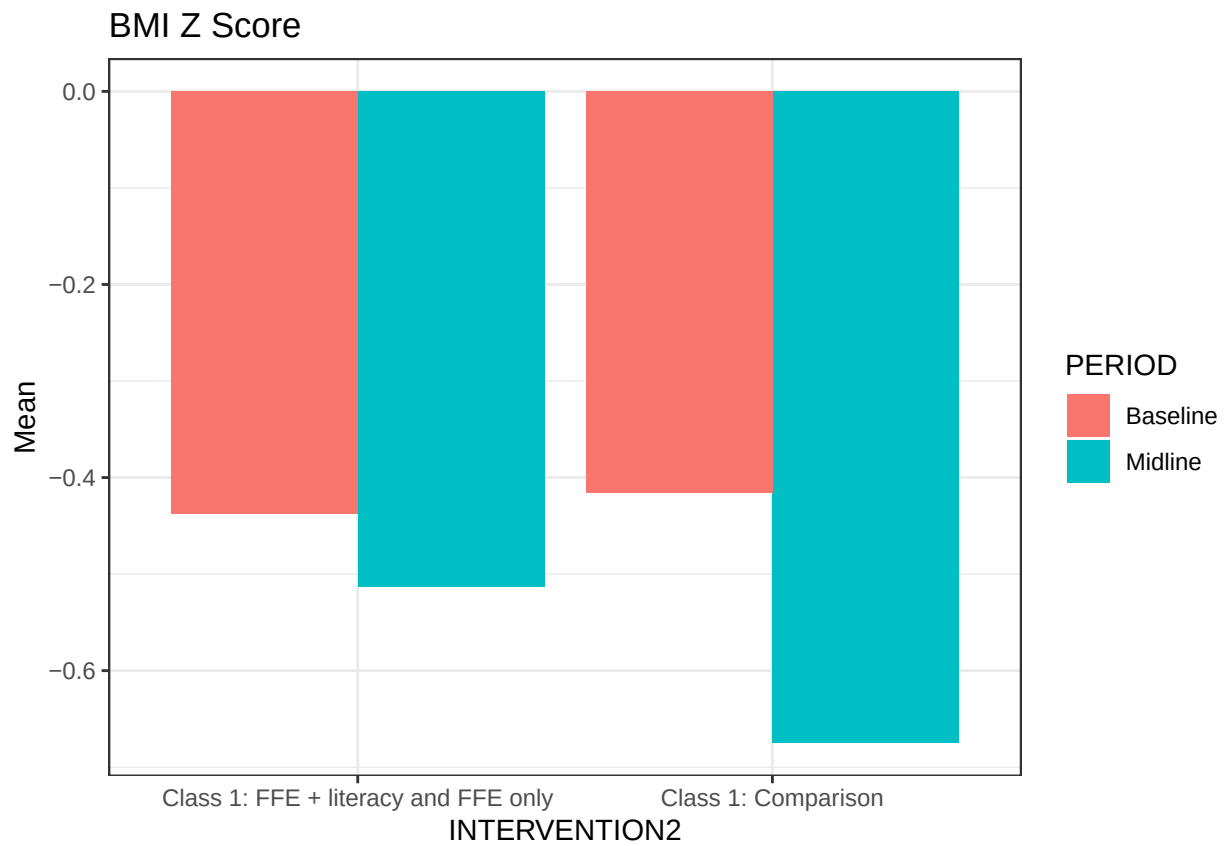


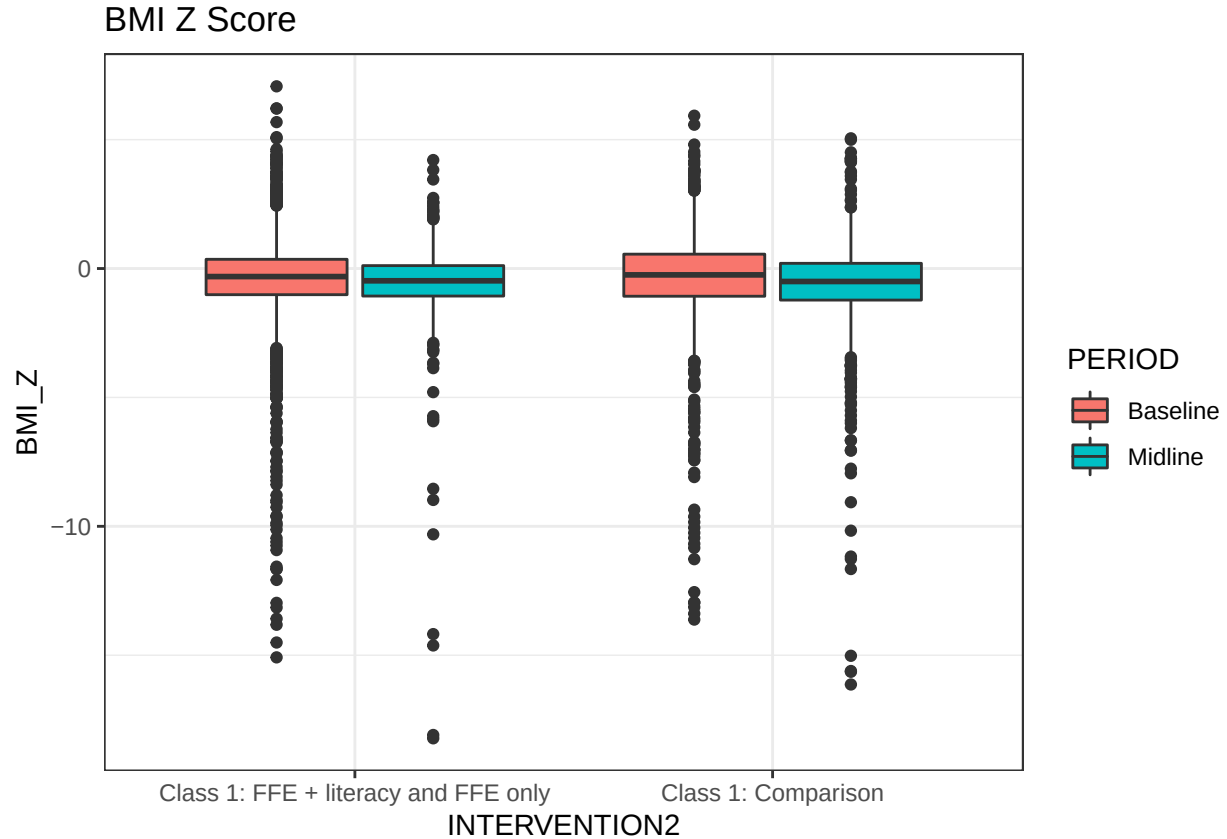
For the Weight the FFE + Literacy and FFE only conditions have been collapsed into a single contrast to evaluate the impact of both of the intervention conditions with a nutritional component against the Comparison. As shown in the table above, for the the Weight, the mean for the Comparison condition at baseline was 20 (SD = 3.62) and the mean for the FFE + Literacy and FFE only condition at baseline was 19.8 (SD = 3.39). The difference between the Comparison and FFE + Literacy and FFE only at baseline was thus of 0.194 points. The p-value for this difference was 0.37. The mean for the Comparison condition at midline was 24.2 (SD = 5.1) and the mean for the FFE + Literacy condition and FFE only at midline was 24.5 (SD = 4.52). The difference between the Comparison and FFE + Literacy and FFE only at midline was thus of 0.343 points. The p-value for this difference was 0.17. The change from the baseline to the midline of 19.1 points for the Comparison condition establishes the counterfactual against which the change for the FFE + Literacy and FFE only can be assessed against. The change for the FFE + Literacy and FFE only from the baseline to the midline was of 4.76 points. Consequently, the change for the Comparison relative to the change for the FFE + Literacy and FFE only condition was of 0.537 points. The p-value for this difference was 0.0861. Accordingly, the difference was not statistically significant ($p < 0.017$ corrected for 3 additional comparisons) and therefore, the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Weight across the Comparison and the FFE + Literacy and FFE only conditions beyond that expected from sampling error. All statistical tests accounted for the schools' intraclass correlation effect on the outcome.

4.3 BMI Z Score

Table 108: BMI Z Score

Intervention	Baseline			Midline		
	Mean	SD	n	Mean	SD	n
Class 1: FFE + literacy and FFE only	-0.437	1.72	2951	-0.514	1.24	1891
Class 1: Comparison	-0.416	2.15	1150	-0.675	1.92	965





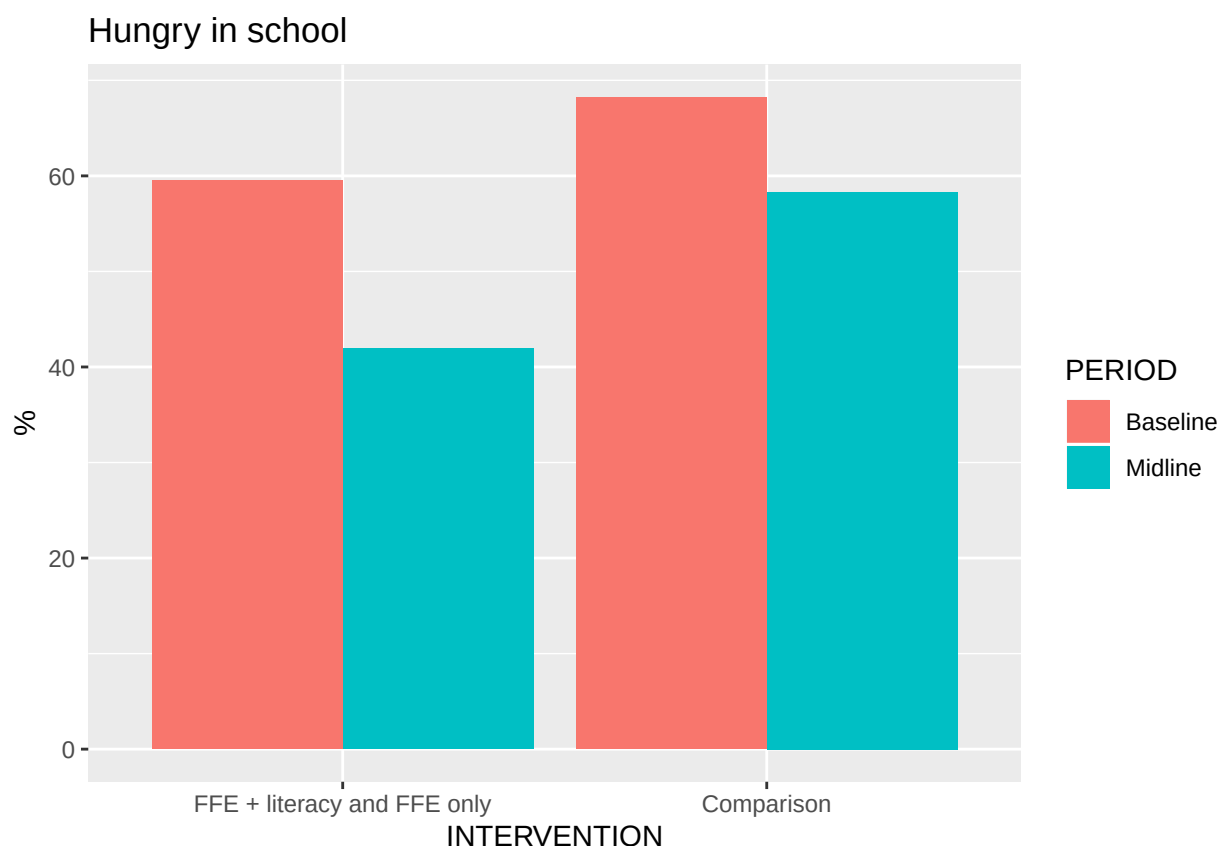
For the BMI Z Score the FFE + Literacy and FFE only conditions have been collapsed into a single contrast to evaluate the impact of both of the intervention conditions with a nutritional component against the Comparison. As shown in the table above, for the the BMI Z Score, the mean for the Comparison condition at baseline was -0.416 (SD = 2.15) and the mean for the FFE + Literacy and FFE only condition at baseline was -0.437 (SD = 1.72). The difference between the Comparison and FFE + Literacy and FFE only at baseline was thus of 0.0214 points. The p-value for this difference was 0.899. The mean for the Comparison condition at midline was -0.675 (SD = 1.92) and the mean for the FFE + Literacy condition and FFE only at midline was -0.514 (SD = 1.24). The difference between the Comparison and FFE + Literacy and FFE only at midline was thus of 0.162 points. The p-value for this difference was 0.164. The change from the baseline to the midline of -2.59 points for the Comparison condition establishes the counterfactual against which the change for the FFE + Literacy and FFE only can be assessed against. The change for the FFE + Literacy and FFE only from the baseline to the midline was of -0.0763 points. Consequently, the change for the Comparison relative to the change for the FFE + Literacy and FFE only condition was of 0.183 points. The p-value for this difference was 0.367. Accordingly, the difference was not statistically significant ($p < 0.017$ corrected for 3 additional comparisons) and therefore, the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the BMI Z Score across the Comparison and the FFE + Literacy and FFE only conditions beyond that expected from sampling error. All statistical tests accounted for the schools' intraclass correlation effect on the outcome.

5 Questions regarding hunger, attention and hygiene

5.1 HUNGER: Hungry in school

Table 109: Hungry in school

Intervention	Baseline			Midline		
	n	%	SD	n	%	SD
FFE + literacy and FFE only	2956	59.6	9.03	2872	41.9	9.21
Comparison	1121	68.2	13.90	1083	58.3	14.98



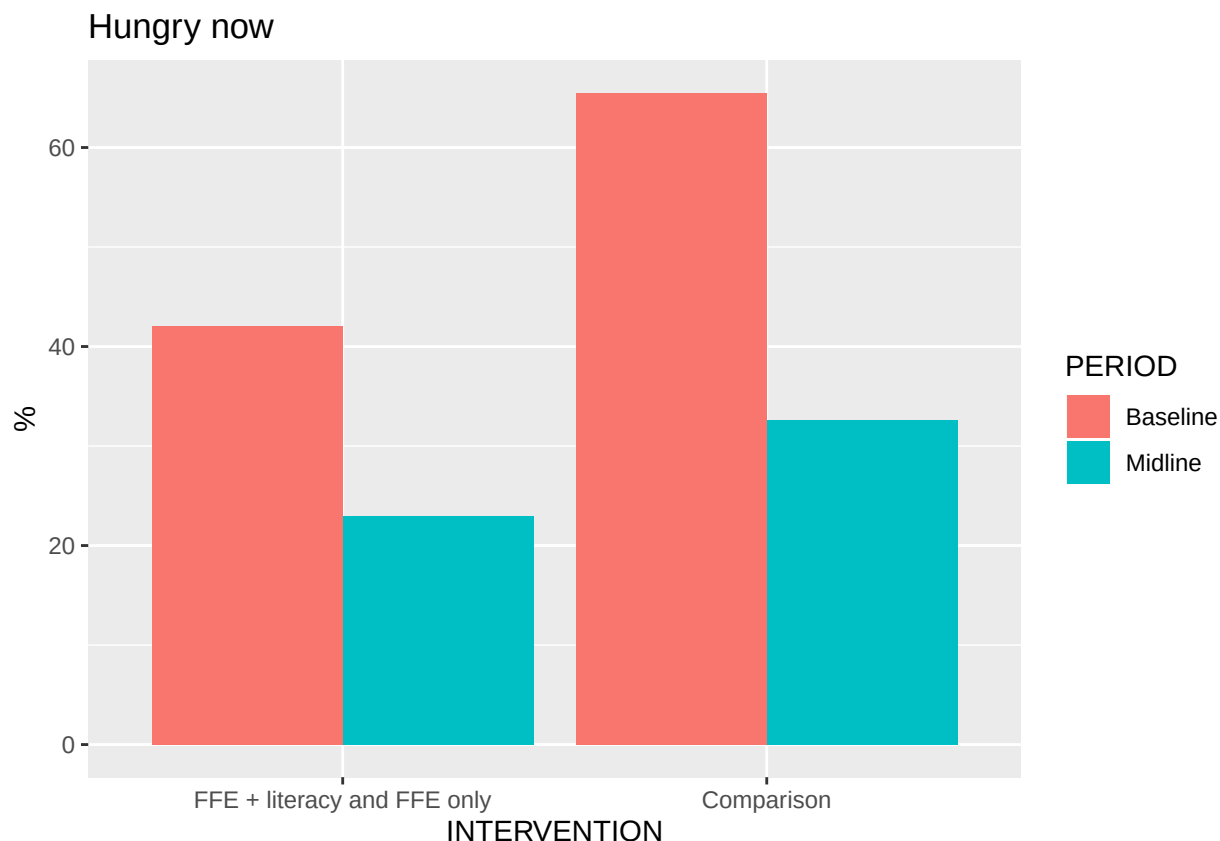
For the Hungry in school significance test, the FFE + Literacy and FFE only have been combined and contrasted against the Comparison so that the impact of the interventions with a food and hygiene component could be compared against the Comparison which received neither. As shown in the table above, for the the Hungry in school, the percentage of success for the Comparison condition at baseline was 68.2 (SD = 13.9) and the percentage for the FFE + Literacy and FFE only condition at baseline was 59.6 (SD = 9.03). The difference between Comparison and FFE + Literacy and FFE only at baseline was thus of 8.69 percentage points. The p-value for this difference was 0.202. The percentage for the Comparison condition at midline was 58.3 (SD = 15) and the percentage for the FFE + Literacy and FFE only condition at midline was 41.9 (SD = 9.21). The difference between Comparison and FFE + Literacy and FFE only at midline was thus of 16.4 points. The p-value for this difference was 1.53e-13. The change in percentage of success for the Comparison condition from baseline to midline was of -9.91 points and establishes the counterfactual against which the change for the FFE + Literacy and FFE only is assessed. The percentage change for the FFE + Literacy and FFE only condition from baseline to midline was of -17.6 points. Consequently, the change for the Comparison relative to the change for the FFE + Literacy and FFE only condition

amounted to -7.74 percentage points. The p-value for this difference was 0.354. Accordingly, the difference was not statistically significant and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Hungry in school outcome across the Comparison and the FFE + Literacy and FFE only conditions beyond that expected from sampling error. All statistical tests accounted for the schools' intraclass correlation effect on the outcome.

5.2 HUNGERNOW: Hungry now

Table 110: Hungry now

Intervention	Baseline			Midline		
	n	%	SD	n	%	SD
FFE + literacy and FFE only	2956	42.1	9.08	2872	22.9	7.84
Comparison	1121	65.5	14.20	1083	32.6	14.25



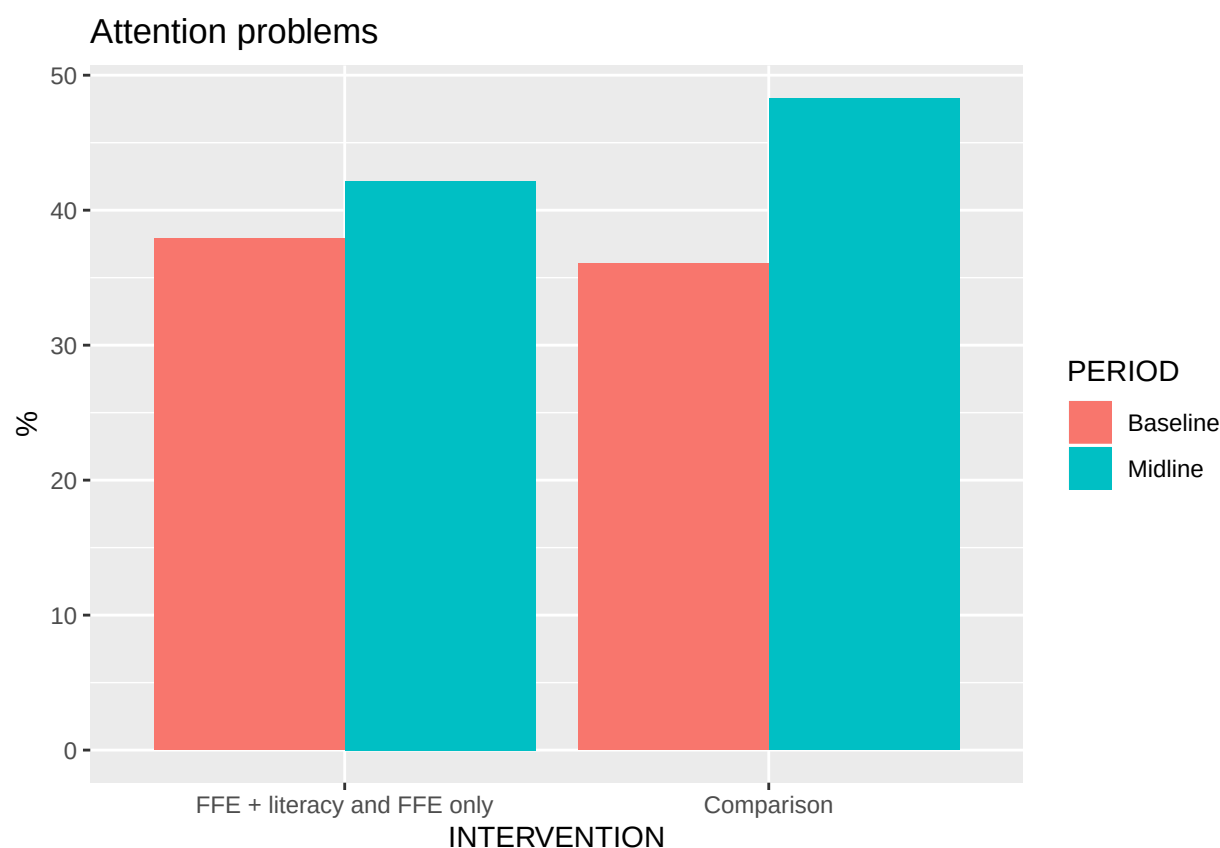
For the Hungry now significance test, the FFE + Literacy and FFE only have been combined and contrasted against the Comparison so that the impact of the interventions with a food and hygiene component could be compared against the Comparison which received neither. As shown in the table above, for the the Hungry now, the percentage of success for the Comparison condition at baseline was 65.5 (SD = 14.2) and the percentage for the FFE + Literacy and FFE only condition at baseline was 42.1 (SD = 9.08). The difference between Comparison and FFE + Literacy and FFE only at baseline was thus of 23.4 percentage points. The p-value for this difference was 0.00057. The percentage for the Comparison condition at midline was 32.6 (SD = 14.2) and the percentage for the FFE + Literacy and FFE only condition at midline was 22.9 (SD = 7.84). The difference between Comparison and FFE + Literacy and FFE only at midline was thus of 9.7 points. The p-value for this difference was 0.00000241. The change in percentage of success for the

Comparison condition from baseline to midline was of -32.9 points and establishes the counterfactual against which the change for the FFE + Literacy and FFE only is assessed. The percentage change for the FFE + Literacy and FFE only condition from baseline to midline was of -19.2 points. Consequently, the change for the Comparison relative to the change for the FFE + Literacy and FFE only condition amounted to 13.7 percentage points. The p-value for this difference was 0.111. Accordingly, the difference was not statistically significant and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Hungry now outcome across the Comparison and the FFE + Literacy and FFE only conditions beyond that expected from sampling error. All statistical tests accounted for the schools' intraclass correlation effect on the outcome.

5.3 ATTENTION: Attention problems

Table 111: Attention problems

Intervention	Baseline			Midline		
	n	%	SD	n	%	SD
FFE + literacy and FFE only	2956	37.9	8.92	2872	42.2	9.21
Comparison	1121	36.1	14.34	1083	48.3	15.19



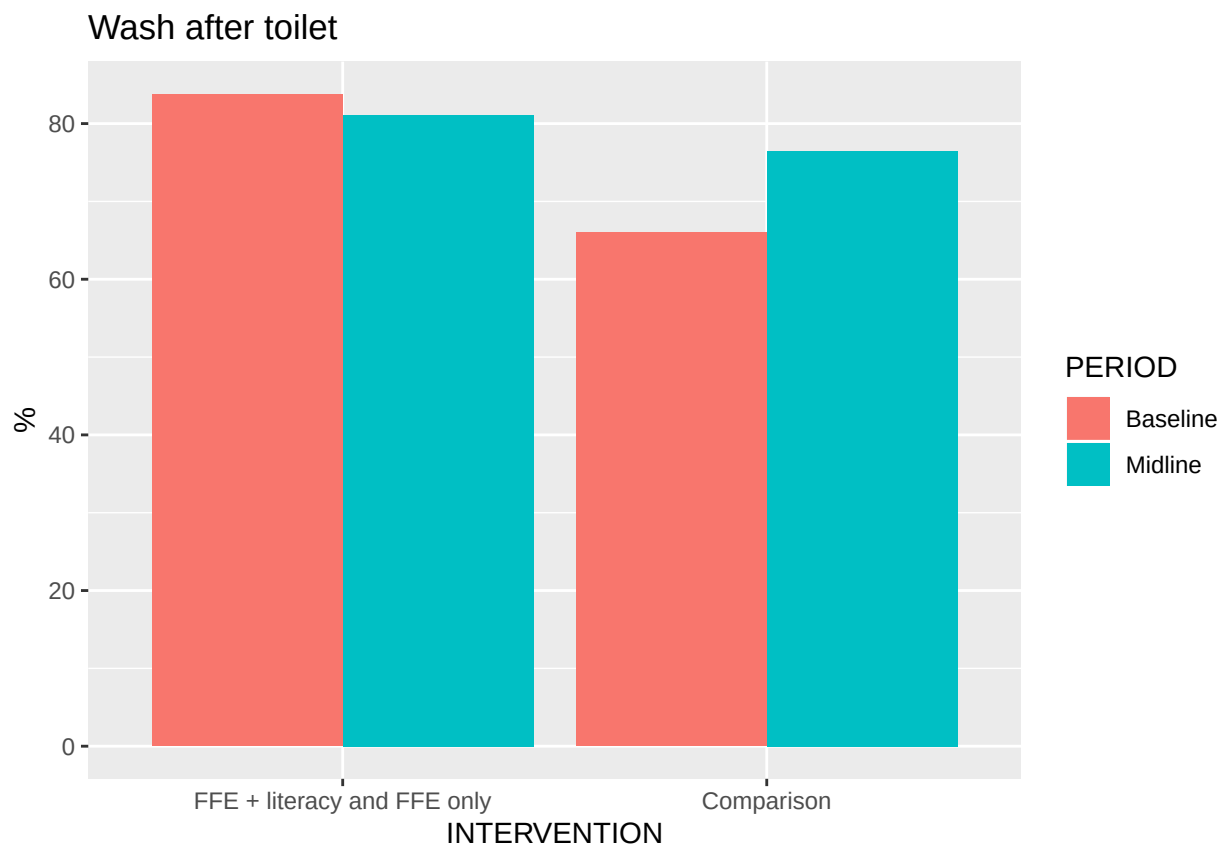
For the Attention problems significance test, the FFE + Literacy and FFE only have been combined and contrasted against the Comparison so that the impact of the interventions with a food and hygiene component could be compared against the Comparison which received neither. As shown in the table above, for the the Attention problems, the percentage of success for the Comparison condition at baseline was 36.1 (SD = 14.3) and the percentage for the FFE + Literacy and FFE only condition at baseline was 37.9 (SD = 8.92). The difference between Comparison and FFE + Literacy and FFE only at baseline was thus of 1.84

percentage points. The p-value for this difference was 0.651. The percentage for the Comparison condition at midline was 48.3 (SD = 15.2) and the percentage for the FFE + Literacy and FFE only condition at midline was 42.2 (SD = 9.21). The difference between Comparison and FFE + Literacy and FFE only at midline was thus of 6.15 points. The p-value for this difference was 0.0263. The change in percentage of success for the Comparison condition from baseline to midline was of 12.2 points and establishes the counterfactual against which the change for the FFE + Literacy and FFE only is assessed. The percentage change for the FFE + Literacy and FFE only condition from baseline to midline was of 4.25 points. Consequently, the change for the Comparison relative to the change for the FFE + Literacy and FFE only condition amounted to -7.99 percentage points. The p-value for this difference was 0.113. Accordingly, the difference was not statistically significant and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Attention problems outcome across the Comparison and the FFE + Literacy and FFE only conditions beyond that expected from sampling error. All statistical tests accounted for the schools' intraclass correlation effect on the outcome.

5.4 TOILET: Wash after toilet

Table 112: Wash after toilet

Intervention	Baseline			Midline		
	n	%	SD	n	%	SD
FFE + literacy and FFE only	2956	83.8	6.78	2872	81.1	7.3
Comparison	1121	66.0	14.15	1083	76.5	12.9



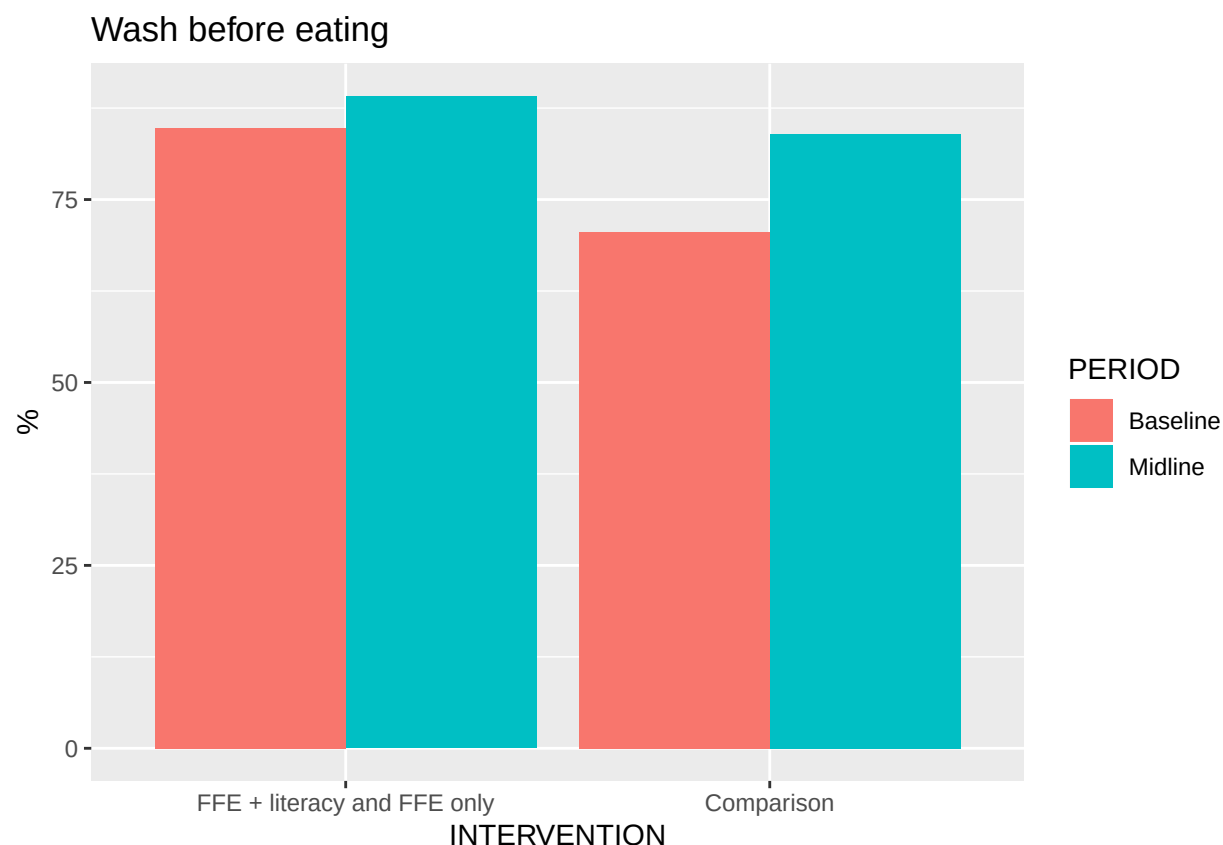
For the Wash after toilet significance test, the FFE + Literacy and FFE only have been combined and contrasted against the Comparison so that the impact of the interventions with a food and hygiene component

could be compared against the Comparison which received neither. As shown in the table above, for the the Wash after toilet, the percentage of success for the Comparison condition at baseline was 66 (SD = 14.2) and the percentage for the FFE + Literacy and FFE only condition at baseline was 83.8 (SD = 6.78). The difference between Comparison and FFE + Literacy and FFE only at baseline was thus of 17.8 percentage points. The p-value for this difference was 1.41e-08. The percentage for the Comparison condition at midline was 76.5 (SD = 12.9) and the percentage for the FFE + Literacy and FFE only condition at midline was 81.1 (SD = 7.3). The difference between Comparison and FFE + Literacy and FFE only at midline was thus of 4.65 points. The p-value for this difference was 0.028. The change in percentage of success for the Comparison condition from baseline to midline was of 10.5 points and establishes the counterfactual against which the change for the FFE + Literacy and FFE only is assessed. The percentage change for the FFE + Literacy and FFE only condition from baseline to midline was of -2.65 points. Consequently, the change for the Comparison relative to the change for the FFE + Literacy and FFE only condition amounted to -13.1 percentage points. The p-value for this difference was 0.000602. Accordingly, the difference was statistically significant and thus there is evidence of a different degree of change from baseline to midline in the FFE + Literacy and FFE only when compared to the Comparison condition. This provides evidence that the intervention (or some other unobserved process) impacted on the Wash after toilet outcome and thus the rate of change of the two conditions relative to their baselines differed. All statistical tests accounted for the schools' intraclass correlation effect on the outcome.

5.5 EATING: Wash before eating

Table 113: Wash before eating

Intervention	Baseline			Midline		
	n	%	SD	n	%	SD
FFE + literacy and FFE only	2956	84.8	6.61	2872	89.1	5.81
Comparison	1121	70.6	13.61	1083	83.9	11.16



For the Wash before eating significance test, the FFE + Literacy and FFE only have been combined and contrasted against the Comparison so that the impact of the interventions with a food and hygiene component could be compared against the Comparison which received neither. As shown in the table above, for the the Wash before eating, the percentage of success for the Comparison condition at baseline was 70.6 (SD = 13.6) and the percentage for the FFE + Literacy and FFE only condition at baseline was 84.8 (SD = 6.61). The difference between Comparison and FFE + Literacy and FFE only at baseline was thus of 14.2 percentage points. The p-value for this difference was 0.00000499. The percentage for the Comparison condition at midline was 83.9 (SD = 11.2) and the percentage for the FFE + Literacy and FFE only condition at midline was 89.1 (SD = 5.81). The difference between Comparison and FFE + Literacy and FFE only at midline was thus of 5.17 points. The p-value for this difference was 0.00172. The change in percentage of success for the Comparison condition from baseline to midline was of 13.3 points and establishes the counterfactual against which the change for the FFE + Literacy and FFE only is assessed. The percentage change for the FFE + Literacy and FFE only condition from baseline to midline was of 4.35 points. Consequently, the change for the Comparison relative to the change for the FFE + Literacy and FFE only condition amounted to -8.99 percentage points. The p-value for this difference was 0.0793. Accordingly, the difference was not statistically significant and therefore the statistical test provides no conclusive evidence in regards to there being a difference in the rate of change between the baseline and the midline for the Wash before eating outcome across the Comparison and the FFE + Literacy and FFE only conditions beyond that expected from sampling error. All statistical tests accounted for the schools' intraclass correlation effect on the outcome.

6 Teacher questions regarding hunger, attention and difficulty in learning

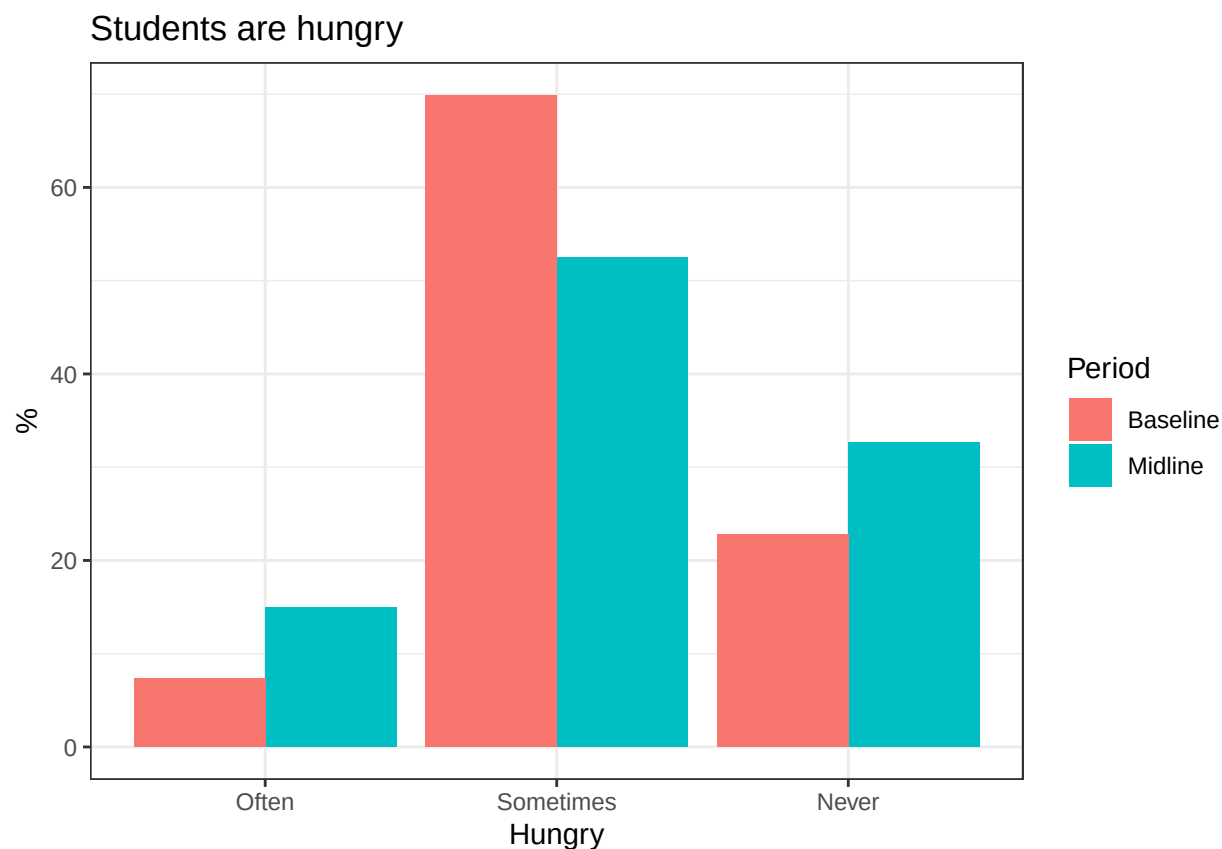
6.1 Students are hungry

Table 114: Students are hungry

	Baseline	Midline
Often	7.3	14.9
Sometimes	69.9	52.5
Never	22.8	32.6

Table 115: Students are hungry

	Baseline	Midline
Often	33	27
Sometimes	316	95
Never	103	59



As can be seen in the table above, in regards to pupils being hungry at baseline, 22.8 of the teacher responded 'Never', 69.9 responded 'Sometimes' and 7.3 responded 'Often.' At midline, 32.6 of the teacher responded 'Never', 52.5 responded 'Sometimes' and 14.9 responded 'Often.' A hypothesis test for differences in ordinal distribution across period (baseline vs midline) did not yield a significant result ($p = 0.335$) and thus there is no evidence that the distribution of responses differed across period.

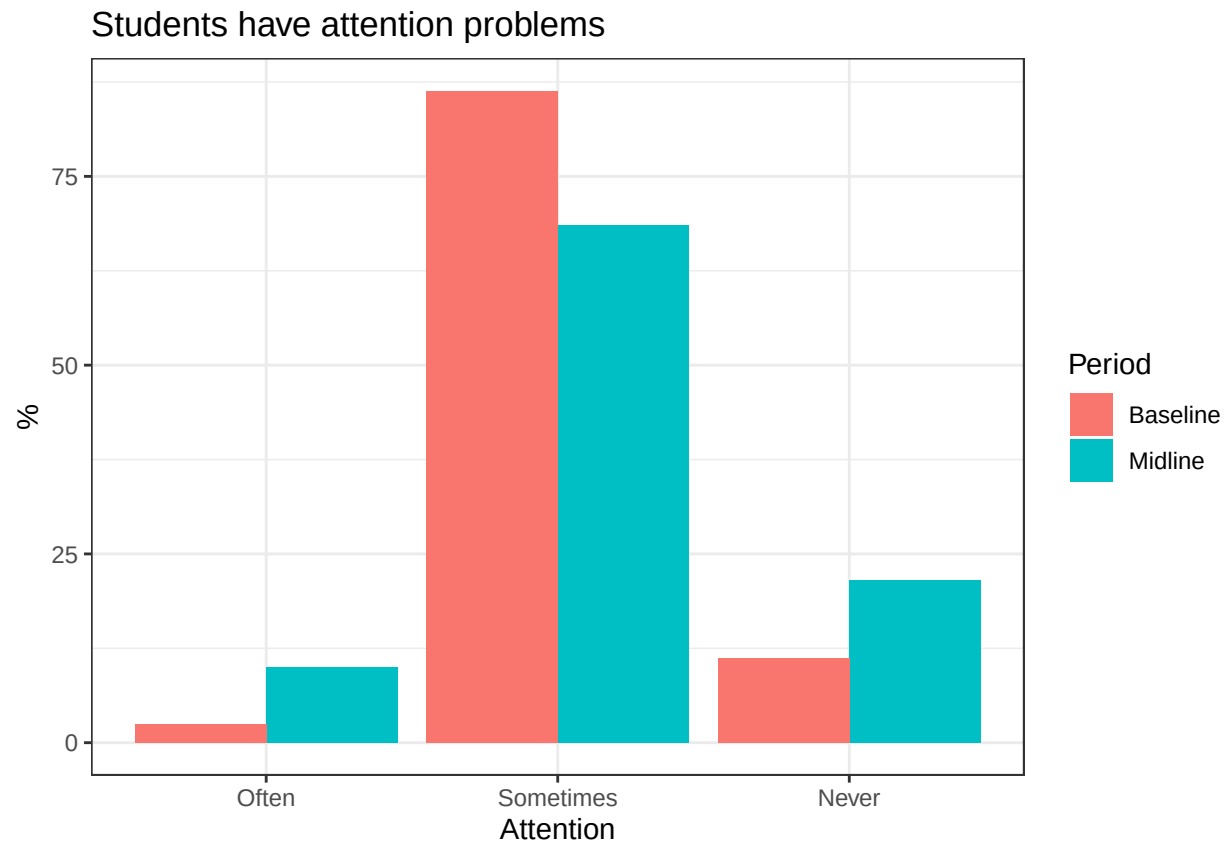
6.2 Students have attention problems

Table 116: Students have attention problems

	Baseline	Midline
Often	2.42	9.94
Sometimes	86.34	68.51
Never	11.23	21.55

Table 117: Students have attention problems

	Baseline	Midline
Often	11	18
Sometimes	392	124
Never	51	39



As can be seen in the table above, in regards to pupils being hungry at baseline, 11.2 of the teacher responded 'Never', 86.3 responded 'Sometimes' and 2.42 responded 'Often.' At midline, 21.6 of the teacher responded 'Never', 68.5 responded 'Sometimes' and 9.94 responded 'Often.' A hypothesis test for differences in ordinal distribution across period (baseline vs midline) did not yield a significant result ($p = 0.635$) and thus there is no evidence that the distribution of responses differed across period.

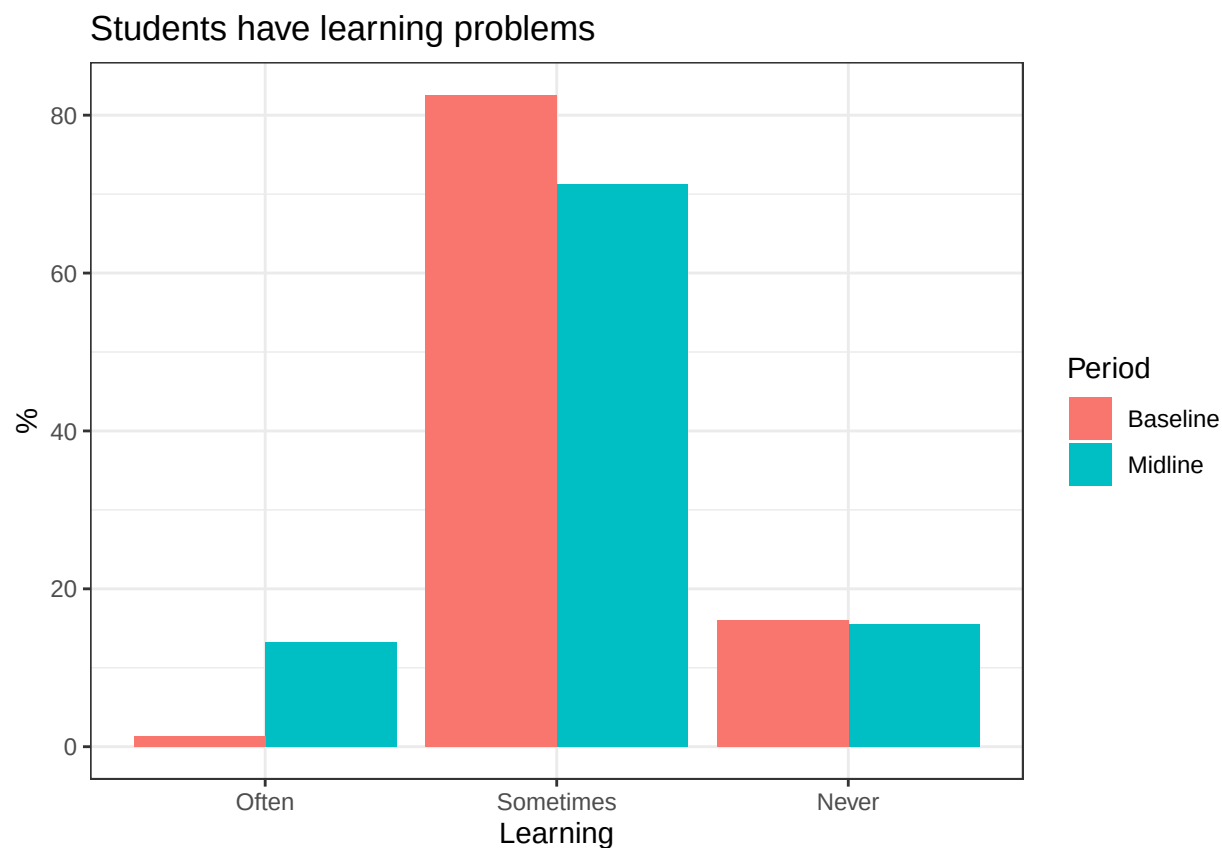
6.3 Students have learning problems

Table 118: Students have learning problems

	Baseline	Midline
Often	1.34	13.3
Sometimes	82.59	71.3
Never	16.07	15.5

Table 119: Students have learning problems

	Baseline	Midline
Often	6	24
Sometimes	370	129
Never	72	28



As can be seen in the table above, in regards to pupils having problems to learn at baseline, 16.1 of the teacher responded 'Never', 82.6 responded 'Sometimes' and 1.34 responded 'Often.' At midline, 15.5 of the teacher responded 'Never', 71.3 responded 'Sometimes' and 13.3 responded 'Often.' A hypothesis test for differences in ordinal distribution across period (baseline vs midline) yielded a significant result ($p = 0.0172$). This provides evidence that the difference in the distribution of percentages across baseline and midline could not be the result of sampling error.