
# USAID/MADAGASCAR IMPACT EVALUATION OF THE MAHAY MAMAKY TENY (MMT) PROJECT

AID-OOA-I-14-00054
TASK ORDER 720-687-18-F-00001

## ENDLINE IMPACT EVALUATION REPORT

**August 2018**

This publication was produced at the request of the United States Agency for International Development. It was prepared independently by Dr. Wael Moussa, Yvonne Cao and Nathalie Louge of FHI 360. The authors' views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

# USAID/MADAGASCAR IMPACT EVALUATION OF THE MAHAY MAMAKY TENY (MMT) PROJECT

## ENDLINE IMPACT EVALUATION REPORT

August 2018

AID-OAA-I-14-00054
Task Order 720-687-18-F-00001

# CONTENTS

# LIST OF TABLES AND FIGURES

# ACKNOWLEDGEMENTS

The impact evaluation of the Mahay Mamaky Teny (MMT) Project would not have been possible without the hard work and collaboration of many people. First, we would like to thank USAID/Madagascar for funding this opportunity and for their ongoing support over the course of the program.

Given the short timeframe of the program, the EGRA adaptation, administration, and analysis had to move quickly. This issue was mitigated by the presence and invaluable contributions of experienced Ministry of Education (MOE) evaluators and analysts from the Direction de la Planification de l'Education (DPE), and Direction des Technologies de l'Information et de la Communication (DTIC). In particular, we'd like to thank the participants of the Early Grade Reading Assessment (EGRA) adaptation workshop: Olivier Razafindranovona (National Evaluation Team Lead, DPE), Jullino Serge Rasamison (Statistics Service, DPE), Andry Nirisoa Rabenasolo (Evaluation team member, Research and Planning Service, DPE), Harison Andriarinivomanana (Service de la Conception, de l'Encadrement et de l'Evaluation - SCEE), Ruphin Santatriniaina (DTIC), Landitiana Miakatra Soamina (World Bank), Vonintsoa Ranivoharimanga (Direction des Curricula et des Intrants; DCI), Mirindra Andriamampianina. We'd also like to thank the people who facilitated the enumerator training and supervised data collection: Olivier Razafindranovona, Jullino Serge Rasamison, Harison Andriarinivomanana, Harimalala Randriambeloarinosy, and Andry Nirisoa Rabenasolo.

# AUTHORS

This report was produced by Principal Investigator Dr. Wael Moussa, Research and Evaluation Technical Advisor, Yvonne Cao, Literacy Advisor, Nathalie Louge, with guidance from FHI 360 Research and Evaluation Director, Dr. Carina Omoeva and quality assurance from Dr. Nina Menezes Cunha.

# LIST OF ACRONYMS

| | |
|---|---|
| CLA | Collaborating, Learning, Adapting |
| DCI | Direction des Curricula et des Intrants |
| DD | Difference-in-Differences; Double Differences |
| DPE | Direction de la Planification de l'Education |
| DREN | Direction Régionale de l'Education Nationale |
| DTIC | Direction des Technologies de l'Information et de la Communication |
| EGR | Early Grade Reading |
| EGRA | Early Grade Reading Assessment |
| FOI | Fidelity of Implementation |
| Lspm | Letter Sounds Per Minute |
| MOE | Ministry of Education |
| MMT | Mahay Mamaky Teny |
| PCA | Principal Components Analysis |
| SCEE | Service de la Conception, de l'Encadrement et de l'Evaluation |
| SES | Socio-economic Status |
| Spm | Syllables per minute |
| USAID | United States Agency for International Development |
| WB | World Bank |
| ZAP | Zone Administrative et Pédagogique |

# EXECUTIVE SUMMARY

## PROJECT BACKGROUND

The USAID/Madagascar Mahay Mamaky Teny (MMT) program is a capacity building activity, designed to strengthen the Madagascar Ministry of Education's (MOE) ability to design teaching and learning materials and to implement and monitor teacher training programs in relation to early grade reading. It was implemented from January to July 2018 with school-based activities starting in April 2018. The project accompanied the MOE to (1) develop and pilot teaching and learning materials (teachers' guides and pupil books) designed to teach Grade 1 children how to read in the Malagasy language, (2) train Grade 1 teachers and headteachers in the use of these teaching and learning materials, (3) train headteachers to coach and supervise teachers, (4) train internal MOE staff in the design of impact evaluations and data collection instruments, (5) train internal MOE staff in the use and administration of data collection instruments. In sum, the project used a "learning-by-doing" approach whereby project staff worked hand-in-hand with MOE staff throughout all phases of the project including both development and implementation of the early grade reading (EGR) interventions but also monitoring and evaluation activities.

The original scope of the MMT program was to pilot this approach in 8 schools and select an additional 8 schools as a comparison group. However, during the early stages of the project, the Ministry of Education expressed interest in expanding the pilot to cover a total of 60 intervention schools and 60 control schools. The World Bank (WB) funded the remaining 52 schools from both treatment and control schools. As a result, this report describes the impact evaluation results for 60 intervention (treatment schools) and a similar number of control schools (59 control schools)[1].

## EVALUATION PURPOSE AND EVALUATION DESIGN

The purpose of this impact evaluation is to document and measure the impact of the MMT project in 60 intervention schools across two Directions Régionales de l'Education Nationale (DREN) of Madagascar in terms of improvement of learning outcomes in reading.

This report seeks to address two main research questions:

1. What is the impact of the EGR intervention on students' performance in reading? What are the differences in performance improvements between student sub-groups?
2. What are the links between student learning growth and teacher characteristics and behavior (especially as it relates to implementation of practices promoted by the EGR intervention)?

The impact evaluation of MMT uses an experimental design whereby treatment and control schools were randomly selected. Students were assessed prior to the start of the intervention in April 2018 and again at the end of the intervention in July 2018. The impact evaluation employs a Difference-in-Differences (DD) model. This method involves comparing the changes between

[1] FHI 360 conducted an analysis with the full sample as well as with the original sample of 8 control and 8 pilot schools. Both analyses yielded very similar results. This report therefore focuses on the larger sample.

baseline and endline test scores in treatment schools to changes between baseline and endline test scores in control schools. In addition, the empirical strategy we employ also disaggregates the DD estimates by whether the teachers were following the specified steps of each lesson activity as designed by the program team. Qualitative data collected via the Collaborating, Learning, and Adapting (CLA) approach is also used to explain the impact evaluation results through the report.

## EVALUATION FINDINGS

This evaluation measures the impact of the MMT intervention on reading outcomes of T1 students in 60 schools where the program is implemented. We measure program impact on the following subtasks of the Early Grade Reading Assessment (EGRA): letter sound identification, syllable identification, familiar word reading, oral reading fluency (ORF) measured with a passage timed at 1-minute and with a passage timed at 2-minutes, reading comprehension, and listening comprehension. The analysis also includes investigating program impacts on the average subtask score, the probability of a zero score, and the probability of meeting minimal proficiency (as defined by benchmarks developed with MOE).

The baseline data collection involved sampling from 1,584 students across 119 schools, 60 of which are treatment schools and 59 are control schools and were tracked longitudinally through the school year. At baseline we found that the two groups were not perfectly balanced along student observable characteristics, letter sound scores, and oral reading fluency. Students randomly selected into the treatment group had a higher likelihood of being high SES (above average on the wealth index), have access to books, and a slightly higher score in letter sounds and oral reading fluency. By endline, 1,028 students were available for a re-interview and a second administration of EGRA. The students that dropped out of the sample were not significantly different from those that remained in terms of their background characteristics but performed slightly worse on two EGRA subtasks (letter and syllable identification) at baseline. However, by endline, the gap between the MMT and control schools widens significantly and is reflected in the regression results.

We estimate that, **on average, students attending MMT schools outperformed similar students in the control schools by 2.73 letter sounds per minute for teachers who implemented some steps of the lesson (or less) and by 6.09 letter sounds per minute for teachers who implemented most or all steps of the lesson for an average treatment effect of 3.24 letter sounds per minute. We also find an average treatment effect of 3.21 syllable identification per minute, 1.42 correct words per minute for familiar word reading and 2 correct words per minute for oral reading fluency.** Students in the treatment group showed more modest gains in reading and listening comprehension over the control group by 5.7 and 2.1 percentage points, neither of which were statistically significant. These impacts correspond to **effect sizes ranging from 0.20 to 0.37 standard deviations** depending on the subtask, which is quite remarkable for a 3-month intervention. Moreover, **when disaggregating the treatment effect among those students whose teachers followed most or all lesson activity steps and those who did not, we find that the former group achieves higher scores than the latter group, implying that teachers who followed the MMT lesson plan with more fidelity obtained better results that teachers who did not**. We also stratify program impacts by

gender and SES where we see that **boys and girls are affected similarly by MMT**, but the program was generally **more impactful among students in the high SES group**. This means that the pre-existing achievement gap between low and high SES students has widened following the treatment.

**The MMT intervention was also successful in improving upon the proportion of students with zero scores and those with scores that meet minimal proficiency benchmarks**. On average, we see improvements in terms of zero scores in syllables, familiar words and oral reading fluency by 14.2, 4.9, and 11.3 percentage points, respectively. On the other hand, **MMT impacts were more consistent in terms of improving students' probability of meeting minimal proficiency across all subtasks by 3.8 percentage points (pp) for letter sounds, 2.6 pp for syllable identification, 9.6 pp for familiar word reading, and 5.1-7.1 pp for oral reading fluency.** At the zero and minimal proficiency margins, we again observe a similar pattern where among the treatment schools, those whose teachers followed the lesson activities with a high level of fidelity exhibited larger gains than those whose teachers did not.

## DISCUSSION AND FINAL COMMENTS

In sum, we found that the EGR intervention had positive and significant impacts on reading outcomes on three different measures of the early grade reading assessment (EGRA) – mean EGRA subtask scores, proportion of zero scores, and probability of meeting minimum benchmarks. Further, when disaggregating the treatment effect by whether the classroom teacher followed the lesson activity steps as laid out by the MMT intervention, we find that those who do have an even higher impact than those who do not. The estimates associated with following the steps of the lesson activity, because they are non-random in nature, may not necessarily be attributed to the steps themselves. As such, we cannot rule out confounders such as teacher attitude toward the intervention, teacher quality, or the student population being different between the group of teachers who implement the intervention with a high degree of fidelity and those who do not. However, **because the overall treatment assignment is random, we can conclude that the net causal effect of receiving the treatment and implementing at a high level of fidelity is positive in boosting T1 student reading outcomes**. Meaning that the combination of the treatment and implementing the treatment with a high level of fidelity (following the lesson steps closely) result in significant reading gains over a short period of time.

One important aspect of the MMT lesson plans concerns the focus on students' independent practice, the "you do" step of the "I do, we do, you do" method. This gives students more opportunity to read letters, syllables, words and/or text. Given this focus, it is not surprising that treatment school students showed significant improvements in the mechanics of reading (decoding). On the other hand, it is also unsurprising that the project did not have an impact on reading comprehension. Given the short duration of the project (only 3 months of school-based implementation), teachers had not yet reached the lessons where skill focus switches from a reading mechanics and decoding to building reading accuracy and fluency through reading leveled text and reading comprehension.

# SOMMAIRE EXECUTIF

## SOMMAIRE DU PROGRAMME MAHAY MAMAKY TENY (MMT)

Débutant en janvier 2018, le programme de Mahay Mamaky Teny (MMT) a été conçu pour fournir sept mois d'assistance technique au Ministère de l'Education National (MEN) pour leur fournir des outils et de l'expertise afin qu'ils puissent décider quelle méthode de lecture à continuer et rependre à Madagascar. Plus spécifiquement, le programme a accompagné le MEN à: 1) développer et piloter des matériels d'enseignement-apprentissage (guide de l'enseignant et cahier de l'apprenant de lecture-écriture) en Malagasy, (2) former des enseignants du T1, des directeurs d'école et des Chefs ZAP à l'utilisation de ces matériels d'enseignement-apprentissage, (3) former des directeurs d'école pour encadrer et superviser les enseignants à l'aide des outils, (4) former le personnel technique du MEN dans la conception d'évaluation d'impact et d'instruments de collecte de données, et (5) former le personnel technique du MEN à l'utilisation et à l'administration des instruments de collecte de données. Les activités de pilotage du matériel et de la formation ont débuté en avril 2018 et ont pris fin en juillet 2018. En résumé, le programme a utilisé une approche «d' apprentissage par la pratique» par laquelle le personnel du programme a travaillé en étroite collaboration avec le personnel du MEN dans toutes les phases du projet, y compris le développement et la mise en œuvre des interventions de lecture-écriture, de la formation, et des activités d'évaluation.

Le plan initial du programme MMT était de piloter ses différentes activités dans 8 écoles et de sélectionner 8 écoles supplémentaires comme groupe de comparaison. Cependant, au début du projet, le MEN a recommandé un élargissement du projet pilote à 60 écoles pilotes et à 60 écoles de contrôle. La Banque mondiale (BM) a financé les 52 écoles supplémentaires des écoles pilote et de contrôle. En conséquence, ce rapport décrit les résultats de l'évaluation de l'impact venant des 60 écoles pilotes (écoles de traitement) et d'un nombre similaire d'écoles de contrôle (59 écoles de contrôle).

## OBJECTIFS ET CONCEPTION DE L'EVALUATION

L'objectif de cette évaluation d'impact est de documenter et de mesurer l'impact du programme de lecture-écriture du T1 de MMT dans 60 écoles pilotes réparties sur deux Directions Régionales de l'Éducation Nationale (DREN) de Madagascar en termes d'amélioration des résultats d'apprentissage en lecture.

Ce rapport vise à répondre à deux questions de recherches principales:

1. Quel est l'impact du programme de lecture-écriture du T1 de MMT vis-à-vis la performance des apprenants du T1 en lecture ? Quelles sont les différences d'amélioration en lecture entre le groupe d'apprenants des écoles pilotes et celui des écoles de contrôle ?
2. Quels sont les corrélations entre l'amélioration des performances de lecture des apprenants et les caractéristiques des enseignants et leurs pratiques (en termes de mise en œuvre des pratiques et stratégies encouragées dans le programme de lecture-écriture MMT) ?

L'évaluation d'impact du MMT utilise un modèle expérimental selon lequel les écoles pilotes et de contrôle ont été sélectionnées de manière aléatoire. Les apprenants ont été évalués avant le début de l'intervention en avril 2018 et à la fin de l'intervention en juillet 2018. L'évaluation d'impact utilise un modèle de différence dans les différences (DD). Cette méthode consiste à comparer les changements de performance des apprenants entre les résultats de l'évaluation de base (baseline) et des résultats de l'évaluation finale (endline) dans les écoles pilotes aux changements entre les résultats de l'évaluation baseline et des résultats de l'évaluation endline des écoles de contrôle. En outre, la stratégie empirique que nous utilisons désagrège également les estimations de DD pour les enseignants qui ont suivi les étapes spécifiées dans les leçons de lecture-écriture du T1 telles que conçues par l'équipe du programme MMT et pour ceux qui ne l'ont pas fait. Les données qualitatives collectées via l'approche Collaborating, Learning, and Adapting (CLA), c'est-à-dire l'évaluation formative sont également utilisées pour expliquer les résultats de l'évaluation d'impact dans ce rapport.

## RESULTATS DE L'EVALUATION

Cette évaluation mesure l'impact des activités MMT sur les résultats en lecture des apprenants du T1 dans 60 écoles où le programme a été mis en œuvre. Nous mesurons l'impact du programme en utilisant les sous-tâches suivantes de l'EGRA: l'identification du son des lettres, l'identification des syllabes, la lecture des mots familiers, la fluidité de la lecture orale mesurée avec un passage chronométré à 1 minute et avec un passage chronométré à 2 minutes, la compréhension en lecture et la compréhension à l'audition. L'analyse comprend également l'étude des impacts du programme MMT sur la moyenne des scores des sous-tâches, la probabilité d'un score zéro et la probabilité qu'un apprenant atteignent les seuils de performances minimaux (telles que définies par les standards et seuils développés avec le MEN).

La collecte de données de base a consisté à échantillonner 1.584 apprenants dans 119 écoles, dont 60 écoles pilotes et 59 écoles de contrôle, suivies longitudinalement pendant toute l'année scolaire. Au baseline, nous avons trouvé que les deux groupes n'étaient pas parfaitement équilibrés sur les caractéristiques observables de l'apprenant, notamment pour les scores d'identification des sons des lettres et la fluidité de la lecture orale. Les apprenants sélectionnés aléatoirement pour le groupe pilote étaient plus susceptibles de venir de milieux plus avantagés (supérieurs à la moyenne sur l'indice de statut socioéconomique), ont accès aux livres et obtiennent un score légèrement supérieur dans les sous-tâches sons des lettres et la fluidité de la lecture orale. A l'évaluation endline, 1.028 apprenants étaient disponibles pour une autre administration de l'EGRA. Les apprenants qui ont quitté l'échantillon n'étaient pas significativement différents de ceux qui sont restés en termes de caractéristiques de base, mais avaient un rendement légèrement inférieur sur deux sous-tâches EGRA (sons de lettres et lecture de syllabes) au baseline. Cependant, à la fin, l'écart entre les écoles pilotes et les écoles de contrôle s'élargit de manière significative et se reflète dans les résultats de régressions.

Nous estimons qu'en **moyenne, les apprenants fréquentant les écoles pilote de MMT ont surpassé les apprenants des écoles de contrôles par 2,73 sons de lettres par minute pour les enseignants qui ont appliqué quelques étapes de la leçon (ou moins) et 6,09 sons de lettre par minute pour les enseignants ayant exécuté toutes**

**les étapes de la leçon pour un effet de traitement moyen de 3,24 sons de lettre par minute. Nous trouvons également un effet de traitement moyen de 3,21 syllabes lues par minute, 1,42 mots correctement lus par minute pour la lecture de mots familiers et 2 mots corrects par minute pour la fluidité de lecture orale**. Les apprenants du groupe pilote ont démontré des gains plus modestes en compréhension de la lecture et à l'audition par rapport au groupe de contrôle de 5,7 et 2,1 points de pourcentage, dont aucun était statistiquement significatif. Ces impacts correspondent à des tailles d'effet allant de **0,20 à 0,37 écart-type** selon la sous-tâche, ce qui est tout à fait remarquable pour une intervention de 3 mois. De plus, **en décomposant l'effet du traitement parmi les apprenants dont les enseignants ont suivi la plupart ou la totalité des étapes de des leçons et ceux qui n'en ont pas suivi, le premier groupe obtient des résultats supérieurs à ceux du deuxième groupe. Les enseignants qui ont mis en œuvre les leçons MMT avec plus de fidélité ont obtenu de meilleurs résultats que les enseignants qui ne l'ont pas fait.** Nous stratifions également les impacts du programme selon le sexe et le statut socioéconomique et constatons que les garçons et les filles sont affectés de la même manière par le programme MMT, mais que le programme a généralement eu plus d'impact chez les apprenants du groupe socioéconomique plus élevé. Cela signifie que l'écart de performance préexistant entre les apprenants de statuts socioéconomiques faible et élevé s'est élargi après le traitement.

**Le programme MMT a également permis de diminuer la proportion d'apprenants avec un score zéro et d'augmenter la proportion qui atteint les seuils de performance en lecture**. En moyenne, nous constatons des diminutions en termes de score zéro dans la lecture des syllabes, de mots familiers et en fluidité de la lecture orale de 14,2, 4,9 et 11,3 points de pourcentage, respectivement. **Les impacts du pilote MMT ont aussi été cohérents en termes d'amélioration de la probabilité qu'un apprenant atteigne les seuils de performance en lecture dans toutes les sous-tâches - de 3,8 points de pourcentage pour les sons de lettre, de 2,6 points pour les syllabes, de 9,6 points pour la lecture des mots familiers et de 5,1 points à 7,1 points pour la fluidité de la lecture orale**. Aux marges de scores zéros et de seuils minimaux de performance, nous observons de nouveau un schéma similaire: parmi les écoles pilotes, celles dont les enseignants suivaient les étapes de la leçon avec un niveau de fidélité élevé affichaient des gains plus importants que celles dont les enseignants ne le faisaient pas.

## DISCUSSION ET COMMENTAIRES FINAUX

En résumé, nous avons constaté que le pilote de lecture-écriture MMT du T1 a obtenu des impacts positifs et significatifs sur les résultats de lecture des apprenants sur trois mesures différentes de l'évaluation de lecture (EGRA): scores moyens des sous-tâches EGRA, proportion de scores zéros et probabilité d'atteinte des seuils de performances minimaux. En outre, lorsque nous désagrégeons l'effet de traitement en tenant compte de la fidélité de mise en œuvre des leçons par l'enseignant (en comparant les enseignant qui ont mis en œuvre la plupart ou toutes les étapes de la leçon avec les enseignants qui ne l'ont pas fait), nous constatons que l'impact de l'intervention est plus élevé pour les enseignants qui ont mis en œuvre les leçons avec fidélité. Les estimations associées au suivi des étapes de la leçon, parce qu'elles ne sont pas de nature aléatoire, ne sont pas nécessairement attribuables aux étapes elles-mêmes. En tant que tel, nous ne pouvons pas exclure les facteurs de confusion tels que

l'attitude des enseignants envers le programme, la qualité de l'enseignement ou la différence entre les groupes d'enseignants qui mettent en œuvre le programme avec un degré élevé de fidélité et ceux qui ne le font pas. **Cependant, comme l'attribution globale du traitement est aléatoire, nous pouvons conclure que l'effet causal net de recevoir le programme de lecture-écriture et de le mettre en œuvre avec un haut niveau de fidélité est positif pour stimuler les résultats de lecture des apprenants du T1. Cela signifie que la combinaison du traitement et la mise en œuvre du traitement avec un haut niveau de fidélité (en suivant les étapes de la leçon de près) entraînent des gains de lecture significatifs sur une courte période de temps.**

Un aspect important des leçons du MMT concerne l'accent mis sur la pratique indépendante des apprenants, l'étape «tu fais» de l'approche «Je fais, nous faisons, tu fais». Cela donne aux apprenants plus d'opportunité pour s'entrainer à la lecture des lettres, des syllabes, des mots et de textes. Compte tenu de cet objectif, il n'est pas surprenant que les apprenants des écoles pilotes aient démontré des améliorations significatives dans les aspects mécaniques de la lecture (décodage). D'autre part, il n'est pas surprenant que le pilote n'ait pas eu d'impact significatif sur la compréhension en lecture. Étant donné la courte durée du projet (seulement 3 mois de mise en œuvre), le contenu et les compétences couverts par la tranche du programme pilotée se focalisaient plus sur le décodage que la compréhension car le décodage est une base nécessaire pour entrainer la compréhension en lecture. Pour une discussion plus détaillée, voir la section discussion et commentaires finaux dans ce rapport.

# A. PROJECT BACKGROUND

The USAID/Madagascar Mahay Mamaky Teny (MMT) program is a capacity building activity, designed to strengthen the Madagascar Ministry of Education's (MOE) ability to design teaching and learning materials and to implement and monitor teacher training programs in relation to early grade reading. It was implemented from January to July 2018 with school-based activities starting in April 2018. The project accompanied the MOE to (1) develop and pilot teaching and learning materials (teachers' guides and pupil books) designed to teach Grade 1 children how to read in the Malagasy language, (2) train Grade 1 teachers and headteachers in the use of these teaching and learning materials, (3) train headteachers to coach and supervise teachers, (4) train internal MOE staff in the design of impact evaluations and data collection instruments, and (5) train internal MOE staff in the use and administration of data collection instruments. In sum, the project used a "learning-by-doing" approach where project staff worked hand-in-hand with MOE staff throughout all phases of the project including both development and implementation of the early grade reading (EGR) interventions but also monitoring and evaluation activities. The culmination of this program was to help the MOE make data-driven decisions around what to scale-up, in the context of their new education reform.

The original scope of the FHI 360 project was to pilot this approach in 8 schools and select an additional 8 schools as a comparison group. However, during the early stages of the project, the Ministry of Education expressed interest in expanding the pilot to cover a total of 60 intervention schools and 60 control schools. The World Bank (WB) funded the remaining 52 schools from both treatment and control schools. As a result, this report describes the impact evaluation results for 60 intervention (treatment schools) and a similar number of control schools (59 control schools)[2].

Grade 1 reading and writing program pilot began in April 2018 and ended in July 2018. Over these four months, Grade 1 pilot school teachers were trained twice on the effective use of a teacher's guide and student book and supported by mentors (school directors). Every Grade 1 pilot school student received a book which included all content for the lessons: letters, syllable, words, decodable and leveled text, and writing exercises. Students were encouraged to take these books home to continue their reading and writing practice. Parents of Grade 1 students were also sensitized to the program and shown how to support their child at home on the use of the student book.

The reading and writing program was based on 33 weeks. These weeks were divided into three sequences that each had their "focus" skills, as shown in Table 1.

[2] FHI 360 conducted an analysis with the full sample as well as with the original sample of 8 control and 8 pilot schools. Both analyses yielded very similar results. This report therefore focuses on the larger sample.

**Table 1. Lesson sequence content**

| | | |
|---|---|---|
| Sequence 1 | 6 weeks | - Learn the names of the letters of the alphabet/<br>- Learn to write lowercase script letters<br>- Develop phonological awareness (manipulate sounds orally – words in sentences, syllables in words, sounds in syllables/words) |
| Sequence 2 | 17 weeks | - Learn to associate letters with sounds to decode syllables and words<br>- Practice decoding decodable texts (simple texts made up of letters that students have already studied)<br>- Learn to write upper case script letters<br>- Practice writing words by using letter-sound associations (spelling) |
| Sequence 3 | 10 weeks | - Learn to read texts accurately (fluency)<br>- Apply reading comprehension strategies<br>- Learn to writer lowercase cursive letters<br>- Practice written expression by using the text conventions and new vocabulary words learned |

Given the program began piloting in the final four months of the school year, teachers began teaching at week 14[3], that is, in the middle of sequence 2 which has a strong focus on letter-sound decoding and decodable text[4] reading. This means that the letters introduced prior to the start of the program were not necessarily taught the same way as prescribed with the MMT instructional approach. The pilot program was projected to run through the end of the school year. However, the MMT program timeline required the EGRA endline to be conducted in early July in order for results to be ready by program close-out. Therefore, 13 weeks of piloting were originally projected. Unfortunately, toward the end of the four-month pilot period, schools in Madagascar were closed for 3 weeks due to national teacher strikes. This diminished the number of lessons students received from the originally planned 13 weeks of lessons to pilot to an average of 9 to10 weeks. This means that most students were never exposed to lessons from sequence 3 which switches from a reading mechanics and decoding focus to building reading accuracy and fluency through reading leveled text[5] and reading comprehension. These elements are important to note when considering the EGRA endline results presented in this report.

The results of this impact evaluation will also serve to help the MOE make an informed decision on which reading approach to scale-up. Early on in the program, the MOE communicated that other reading initiatives existed but that no program had yet been identified to scale-up. At this time, the MOE requested that all partners active in reading like engage in experimentation of their approach so that the MOE could use the results to make this decision. This report aligns

[3] The choice of where to start in the pilot program was made after visits to the 8 MMT schools to determine where teachers were in the national program. Following this visit, it was determined that 90% of teachers had attained week 13.

[4] Decodable texts are simple text which are designed to focus on specific letter-sound correspondences. For transparent languages like Malagasy, texts contain only the letters that students have already learned previously so that students can practice applying their decoding skills to the letter-sounds they've already learned.

[5] Leveled texts are written to specific leveling criteria determined by a leveling rubric. As text levels increase, the difficulty of the text also increases in terms of sentence length, word length, number of words, number of unique words, complexity of vocabulary and text structure, etc. In Madagascar, students at the end of Grade 1 are expected to read a level 4 text (33 words, 20 unique words, maximum of 6 sentences).

with the MOE's request for data. A parallel report written by the MOE evaluation specialists will also be produced as a way to validate the present analysis. It is hoped that both reports in addition to the final report will provide the MOE with sufficient data to decide on a reading and writing program to fund at scale.

# B. EVALUATION PURPOSE & EVALUATION DESIGN

## B1. EVALUATION PURPOSE

The purpose of this impact evaluation is to document and measure the impact of the four-month MMT reading and writing program pilot in 60 intervention schools across two Directions Régionales de l'Education Nationale (DREN) of Madagascar in terms of improvement of Grade 1 student learning outcomes in reading.

This report seeks to address two main research questions:

1. What is the impact of the EGR intervention on students' performance in reading? What are the differences in performance improvements between student sub-groups?
2. What are the links between student learning outcomes and teacher characteristics and behavior (especially as it relates to implementation of practices promoted by the EGR intervention)?

## B2. EVALUATION DESIGN

In any impact evaluation constructing a valid counterfactual constitutes the main methodological challenge. The ideal comparison group stems from the use of experimental methods in which eligible participants are randomly assigned to receive the intervention or not. The impact evaluation of MMT uses such an experimental design whereby treatment and control schools were randomly selected. Students were assessed prior to the start of the intervention in April 2018 and again at the end of the intervention in July 2018. The impact evaluation employs a Difference-in-Differences (DD) model. This method involves comparing the changes between baseline and endline test scores in treatment schools to changes between baseline and endline test scores in control schools. Table 2 presents an overview of indicators, data sources and analysis methodology associated with each research question.

**Table 2. Evaluation matrix**

| Research Question | Indicator | Data Source | Methodology |
|---|---|---|---|
| Question 1 | | | |
| 1. What is the impact of the EGR intervention on students' performance in reading? What are the differences in performance improvements between student sub-groups? | Early Grade Reading Assessment (EGRA) scores | EGRA instruments | Dfference-in-differences model |

| Research Question | Indicator | Data Source | Methodology |
|---|---|---|---|
| Question 2 | | | |
| 2. What are the links between student learning growth and teacher characteristics and behavior (especially as it relates to implementation of practices promoted by the EGR intervention)? | • EGRA scores<br>• Teacher characteristics<br>• Teacher behavior, particularly fidelity of implementation of the MMT program | • EGRA instruments<br>• Teacher questionnaire<br>• Classroom observation tool | Multivariate analysis (multi-level hierarchical linear models (HLM)) |

## B2.1 Indicator Measurement and Instruments

The evaluation focuses on two types of outcomes:

1. Students' test scores on Early Grade Reading Assessment (EGRA) subtasks
2. Proportions of students who meet reading benchmarks

In addition, the evaluation also looks at the links between EGRA scores and teacher characteristics and behaviors. In this section, we describe the process taken to measure these indicators.

### *Adaptation of the Early Grade Reading Assessment*

At the beginning of the project, in February 2018, MMT led a standards and benchmark setting workshop involving MOE staff, education partners, linguists, Grade 1-Grade 3 teachers and school directors to set standards, benchmarks and text-leveling criteria. Standards are observable and measurable descriptions of what students should be able to do for each subject or component skill at each grade level. To be operational, standards need to be accompanied by grade-specific benchmarks or specific measures of expected level of performance. The standards and their accompanying benchmarks provide a roadmap for the specific reading and writing tasks all learners should be able to accomplish at specific points along their learning continuum.

Following the standard and benchmarking setting workshop, MMT led a four-day adaptation workshop of the Early Grade Reading Assessment (EGRA) with MOE stakeholders in order to modify the EGRA which was administered in 2015 to ensure that the stories aligned with the Grade 1 standards, and particularly that the reading and listening comprehension passages presented in the EGRA aligned with the text-levelling criteria set for Grade 1. The workshop attendees all participated in the piloting of the EGRA tool at two nearby schools around Antananarivo. This resulted in the creation of two forms of the EGRA, i.e. two equated versions; one used for the baseline and the other used for the endline. A more detailed description of the adaptation process and subsequent psychometric analysis is presented in Annex I.

As mentioned, for this evaluation we use both the students' test scores on the EGRA subtasks, such as the number of items read correctly in one minute (for timed tasks), or the percentage of comprehension questions answered correctly. We also look at the percentage of zero scores for each subtask. Additionally, we also compute the proportion of students who achieved the grade-level benchmarks set during the standard and benchmarking setting workshop, for each task.

The EGRA used in this evaluation includes the following subtasks.

**Table 3. MMT EGRA content**

| Early Literacy Skill | Sub-test | Measurement |
|---|---|---|
| Alphabetic Knowledge | Letter Sound Recognition | Number of letter sounds correctly identified out of 100 in 60 seconds |
| Phonics/Alphabetic Principle | Syllable identification | Number of syllables correctly read out of 50 in 60 seconds |
| Phonics/Alphabetic Principle | Familiar word reading | Number of familiar words correctly read out of 50 in 60 seconds |
| Oral Reading Fluency | Oral Passage Reading timed at one minute | Number of words in a reading passage read with accuracy in one minute |
| Oral Reading Fluency with Reading Comprehension | Oral Passage Reading timed at two minutes, followed by comprehension questions (text stays in front of student) | Number of questions (out of 5) about a reading passage read by the student answered correctly |
| Listening Comprehension | Oral Recall | Number of questions (out of 5) about a reading passage read aloud by an enumerator answered correctly |

***Data collected using the Collaborating, Learning, and Adapting (CLA) approach or formative evaluation***

CLA was MMT's core method to piloting the Grade 1 reading program, teacher training, coaching model, and parent sensitization/communication strategy. Several tools were employed to collect data on these pilot interventions. In particular, one important tool is the classroom observation designed to measure the extent to which teachers implemented the MMT teacher guides with fidelity; i.e. designed to measure Fidelity of Implementation (FOI). All formative evaluation tools are described in Table 4.

**Table 4. CLA data collection instruments**

| Tool | Description | With whom ? |
|---|---|---|
| Fidelity of implementation | Tablet-based classroom observation tool used to observe a reading or writing lesson. | Each Grade 1 teacher in pilot schools |
| Teacher and student questionnaire | Paper tool to administer directly after each classroom observation. This questionnaire aims to obtain feedback on the teacher guide and student book from their users. | Each Grade 1 teacher in pilot schools and a group of Grade 1 students per classroom observed |
| School director questionnaire | Paper tool which aims to collect feedback from the director on his/her responsibilities as a pedagogical mentor and on parent sensitization. | Each school director from each pilot school |

| Tool | Description | With whom ? |
|---|---|---|
| Chef ZAP questionnaire | Paper tool to collect feedback from the Chef ZAP on his/her responsibilities as a support person to the school director in his/her role of pedagogical mentor | Each Chef ZAP from each pilot zone |
| Focus group with parents of Grade 1 students | Paper tool to collect feedback from parents on their participation in sensitization sessions and to gauge their involvement in their child's education, and ideas for how to engage them more effectively. | A group of parents of Grade 1 students per school |

Data from the FOI classroom observation were used to measure teacher practices in treatment schools. This allowed the project to track whether teachers were using the MMT teacher guide as intended. The data from these tools were linked with EGRA scores to understand whether the extent to which teachers implementing the program (i.e. the "dosage effect") had an impact on student learning outcomes. Data from all other tools were used to explain the results of the EGRA impact evaluation findings.

## B2.2 Sample Description and Data Collection

***Sampling Frame and Sampling Plan***

The study took place in two regions of Madagascar: Analamanga and Boeny and involved 60 treatment schools and 59 control schools.

The schools were selected by the Ministry of Education using a multi-stage cluster sampling strategy. First two regions of Madagascar were selected, Analamanga and Boeny. Then within each region, two school districts (CISCOs) were randomly selected. Within each CISCO, four Zones Administratives et Pédagogiques (ZAP) were randomly assigned to either the treatment and control group, resulting in 2 treatment ZAPs and 2 control ZAPs per CISCO, for a total of eight treatment ZAPs and eight control ZAPs across the entire sample.

ZAP are equivalent to clusters and were selected as the final sampling unit so as to ensure no contamination across schools in a cluster. Randomization at the ZAP level was necessary because MMT was also piloting a coaching program led by school head teachers and supported by the ZAP supervisors (called "Chef ZAPs") who support all schools within their ZAP. This means that all schools within a treatment ZAP received the treatment, and all schools within a control ZAP remained untouched by MMT. Prior to ZAP selection, all ZAPs within the CISCO that did not meet the accessibility criteria had been eliminated. Table 5 shows the sample selection process and the resulting number of schools in the sample.

**Table 5. Sample Selection**

| **REGION** | **Analamanga** | | | | **Boeny (West)** | | | |
|---|---|---|---|---|---|---|---|---|
| **SCHOOL DISTRICT (CISCO)** | **Ambohibhatrimo** | | **Manjakandriana** | | **Minj II** | | **Maravoay** | |
| **ZAP** | 2 pilot | 2 control | 2 pilot | 2 control | 2 pilot | 2 control | 2 pilot | 2 control |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Schools** | 15 across the 2 ZAPs | 15 across the 2 ZAPs | 15 across the 2 ZAPs | 15 across the 2 ZAPs | 15 across the 2 ZAPs | 15 across the 2 ZAPs | 15 across the 2 ZAPs | 15 across the 2 ZAPs |

In each school, the data collection targeted a random selection of 15 Grade 1 students which would be followed to endline (without replacement) for a longitudinal sample of 1800 students. Table 6 shows the target and achieved sample sizes for schools, students, teachers and classroom observations.

**Table 6. Target and achieved sample sizes**

| | Treatment | Control |
|---|---|---|
| ***Target*** | | |
| Schools | 60 schools | 60 schools |
| Students | 900 Grade 1 students | 900 Grade 1 students |
| Teachers | 60 Grade 1 teachers | N/A |
| Classroom observations | 60 Grade 1 classrooms | N/A |
| ***Achieved*** | | |
| Schools | 60 schools | 59 schools |
| Students | 798 (baseline) \| 492 (endline) | 818 (baseline) \| 426 (endline) |
| Teachers | 55 (endline) | N/A |
| Classroom observations | 55 (endline) | N/A |

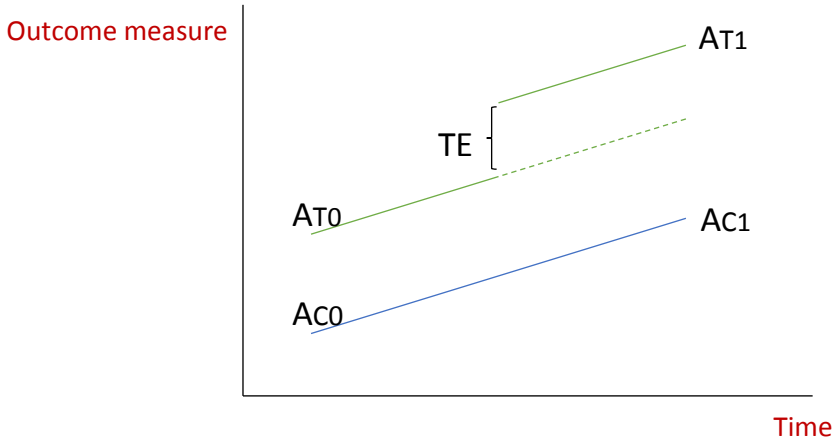
### *Data Collection*

The MOE led a comprehensive week-long training on how to administer the EGRA for all enumerators. The training was followed by a field practice in schools. The MOE provided supervision of the fieldwork throughout the data collection period.

## B2.3 Evaluation Approach and Methodology

A graphical representation of the difference-in differences methodology is depicted by Figure 1.

**Figure 1. Difference-in-Difference Estimator**




where,

$A_{T0}$ is the average test score at baseline in the treatment group

$A_{C0}$ is the average test score at baseline in the control group
$A_{T1}$ is the average test score at endline in the treatment group
$A_{C1}$ is the average test score at endline in the control group
TE is the treatment effect for the corresponding grade

In words, the proposed approach measures the difference between mean test scores for a given grade between baseline and endline, and then compares these differences between treatment and control groups.

***Analysis on timed reading scores***

Formally, we estimate

$$A_{it} = \alpha + \beta D_i + \gamma E_t + \delta D_i E_t + \mu S_i + \theta_i + u_{it} \qquad (1)$$

Where $A_{it}$ measures student achievement for student $i$ in period $t$; $D_i$ is a dummy variable for treatment status; $E_t$ is a dummy variable for the endline, $S_i$ indicates a matrix of observable characteristics for student $i$; $\theta_i$ is the student-specific time invariant effect, and $u_{it}$ is the error term and $\alpha$, $\beta, \gamma, \delta$ and $\mu$ are parameters to be estimated. The parameter of interest is $\delta$, which captures the effect of the program on students' outcomes at endline.

The identification assumption of this approach is that, in the absence of treatment, students in pilot and control schools would experience the same changes in the outcomes of interest and, therefore, any differential change between the two groups can be attributed to the program. The inclusion of student fixed effects to equation (1) enables us to account for individual level heterogeneity in observable and unobservable factors that are time-invariant.

***Analysis on proportion of students who score zero and who meet standards***

In addition to looking at changes in test scores, we also analyze the effect of the program on the probability that students who score zero and those who reach minimal proficiency. For this analysis we implement a model in the same fashion as equation (1), except that the dependent variable is a dummy for reaching a certain achievement threshold. Formally, we estimate:

$$\mathbf{I}(A_{it} > A^*) = \alpha + \beta D_i + \gamma E_t + \delta D_i E_t + \mu S_i + \theta_i + u_{it} \qquad (2)$$

Where $A^*$ is the score needed to achieve the minimum requirement for each grade. This model will allow us to estimate the effect of the program on the probability that students reach a meaningful achievement threshold.

***Analysis by Fidelity of Implementation***

Finally, we disaggregated the overall treatment effect by levels of program fidelity as implemented by the classroom teachers in schools receiving the treatment. In this case and among the treatment schools only, we formally test whether teachers who complete most or all steps associated with lesson activities as designed by the MMT intervention exhibit larger reading gains than their counterparts who do not. This analysis sheds light on the notion that certain aspects

of the program are an important component of the success of the program in boosting T1 reading performance.

As a result, equations (1) and (2) are estimated as follows.

$$A_{it} = \alpha + \beta^0 D_i + \beta^1 D_i * FOI_{\text{i}} + \gamma E_t + \delta D_i E_t + \mu S_i + \theta_i + u_{it} \quad (3)$$

$$\mathbf{I}(A_{it} > A^*) = \alpha + \beta^0 D_i + \beta^1 D_i * FOI_{\text{i}} + \gamma E_t + \delta D_i E_t + \mu S_i + \theta_i + u_{it} \quad (4)$$

where the term $FOI_i$ takes on a value of 1 if the classroom teacher tending to student $i$ follows most or all lesson activity steps, zero otherwise. As such, the parameter $\beta^0$ is the impact of the program when teachers do not follow all steps relative to the control group. $\beta^1$ denotes the impact of the program when teachers follow most or all steps relative to the treatment group where the teacher that does not follow the steps. Therefore, the program impact when lessons are followed at a high level of fidelity is the sum of $\beta^0$ and $\beta^1$. In other words,

- Impact of MMT when teachers do not follow the lesson steps = $\beta^0$
- Impact of MMT when teachers follow the lesson steps = $\beta^0 + \beta^1$

This approach allows to test for statistical significance whether following the lesson steps is associated with higher reading scores.

***Analysis by subgroup***

When sample size allows for it, we conduct separate regressions to analyze heterogeneity, i.e. differential impacts on different subgroups (for example, to study how boys and girls are differently affected by treatment). The approach is similar to those presented in equations (1) and (2), where we estimate each of these equations separately for each subgroup of students. In this case, we stratify by gender and by socioeconomic status (SES). We determine students' SES using a Principal Components Analysis (PCA) using variables denoting household item ownership.

## B2.4 Study Limitations

First, we tested for student attrition to ascertain whether attrition was non-random. We did not find any correlation between attrition and treatment status, so we do not control for attrition in our main impact evaluation results. However, we did find that attrition was negatively correlated with baseline letter and syllable identification test scores, which could imply that the estimated results are only valid for the type of students that are less likely to drop from the sample.

Second, we tested for student sample balance at baseline and found slight differences in student characteristics in terms of SES and having non-textbooks at home in favor of the treatment group. We also find slight differences in letter sound identification and oral reading fluency 1-minute administration. In terms of implications on the findings of this report, the imbalance in baseline characteristics is not detrimental to the identification strategy as balance is not a necessary condition for the difference-in-differences approach to produce causal results. Further, differences in SES are also mitigated with the inclusion of student-level covariates and a

stratification of the program impact by SES. As such, the standard DD assumptions still apply to this analysis.

Lastly, when disaggregating the program impact by fidelity (teachers who follow the lesson steps relative to those who do not), we emphasize that differences in outcomes between the two groups are not necessarily only because of the steps themselves but could also represent a signal of teacher quality, buy-in to the objectives of the intervention, and that the intervention was appropriately designed for teaching Malagasy at this level. It is therefore, important to make this distinction when interpreting the disaggregated evaluation results.

# C. EVALUATION FINDINGS

The impact evaluation analysis follows two stages. First, we provide a descriptive analysis of the overall student sample, their observed demographic characteristics, home environment, socioeconomic status, and performance on the different EGRA subtasks, and by time period. The descriptive portion of the analysis will include a summary of the overall sample, inclusive of all 119 schools, followed by a straightforward difference in means between treatment and control schools along the student characteristics and EGRA scores to test for balance between the two groups. The analysis of the EGRA results at baseline and endline will also include a distributional analysis of student performance in oral reading fluency (ORF) and reading comprehension.

The second stage of the analysis completes a multivariate regression analytic framework.[6] The regression analysis employs a difference-in-differences approach to identify the program's impact on reading outcomes. However, as we detail later in the report, the identification strategy that we use in this evaluation requires an assumption that we are unable to test with the available data and the result of the randomization of the treatment assignment. Although we are still able to inform on the program's performance relative to the control group, we need to assume that pre-treatment trends and post-treatment trajectories are parallel for both the treatment and control groups to attribute causality to the program.

[6] We note that this analysis is necessary as sample balance may not be achieved, especially when randomization is based on a small number of clusters.

## C1. DESCRIPTIVE ANALYSIS

**Key Findings**

- **Of the 1,584 students sampled at baseline, 1,028 were available to complete the endline data collection resulting in a 35% attrition rate**
  - Attrition was not found to be significantly correlated with observed student characteristics
  - However, students who were not available at endline had slightly lower letter and syllable sound scores on the baseline EGRA
- **At baseline, treatment and control group samples were not perfectly balanced**
  - Students in the treatment schools were more likely to be allowed to take books home, have books other than textbooks at home, and more likely to be in the high SES group
  - The treatment group also had slightly higher letter sound and oral reading fluency scores at baseline
- **By endline, average scores of the treatment group were significantly higher than the control group on all EGRA subtasks**
  - While the control group exhibited some gains between baseline and endline, the treatment group showed learning gains that were significantly larger in magnitude
  - The treatment group outperformed their counterparts in terms of average scores, proportion with minimal proficiency, and proportion with zero scores

### C1.1 Sample Summary

The data collected included a detailed student questionnaire administered with the Early Grade Reading Assessment to ascertain students' background demographic information including their sex, age, learning materials at school, home literacy environment, and socioeconomic status. Students are asked whether there are books in the classroom and whether the books can be taken home, and whether their teacher assigns homework to assess the classroom learning environment. They are also asked whether they own any books at home that are not from school, whether there are family members who can read, whether they receive help with their homework, and the number of meals they eat per day to determine the home environment of each student in the sample. Lastly, students are asked a series of questions about owning specific household items which are then used in a Principal Components Analysis (PCA) to create a wealth index. Higher scores on the index denote students who are in a higher socioeconomic standing and vice versa.

Table 7 presents the average student characteristics across the entire sample from the 119 schools covering 1,616 Grade 1 students in total at baseline, 1,584 of whom have valid data

records. The table also includes a full sample description at endline to show any potential shifts in the remaining student characteristics.

At baseline, only 40% of the students have access to books not from the classroom and about 42% report receiving help from a family member with their homework. By endline, 43% of the remaining students report having access to books not from the classroom and 63% report receiving help with homework. At both baseline and endline, about 90% of the students report that there is at least one family member who can read, and students have access to 2.8 meals per day. One major shift between baseline and endline is in the proportion of students who report being able to take textbooks home, from 12% to 43%. This is not surprising given that student books created under MMT were designed to be taken home. Lastly, we summarize the results of the creation of the wealth index, our proxy for socioeconomic status. We construct the wealth index using PCA, which means that the index is measured on a standard normal scale with mean zero and is measured in standard deviations, i.e. the wealth index is measured as a z-score.

**Table 7. Summary of student characteristics**

| | **Baseline** | | | **Endline** | | |
|---|---|---|---|---|---|---|
| | Obs | Mean | Std. Dev. | Obs | Mean | Std. Dev. |
| Female | 1584 | 0.47 | 0.50 | 1028 | 0.46 | 0.50 |
| Male | 1584 | 0.53 | 0.50 | 1028 | 0.54 | 0.50 |
| Age | 1584 | 7.12 | 1.39 | 1028 | 7.25 | 1.39 |
| Books in the classroom | 1584 | 0.89 | 0.31 | 1028 | 0.94 | 0.24 |
| Books can be taken home | 1403 | 0.12 | 0.32 | 963 | 0.43 | 0.50 |
| Other books at home | 1550 | 0.40 | 0.49 | 1022 | 0.49 | 0.50 |
| Other family member who can read | 1557 | 0.91 | 0.28 | 1023 | 0.93 | 0.25 |
| Teacher assigns homework | 1540 | 0.62 | 0.48 | 1021 | 0.68 | 0.47 |
| Receives help with homework | 1481 | 0.42 | 0.49 | 693 | 0.63 | 0.48 |
| Number of meals per day | 1584 | 2.83 | 0.45 | 1028 | 2.92 | 0.31 |
| Attended kindergarten | 446 | 0.33 | 0.47 | 275 | 0.29 | 0.45 |
| Wealth Index | 1584 | 0.00 | 0.85 | 1028 | -0.00 | 0.83 |
| Low SES | 1584 | 0.20 | 0.40 | 1028 | 0.23 | 0.42 |
| High SES | 1584 | 0.20 | 0.40 | 1028 | 0.18 | 0.39 |

## C1.2 Sample Balance

Of the 1,616 students sampled at baseline, 798 students attended treatment schools (60 schools) and 818 attended control schools (59 schools). At endline, 918 students of the full sample were interviewed again with 492 from treatment schools and 426 from control schools. These students form the panel sample. Because attrition was not found to be significantly correlated with observed student characteristics, the analysis from this point forward will focus on the panel sample of 918 students only. We test for sample balance at baseline between the treatment group and the control group along observable student characteristics using simple t-tests to determine whether differences in means are statistically significant. Table 8 displays the student composition of the treatment and control groups in terms of their sex, age, classroom environment, home environment, meals, kindergarten attendance rate, and SES proxied by the wealth index. The results of this analysis show that the samples are balanced along most student demographic, and home environment characteristics. However, we observe statistically significant differences in several aspects of the classroom environment and socioeconomic standing. Specifically, students in the pilot schools are about 12% more likely to have books that are not from school at home. Students in the treatment group are also around .26 of a standard deviation higher on the SES

spectrum. Moreover, the treatment group is about 13 percentage points more likely to be in the highest quintile of the wealth index. This shows, that of the students who were present for both rounds of data collection, the treatment group is composed of children from relatively better off backgrounds which is reflected in the differences in the classroom environment as well as their SES standing.

**Table 8. Mean differences in student characteristics, by treatment status**

| | Control | Treatment | Diff(T-C) |
|---|---|---|---|
| Female | 0.45 | 0.46 | 0.01 |
| Male | 0.55 | 0.54 | -0.01 |
| Age | 7.00 | 7.07 | 0.07 |
| Books in the classroom | 0.90 | 0.90 | -0.01 |
| Other books at home | 0.33 | 0.46 | 0.12* |
| Other family member who can read | 0.92 | 0.92 | -0.00 |
| Teacher assigns homework | 0.61 | 0.62 | 0.01 |
| Receives help with homework | 0.39 | 0.43 | 0.04 |
| Number of meals per day | 2.84 | 2.85 | 0.01 |
| Attended kindergarten | 0.33 | 0.38 | 0.05 |
| Wealth Index | -0.12 | 0.14 | 0.26* |
| Low SES | 0.21 | 0.17 | -0.04 |
| High SES | 0.14 | 0.27 | 0.13* |
| Observations | 916 | | |

Notes: Significance is denoted as: * $p < 0.05$

It is plausible to conclude that the samples are not balanced at baseline, especially in terms of classroom environment which can be significant determinants of student performance. In other words, any differences in performance at endline between the treatment and control group cannot be fully attributable to the pilot program unless we assume that the treatment group would behave similarly to the control group in the absence of the intervention.

### C1.3 Comparison of EGRA results

Next, we compute mean EGRA subtask scores by treatment status and test for statistical significance of mean differences between the groups at baseline and again at endline. This exercise enables us to show simple double-difference (DD) estimates of the program effects without controlling for differences in student observed characteristics or their time invariant unobserved characteristics. Table 9 displays the number of observations, the mean score for the control group, the mean score for the treatment group, and the difference in means between the two groups at baseline and endline, separately.

**Table 9. Mean differences in EGRA scores, by treatment status**

| | **Baseline** | | | | **Endline** | | | |
|---|---|---|---|---|---|---|---|---|
| | N | C | T | T-C | N | C | T | T-C |
| **Timed Scores:** | | | | | | | | |
| Letter sounds | 918 | 4.03 | 4.76 | 0.73* | 916 | 7.84 | 13.89 | 6.06* |
| Syllable identification | 918 | 3.38 | 3.97 | 0.60 | 917 | 5.95 | 10.69 | 4.74* |
| Familiar words | 915 | 2.08 | 2.56 | 0.48 | 915 | 4.33 | 7.04 | 2.71* |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Oral reading fluency - 1 min | 909 | 2.02 | 2.69 | 0.66* | 916 | 4.52 | 7.27 | 2.75* |
| Oral reading fluency - 2 min | 892 | 2.23 | 2.56 | 0.34 | 899 | 3.66 | 6.11 | 2.44* |
| **Percent of Zeros:** | | | | | | | | |
| Letter sounds | 918 | 19.72 | 16.67 | -3.05 | 916 | 14.59 | 3.67 | -10.92* |
| Syllable identification | 918 | 58.45 | 52.24 | -6.21* | 917 | 43.19 | 25.87 | -17.33* |
| Familiar words | 915 | 71.70 | 64.56 | -7.14* | 915 | 52.47 | 35.71 | -16.76* |
| Oral reading fluency - 1 min | 909 | 38.95 | 34.22 | -4.73 | 916 | 23.94 | 17.55 | -6.39* |
| Oral reading fluency - 2 min | 892 | 34.21 | 32.28 | -1.93 | 899 | 27.32 | 20.29 | -7.02* |
| **Percent Minimal Proficiency:** | | | | | | | | |
| Letter sounds | 918 | 0.01 | 0.01 | 0.01 | 917 | 0.03 | 0.10 | 0.07* |
| Syllable identification | 918 | 0.01 | 0.01 | -0.00 | 917 | 0.02 | 0.06 | 0.04* |
| Familiar words | 918 | 0.03 | 0.05 | 0.02 | 917 | 0.10 | 0.19 | 0.10* |
| Oral reading fluency - 1 min | 918 | 0.03 | 0.05 | 0.02 | 917 | 0.07 | 0.16 | 0.09* |
| Oral reading fluency - 2 min | 913 | 0.03 | 0.03 | 0.00 | 911 | 0.06 | 0.14 | 0.08* |
| **Comprehension:** | | | | | | | | |
| Reading Comprehension Score | 433 | 5.00 | 7.22 | 2.22 | 579 | 9.15 | 20.00 | 10.85* |
| Listening Comprehension Score | 891 | 43.37 | 51.31 | 7.94* | 864 | 35.36 | 41.95 | 6.59* |

Notes: The column headings N, C, T, and T-C refer to the number of observations, mean for control group, mean for treatment group, and the mean difference, respectively. Asterisks denote statistical significance at the 5% level.

Across all subtasks, as well as all the different ways the subtasks are scored, **we observe a consistent pattern of EGRA performance gaps between the two groups in favor of the treatment group at baseline that gets larger by endline, the latter being the intended effect of the intervention**. This means that students in the treatment group have a slightly higher baseline reading ability level, and by endline the difference in mean performance increases substantially in magnitude by endline. Specifically, we determine statistically significant differences in baseline letter sounds, oral reading fluency, and listening comprehension. This again reinforces the need to rely on multivariate regression analysis to mitigate the potential biases in these mean differences that are due to the differences in the observed and unobserved student characteristics. In our regression analysis, we control for individual student fixed effects which account for all time invariant characteristics which may include base aptitude among other factors that could influence reading performance.

Figure 2, below, plots the distribution of letter sound scores for students in the pilot and control schools and highlight the mean score for each group. This is useful to graphically display the information portrayed in Table 9 and provides some perspective on the magnitude of the baseline gap in reading. Again, we observe the gap in performance widening between the treatment and control group between baseline and endline.

**Figure 2. Distribution of letter sound scores, by treatment status**

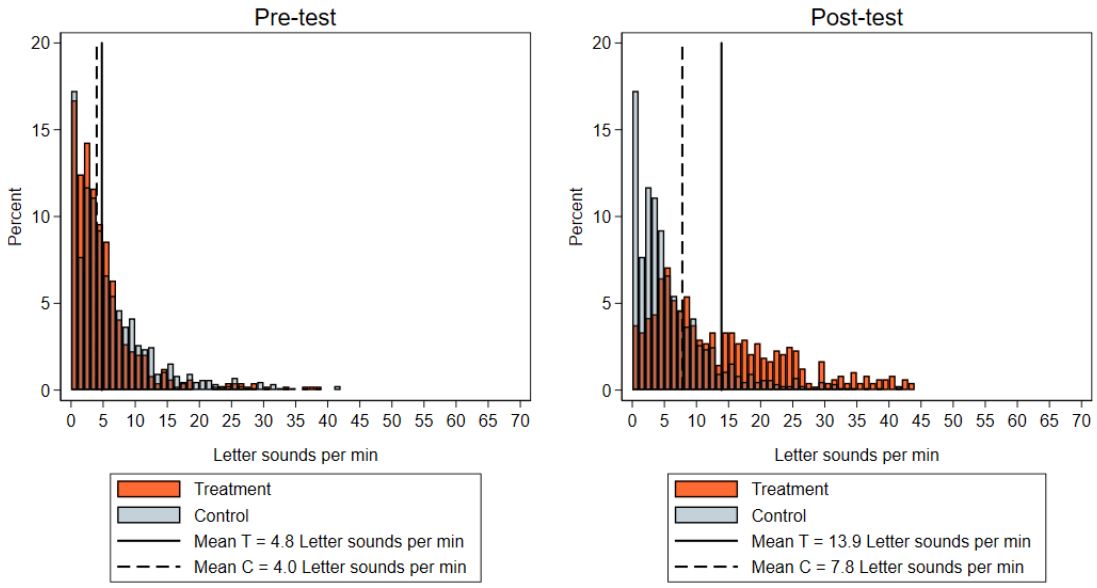
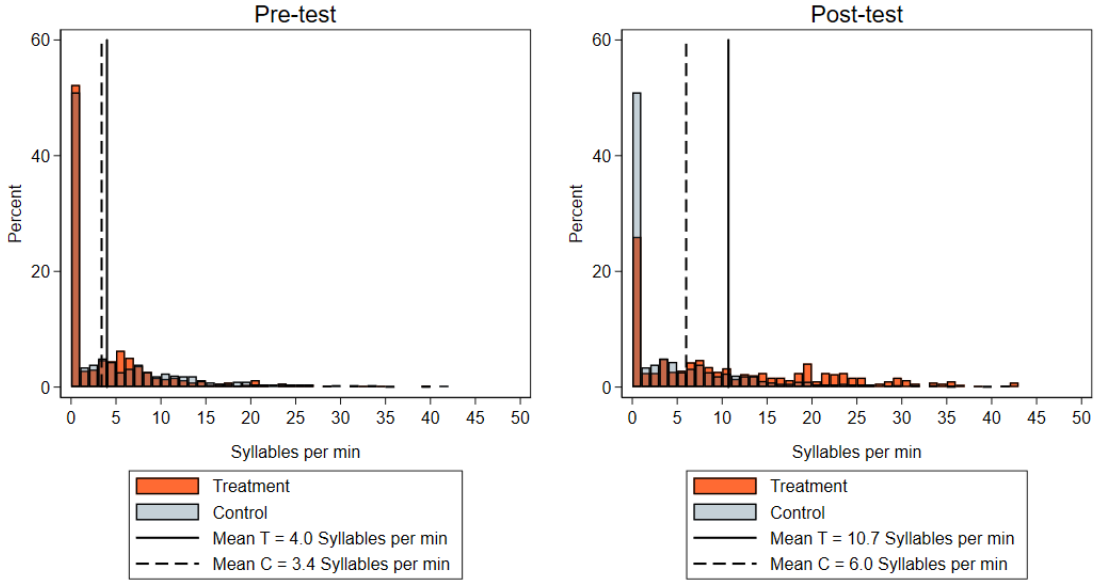


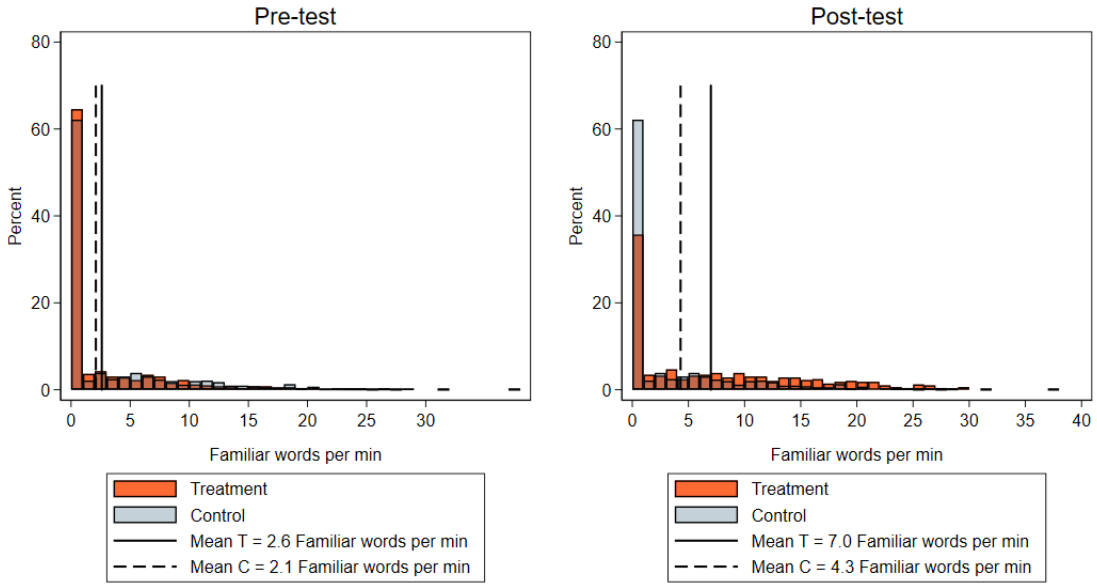

Figure 3 displays a similar histogram for syllable identification scores for both the treatment and control groups at baseline and again at endline. We find that syllable identification scores are skewed for both groups toward zero with both groups exhibiting somewhat similar baseline mean scores. By endline, we observe little change in the control group's distribution while that of the treatment group shifts to the right as exhibited by a lower percentage of zero scores and higher mean score.

**Figure 3. Distribution of syllable identification scores, by treatment status**



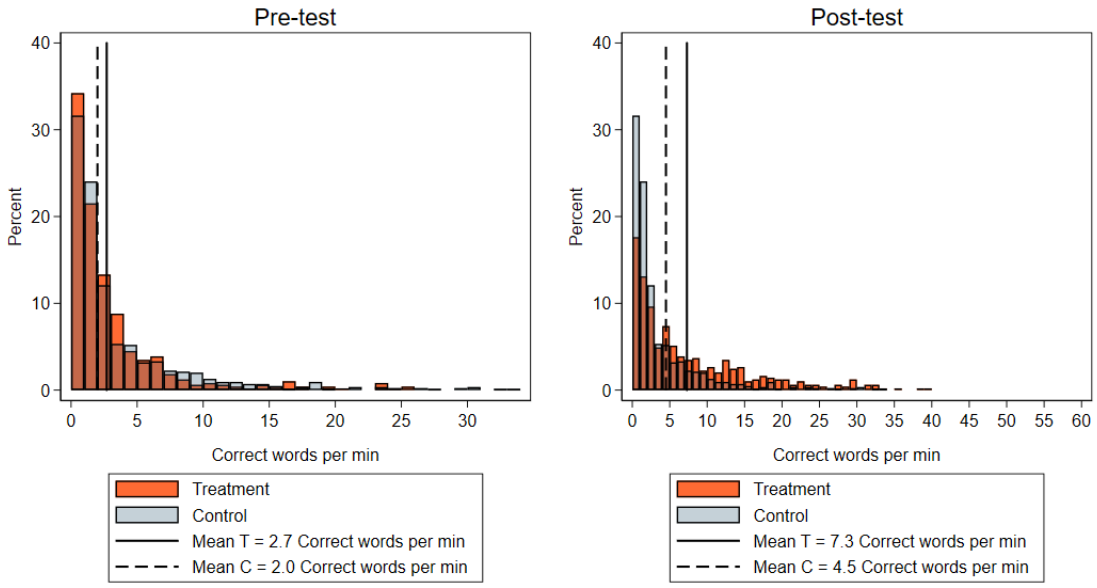
Next, we examine the distribution of familiar word reading among the treatment and control groups and plot their respective histograms in Figure 4. The modal score is again zero for both groups and is higher for the control group than the treatment group. The baseline familiar word score for the treatment group is only slightly higher at baseline but increases at a higher rate by endline.

**Figure 4. Distribution of familiar word scores, by treatment status**




Lastly, we present the distributions of oral reading fluency (ORF) scores for the treatment and control groups. As described earlier, students were given two reading passages. Students were given one minute to read the first reading passage, and two minutes to read the second reading passage. We then calculate the number of correct words per minute – i.e. oral reading fluency – for each of those passages and label then as "ORF scores – 1 minute" and "ORF scores – 2 minutes" respectively. The distribution of these ORF scores are displayed in Figure 5 and Figure 6, respectively.

**Figure 5. Distribution of ORF scores – 1 minute, by treatment status**




In both figures, we observe the same pattern as in the other timed subtasks, however, we see that the distribution of scores for ORF-2 minutes has a shorter right tail for both treatment and control groups. This might suggest that on average both passages yield similar results, but the 2-minute administration yields a narrower range of scores around the mean for the children attempting the passages.

**Figure 6. Distribution of ORF scores – 2 minutes, by treatment status**

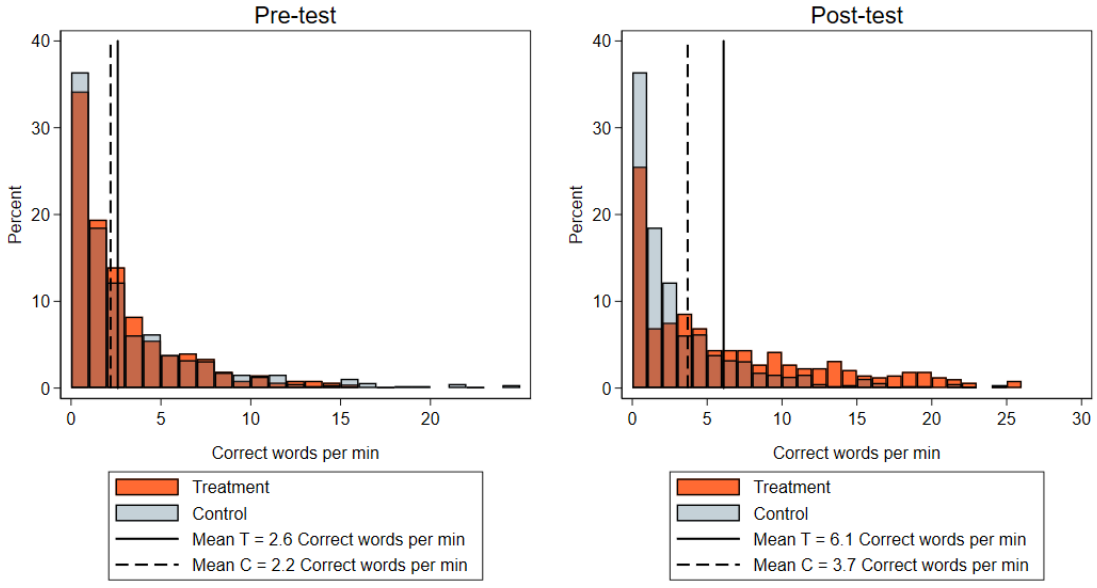
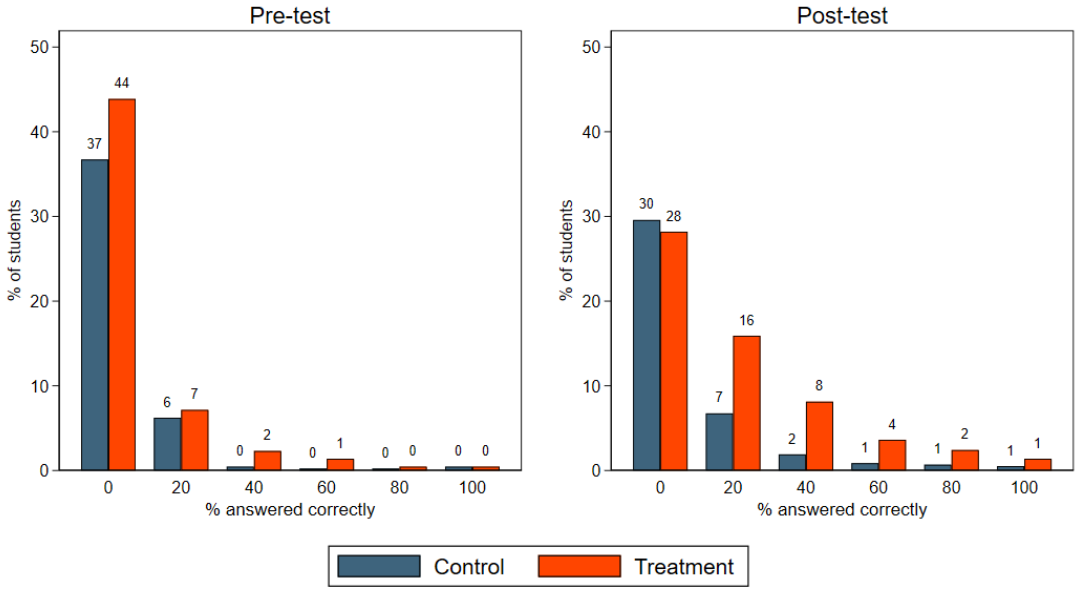



Figure 7 plots the distribution of students in both the treatment and control groups by reading comprehension score (as a reminder, the comprehension questions were only asked for the

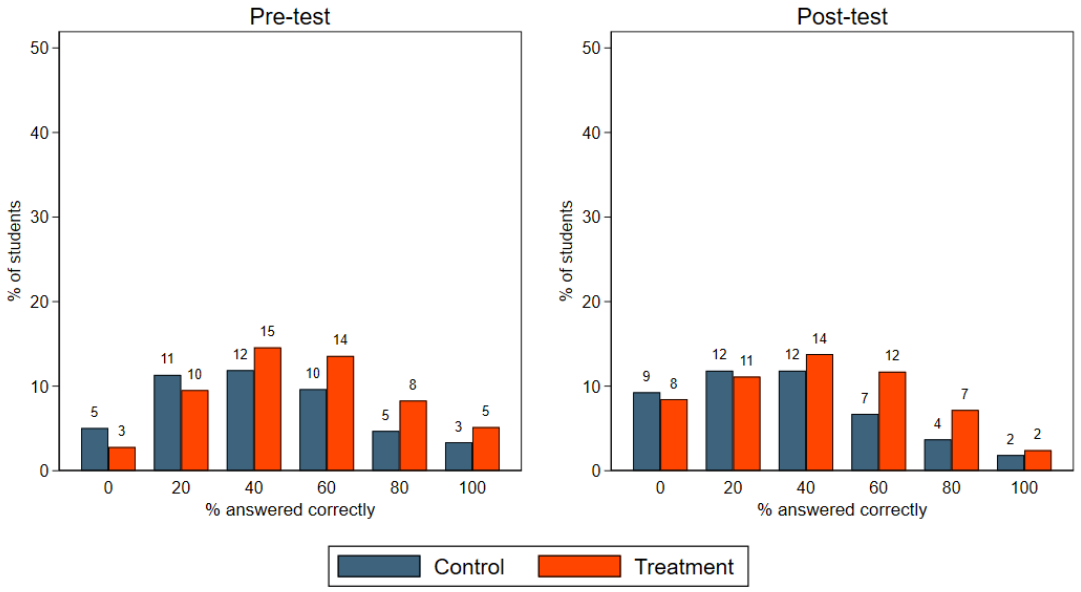
reading passage that was timed at 2 minutes). At baseline, the treatment and control groups display similar distributions with the majority of students receiving a score of zero. By endline, however, even though the modal score is still zero for both groups, the treatment group exhibits a substantial shift in the distribution toward higher reading comprehension scores. On the other hand, the control group distribution does not exhibit much change between baseline and endline.

**Figure 7. Distribution of Reading Comprehension scores, by treatment status**




Lastly, Figure 8 plots the distribution of students based on their listening comprehension scores at baseline and endline, respectively. In this case, we see that the distributions are clustered away from zero and more toward the 40%-60% range. Additionally, from examining the graphs we do not observe any noticeable shifts or changes in the overall distribution of scores for both the treatment and control groups.

**Figure 8. Distribution of Listening Comprehension scores, by treatment status**




## C1.4 Comparison of EGRA benchmarks

After calculating EGRA scores, we classify students based on performance benchmarks based on the outcomes of the standard and benchmark setting workshop conducted with MOE for each of the 7 subtasks identified in this report. The Grade 1 cutoff-scores for each subtask are shown in Table 10. The full benchmarks are presented in Annex III of this report.

**Table 10. Performance benchmarks for reading (Malagasy), Grade 1**

| Component skills | | how to measure it | Performance Categories | | | | Benchmark |
|---|---|---|---|---|---|---|---|
| | | | Beginner | Emergent | Proficient* | Fluent | |
| Comprehension | Listening comprehension | % of information that a child can extract from a text that is read aloud to him. The evaluation is based on the child's response to comprehension questions asked by an evaluator on the text. | 0 to 20% | 20 to 60% | 80% | 100% | 80% |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Reading comprehension | % of information that a child can extract from a text a his/her level (level 4) that he/she reads for the first time. The evaluation is based on the child's response to comprehension questions asked by an evaluator on the text. | 0 to 20% | 20 to 60% | 80% | 100% | 80% |
| Reading of a connected text | Fluency in reading of a connected text | Number of words correctly read aloud in a connected text of a text a his/her level (level 4) in the period of one minute | 0 to 5 wcpm | 6 to 14 wcpm | 15 to 30 wcpm | 31+ wcpm | 15 wcpm |
| Familiar word reading | Fluency in reading familiar words | Number of familiar words correctly read aloud in one minute. | 0 to 5 wcpm | 6 to 14 wcpm | 15 to 30 mlcm | 31 + wcpm | 15 mlcm |
| Alphabetic awareness | Fluency of letter and syllable reading | Number of letters, consonant clusters or syllables correctly read aloud in one minute. | 0 to 14 lspm | 15 to 39 lspm | 30 to 44 lspm | 45+lspm | 30 lspm |

**NB. "Proficient" means that students have hit the benchmark.*

Similar to the analysis in the previous section, we find similar patterns in the findings across all the subtasks (Figure 9 to Figure 13). Across all subtasks, except listening comprehension, we find that the modal performance category is 'beginner.' This mimics the fact that the modal score across most subtasks is zero, followed closely by low scores that are close to zero. Figure 9 plots the distribution of students on letter sounds by performance categories at baseline and endline, separately. At baseline we can see that students from the treatment and control groups are similar at baseline, whereas at endline we observe a noticeable shift in the treatment group toward emergent, proficient, and fluent categories. By MOE standards, students in proficient and fluent categories have hit the benchmark.

**Figure 9. Distribution of letter sound scores, by performance category and treatment status**

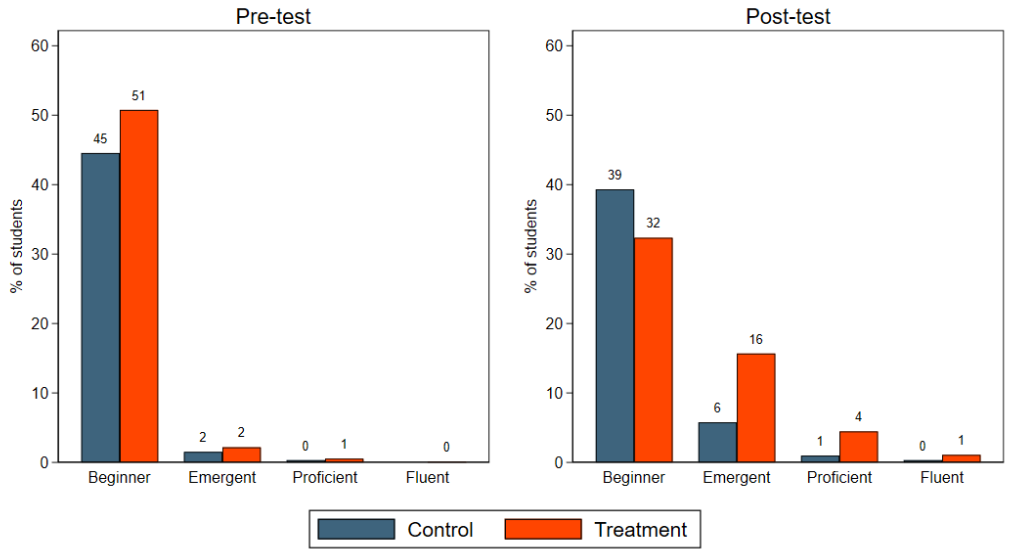
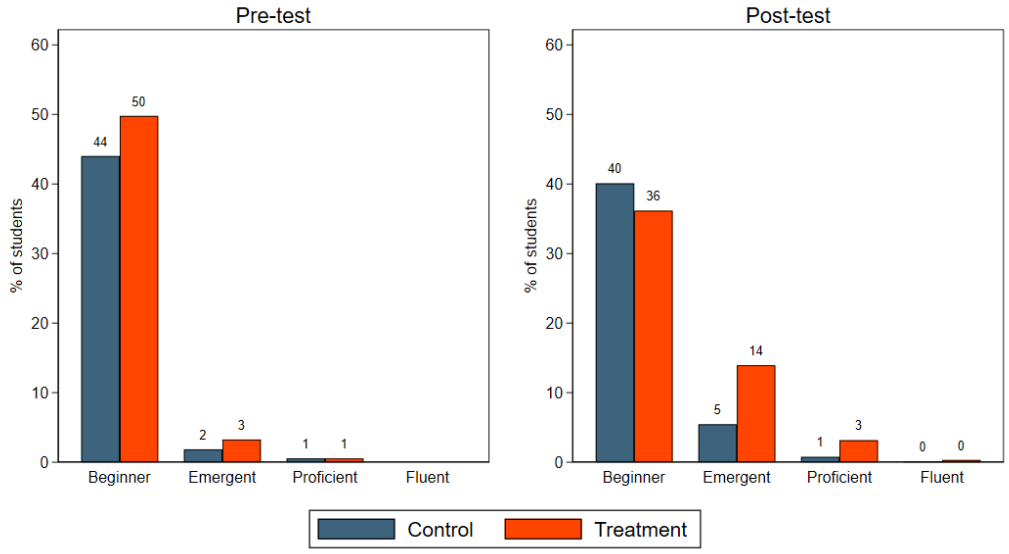



Figure 10 displays the distribution of student performance on syllable identification by performance category and by treatment status, in both the baseline and endline assessments. Again, we observe the same shift in the distribution for the treatment group, where the percentage of students who were in the beginner group declines to 36 percent, which means that 14 percent of students moved into emergent and proficient status.

**Figure 10. Distribution of syllable identification scores, by performance category and treatment status**




We plot the distribution of students across performance categories for familiar word reading in the treatment and control group, at baseline and endline. In this case, we see that about 16 percent of students in the treatment group improving from beginner to at least emergent. We

also see the percentage of students identified as emergent or proficient increase from 12 percent to 27 percent.

**Figure 11. Distribution of familiar word scores, by performance category and treatment status**

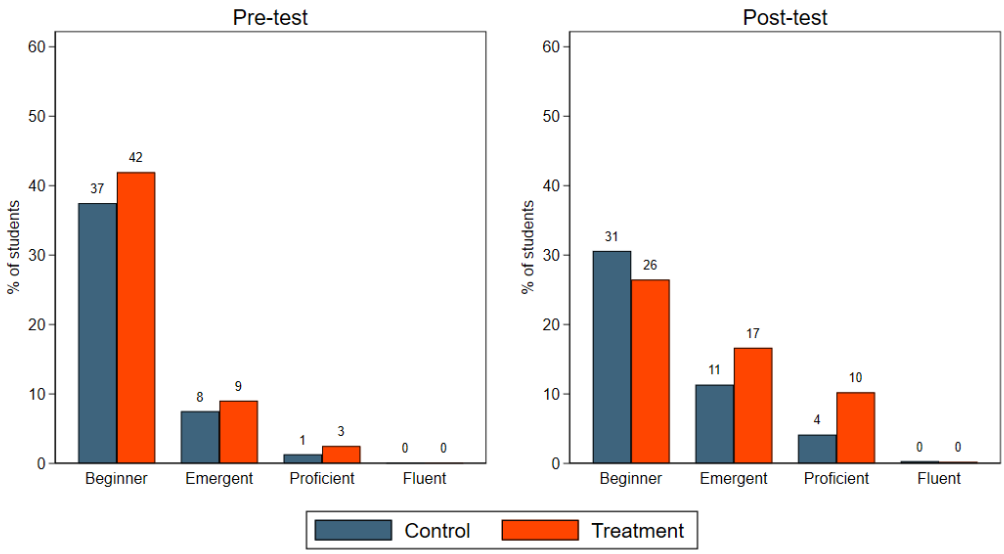



Figure 12 and Figure 13 display the distributions of students based on their performance on the 1- and 2- minute administrations of the oral reading fluency subtask. Between the two methods of administering oral reading fluency, we find a congruent pattern between both distributions. Again, the modal category is beginner reader. However, we observe a 16 percentage point shift from beginner to at least emergent from baseline to endline among students in the treatment schools, and an overall increase of 7 percentage points in the proportion of proficient and fluent students.

**Figure 12. Distribution of oral reading fluency (1 minute) scores, by performance category and treatment status**

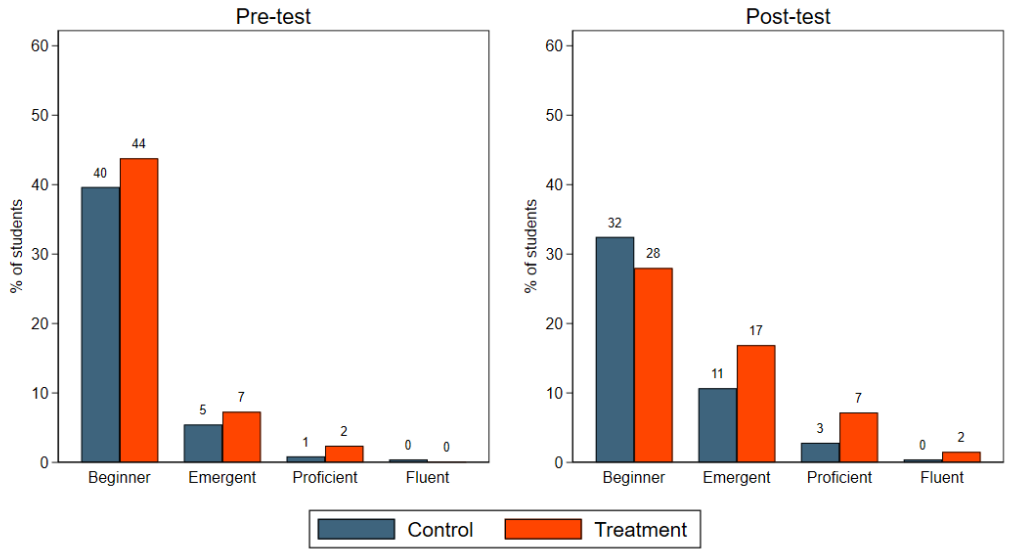
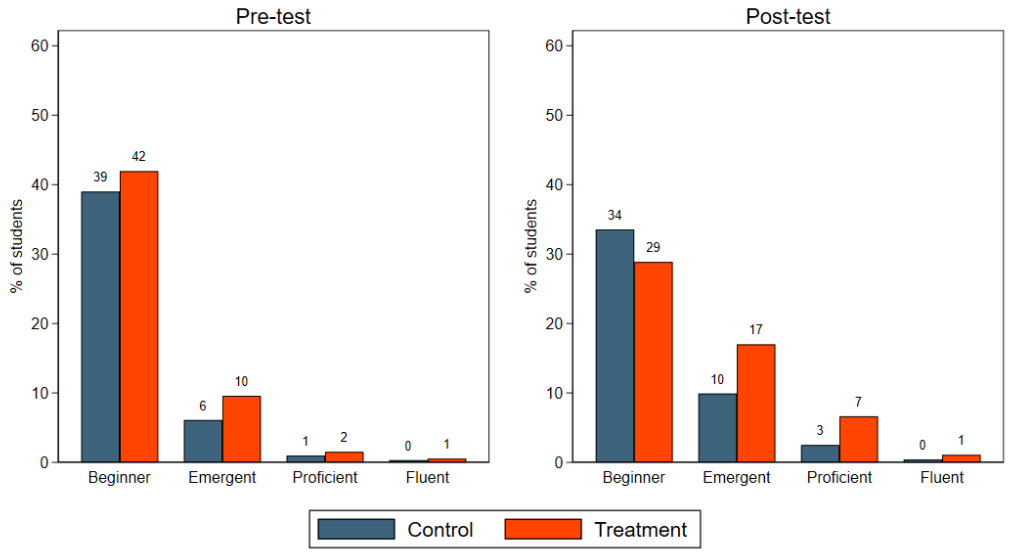



**Figure 13. Distribution of oral reading fluency (2 minute) scores, by performance category and treatment status**



## C2. FIDELITY OF IMPLEMENTATION (FOI) ANALYSIS

**Key Findings**

- **An important part of the design of the intervention is lesson scripting with specified steps for each activity**
  - The intervention had a high level of fidelity in this regard as 83 percent of the sampled teachers were following most or all the steps associated with the lesson activities
- **Students whose teachers followed the steps achieve higher EGRA scores in all subtasks but the listening comprehension subtask**
  - On average, students whose teachers followed the steps are able to identify 12.5 letter sounds per minute relative to 6.9 letter sounds per minute among students whose teachers did not follow the steps
  - Syllable identification scores are higher for the first group by 3.6 syllables per minute
  - Again, the group exposed to completion of most or all lesson steps outperforms their counterpart in oral reading fluency by 1.2 to 2.1 correct words per minute
  - Reading comprehension scores are also higher by 19 percentage points (equivalent to an additional question answered correctly out of 5)
- **Only one third of schools in the treatment group use the local dialect when teaching reading in Malagasy, while the rest does not**
  - We find little to no correlation between use of the local dialect and EGRA performance on the letter sound, syllable sound, and comprehension subtasks
  - Classrooms that incorporate the local dialect underperform their counterparts slightly in terms of familiar word reading and oral reading fluency
  - This might suggest that there is selection at the school level in terms of where the teacher choose to incorporate the local language

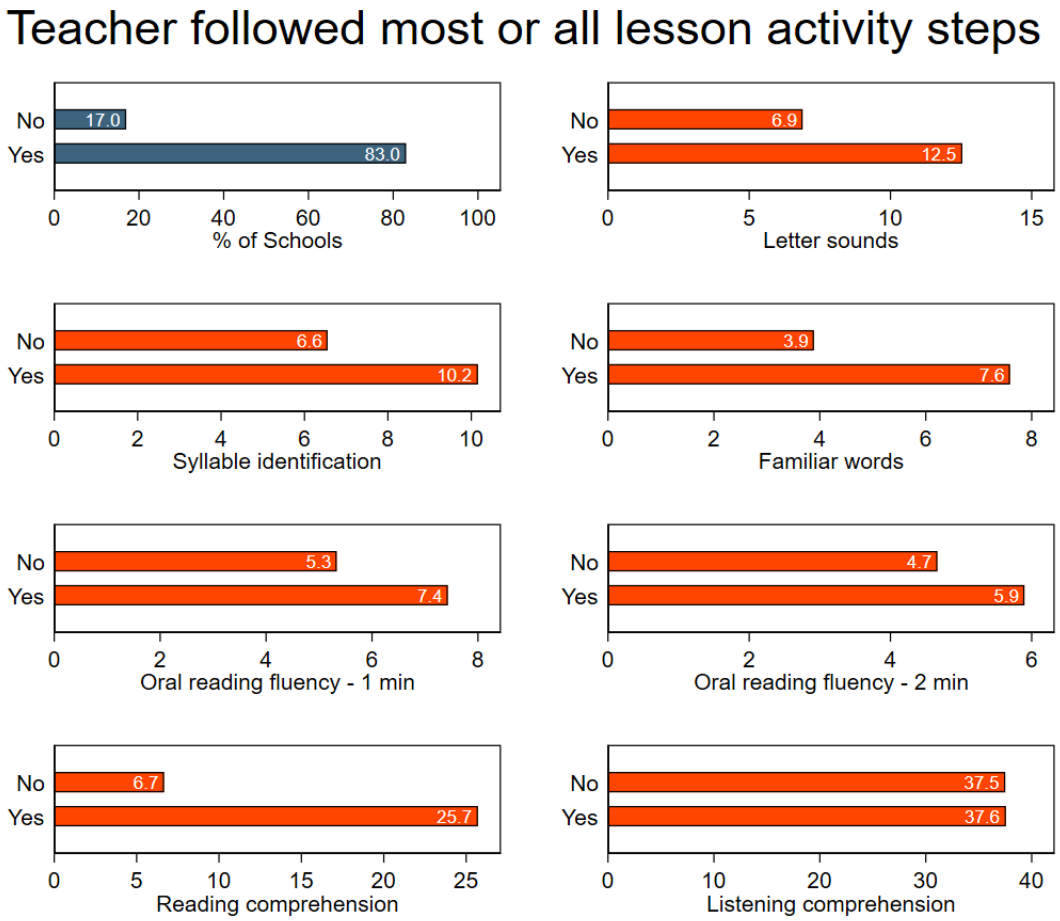
In this section, we explore whether there is a treatment dosage effect by using fidelity of implementation (FOI) as a proxy measure. In other words, we explore whether teachers who implement the project to a fuller extent achieve better student learning outcomes that those who do not. In particular, one important aspect of the MMT program concerns the scripted lesson plans and the focus of teacher trainings in following the steps outlined in each lesson plan. The MMT reading and writing program in Official Malagasy was designed to help Grade 1 teachers to understand and apply effective instructional practices and strategies and to

maximize students' reading and writing practice time. To do so, MMT provided teachers with a structured and scripted teachers' guide which included daily lesson plans for all reading, writing, oral language development and drawing lessons, daily continuous evaluation, and proposed remedial activities for students who did not meet lesson objectives. In the trainings, teachers were taught how to use the guide and student book effectively and to follow the lesson steps as presented in the guide. A teacher who implemented most or all steps of the lesson plan is considered to have implemented the MMT teacher guides with more fidelity (i.e. "higher dosage") than a teacher who did not. Because FOI is used as a proxy measure of treatment dosage, we only collected this information from treatment schools. The FOI data used in the following analysis was collected through classroom observations shortly before the endline EGRA data collection.

Furthermore, all teaching and learning materials were written in Official Malagasy as per the language policy. However, for the purpose of informing the MOE of language use, we sought to capture whether or not teachers employed dialectical variants of Malagasy in their lessons in addition to Official Malagasy. In this section, we also explore whether using the dialectical variant of Malagasy during the reading or writing lessons to improvements in students' reading ability.

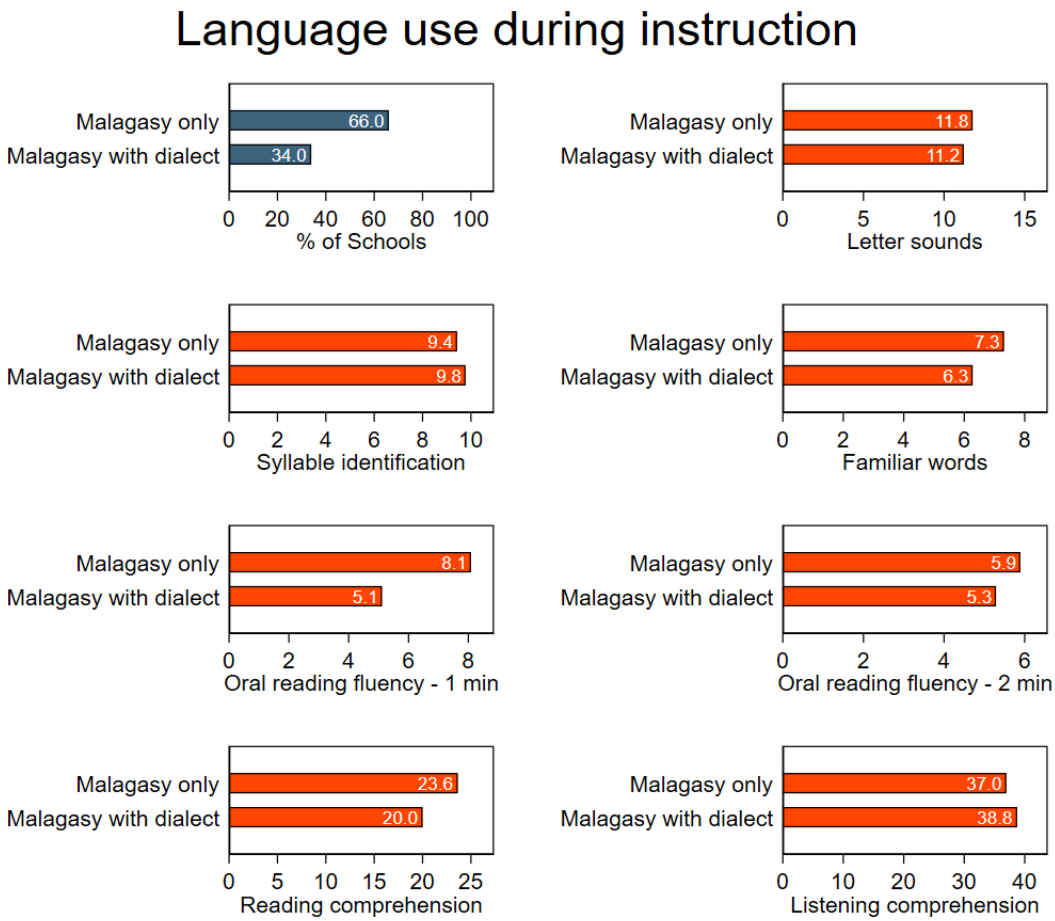
The blue graph (top left) in Figure 14 shows the distribution of teachers who were observed that completed most or all of the activity steps. Overall, we find about 83 percent of teachers followed most or all activity steps, while the rest are considered to have completed none or only some of the steps as judged by the enumerator administering the lesson observation. In terms of correlations with EGRA, we find that students whose teacher follows most/all activity steps outperform their counterparts whose teacher does not across most subtasks. The most notable difference can be seen in reading comprehension.

**Figure 14. Lesson activity steps, school distribution and average EGRA scores**




Next, we examine language use in the classroom to determine whether the reading teacher uses a dialectical variant of Malagasy in addition to Malagasy to teach reading or teaches reading without the use of a dialect. We find that about a third of teachers in the treatment group use a dialectical variant of Malagasy when teaching reading in class. Moreoever, we find that, on average, EGRA scores across most subtasks are similar between both groups of teachers, except for the 1-minute oral reading fluency and reading comprehension where the group that does not use a dialect in teaching have higher scores.

**Figure 15. Language use in teaching reading, school distribution and average EGRA scores**




The differences in test scores that we find in this analysis are not necessarily attributable to the fact that teachers followed the steps or used a local dialect in teacher reading because teachers who chose to follow the steps or use a dialect did not do so at random. That is, teachers who choose to implement the MMT curriculum with a high degree of fidelity might share certain background and/or unobservable characteristics that are fundamentally different from those teachers who do not follow the steps. For instance, about 20 percent more likely to be low SES, meaning that a major confounder of language use is SES as well as other correlates of poverty.

At face value, we find that students whose teachers follow the steps with a high level of fidelity have higher mean scores than those who do not. Whereas in terms of language use, we find that the schools that do not use a dialect have better ORF 1-minute and reading comprehension scores. As such, we rely on regression analyses to account for potential differences in the student composition among those who follow the steps and those who do not.

## C3. REGRESSION RESULTS

**Key Findings**

- **The MMT intervention showed significant gains in improving reading outcomes, overall and across most student subgroups. Teachers who followed most or all steps of the lesson plans had better learning outcomes than teachers who did not.**
  - The program showed effect sizes of .35 and .37 on letter and syllable identification, respectively
  - Familiar words increased by .22, and oral reading fluency improved by .28-.31 depending on the duration of the reading passage
  - Listening comprehension scores improved by 5.7 percentage points
  - Gains in reading comprehension were modest with an increase of only 2.1 percentage points
  - Boys and girls were impacted similarly by the program in terms of their overall EGRA scores
  - The program was more beneficial to students in the high SES group
- **Results were more mixed at the extensive margin**
  - Students in the treatment group had a lower probability of receiving a score of zero in syllable identification, familiar word reading, and the 2-minute oral reading fluency subtasks
  - The program showed gains in the 1-minute administration among boys and students in the low SES group
  - Students in the treatment group had higher probabilities of receiving zero scores on letter sounds than the control group
- **The intervention exhibited more consistency in increasing the proportion of students with minimal proficiency across all subtasks**
  - The largest effects were in familiar words followed by both the 1- and 2-minute administrations of oral reading fluency by 9.6 and 5.1-7.1 percentage points
  - The most positively affected student subgroups were girls and students in the high SES group
  - The impacts on letter and syllable sounds were both positive with little variation in terms of the effects on gender and SES groups

The final stage presents the results from the regression analyses where we estimate program impacts using a difference-in-differences (DD) approach. The DD approach enables us to identify

the impact of the program/treatment by estimating the change in EGRA scores from baseline to endline for the treatment group relative to that of the control group, while accounting for differences in observable student characteristics. This means that the DD estimate measures the average gains made by students in the treatment group over and above any gains made by the control group. For example, if the control group exhibited learning gains of 4 words per minute from baseline to endline and the treatment group exhibited a 10 word per minute gain, then the program impact is +6 words per minute.

The main advantage of the DD approach is that treatment assignment does not always have to be random to produce causal estimates. To support causality in such a framework, we assume that both treatment and control schools exhibit parallel trends, even well before the intervention has started. Testing this assumption would require multiple pre-treatment periods of data collection—which in this case are not available. One method to circumvent a lack of adequate pre-treatment data is randomization of the treatment. This ensures that both groups should have similar pre-treatment trends as well as post-treatment trajectories in the absence of the treatment.

As we have shown earlier, even though assignment of the treatment has been randomized, the samples do not balance perfectly between treatment and control groups in terms of certain observable student background characteristics and baseline EGRA performance. The randomization of the treatment assignment ensures that students (or their parents) did not self-select, or get sorted, into attending the treatment schools. Further, the inclusion of student covariates to account for the differences in observed characteristics ensures that our impact estimates are free from selection bias. However, as discussed earlier, the generalizability of these findings to the larger population may be questioned as sample attrition was non-random and may apply only to students with similar characteristics as those in the final analytic sample.

This impact is thus identified on the assumption that the treatment and control group would have behaved similarly had the program not been implemented. Additionally, we include measures of treatment dosage using proxies for program FOI. These are covariates that are measured for the treatment group only in the post-treatment period. As such, the inclusion of the FOI parameters enables us to measure the base program impact and the incremental impacts from implementing the program at a higher level of fidelity.[7]

For the purposes of this impact evaluation we restrict the FOI analysis to include only whether a teacher followed most or all lesson activity steps.[8] This means that the regression output tables will show the impact of the treatment when only some or none of the activity steps were completed by the teacher (this will be called the base treatment effect throughout the rest of the report) and the impact of completing most or all steps is added on to the base treatment effect. This means that if the base treatment effect is +3 words per minute and the incremental effect of completing all steps is +4, then the treatment effect from completing all steps is +7 words per

[7] Note that a score of zero on any FOI measure is not equivalent to a school receiving no treatment as these measures only proxy for the actual components of the intervention.

[8] We also explored other measures FOI by constructing FOI indices. However, all other FOI indices were not found to be statistically significant determinants of student EGRA performance. All regression outputs that include all FOI indices are included in Annex IV.

minute. Lastly, we measure program (and fidelity) impacts on girls, boys, low SES, and high SES students, separately, enabling us to test for potential heterogeneity in the treatment effect.

### C3.1 Timed Subtask Scores

The first set of results from the DD analyses represent the impact estimates of MMT on timed subtask scores and are presented in Table 11. The first panel of Table 11 shows that **average letter sounds per minute (lspm) scores grew by about 2.73 lspm more for the treatment group between baseline and endline relative to the control group We also find that students whose teachers do follow most or all steps score an additional 3.36 lspm.** This means that relative to the control group, students whose teachers followed most steps scored 6.09 lspm higher (2.73+3.36 lspm). **The overall average effect for the letter identification subtask is thus 3.24 letter sounds per minute (weighted average of the base treatment effect and the effect for teachers who implemented most/all lesson steps).** In terms of differential impacts between genders, we find that boys and girls are affected similarly by the program, thus the program has no equalizing or dis-equalizing effect. However, we find that the program had a larger effect on students in the high SES group relative to the low SES group. We estimate a base treatment effect of 2.9 lspm, and 4.09 lspm with completion of most/all steps on the low SES subgroup, both estimates are not statistically significant, however. Whereas the equivalent effect for the high SES group is 1.64 lspm and 6.72 lspm, respectively, meaning that the SES gap was exacerbated through the intervention.

For syllable identification scores, when the teacher does not follow all steps, however, we find no significant effect of the program ranging between .28 and 1.23 syllables per minute (spm) depending on the subgroup. Whereas **when the teacher follows most or all steps in the lesson activity students in the treatment schools gained an additional 4.65 spm relative to students in the control group, on average**. **The average treatment effect for syllable identification is 3.21 syllables per minute.** We find similar effects on boys and girls of about 4.72-4.73 spm. Again, we see that the program was more effective on the high SES student group relative to the low SES group. High SES students whose teachers complete the lesson steps exhibit gains of about 6.3 spm over the control group, and the low SES group gain is 2.73 spm over the same group. The pattern repeats when investigating **the effects of MMT on familiar word reading where students whose teachers complete most or all steps gain between 1.43 and 2.84 words per minute relative to the control group**, depending on the subgroup. Whereas, students whose teachers do not follow most steps exhibit low to no gains in familiar word reading. **The average treatment effect for familiar word reading is 1.42 correct words per minute.**

In terms of oral reading fluency, we find little difference in administering the 1- or 2-minute versions of the subtask in terms of program effects. Again, we estimate that the program is more successful when the teacher follows most or all steps in the lesson activities. Only for the low SES group we find that the group whose teachers do not follow most steps outperform the students whose teachers do. However, **overall, students whose teachers follow most or all steps in the lesson activities outperform their control group counterparts by between 1.43 and 3.29 correct words per minute**. We also find that boys and girls perform similarly in the MMT schools while low SES students are less successful than the high SES students. Finally, **we do not find a statistically significant relationship between the program and**

**reading comprehension outcomes with modest effect sizes for most subgroups except the high SES group who exhibit close to zero gains**. In terms of listening comprehension, we find that the group whose teacher does not follow most lesson steps underperform relative to the control group with a decline in scores by between -8.7 and -13.76 percent, although only the full sample effect is statistically significant. However, the group whose teachers do follow at least most lesson steps perform similarly to the control group, i.e. the program has no effect in this situation.

**Table 11. Regression results – MMT marginal effects on timed subtasks**

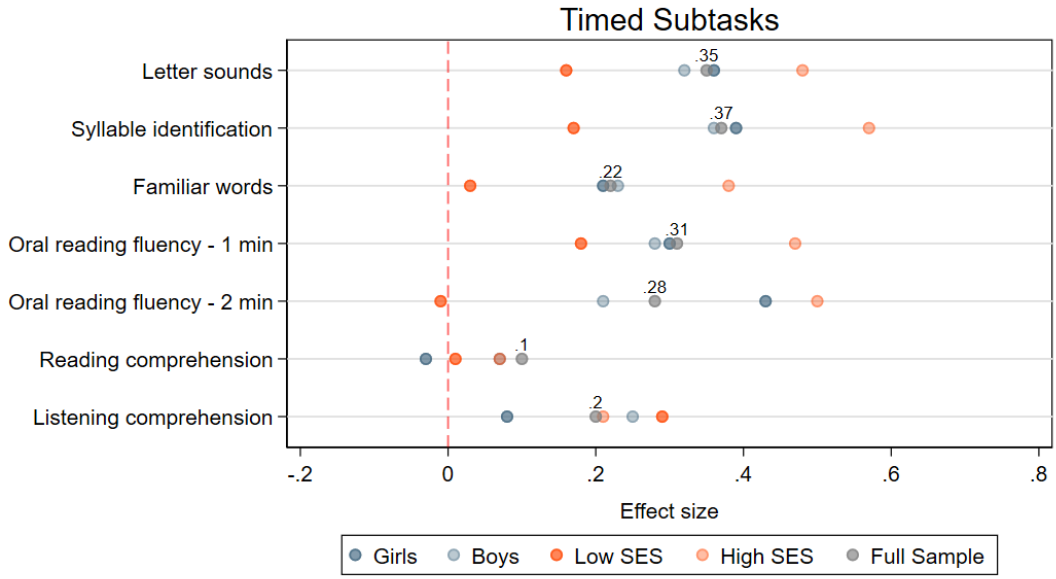
| **Letter sounds** | **Full Sample** | **Girls** | **Boys** | **Low SES** | **High SES** |
|---|---|---|---|---|---|
| Treatment Effect | 2.73*** | 2.72* | 3.13** | 2.90 | 1.64 |
| Teacher followed most/all activity steps | 3.36*** | 3.49** | 2.92** | 1.19 | 5.08** |
| *Observations* | *1,776* | *805* | *968* | *1,017* | *756* |
| **Syllable identification** | | | | | |
| Treatment Effect | 0.91 | 0.81 | 1.15 | 1.23 | 0.28 |
| Teacher followed most/all activity steps | 3.74*** | 3.92*** | 3.57*** | 1.50 | 5.97*** |
| *Observations* | *1,777* | *806* | *968* | *1,018* | *756* |
| **Familiar words** | | | | | |
| Treatment Effect | 0.81 | 0.67 | 0.95 | 1.60 | -0.16 |
| Teacher followed most/all activity steps | 1.56*** | 1.47 | 1.57** | -0.17 | 3.00** |
| *Observations* | *1,772* | *801* | *968* | *1,015* | *754* |
| **Oral reading fluency - 1 min** | | | | | |
| Treatment Effect | 0.24 | 0.30 | 0.34 | 0.45 | -1.10 |
| Teacher followed most/all activity steps | 2.41*** | 2.35* | 2.13*** | 1.34 | 3.98*** |
| *Observations* | *1,767* | *800* | *964* | *1,011* | *753* |
| **Oral reading fluency - 2 min** | | | | | |
| Treatment Effect | 1.07** | -0.17 | 1.83*** | 1.91* | 0.44 |
| Teacher followed most/all activity steps | 1.43** | 2.57** | 0.84 | -0.48 | 2.85* |
| *Observations* | *1,734* | *776* | *955* | *991* | *740* |
| **Reading comprehension** | | | | | |
| Treatment Effect | 7.50 | 9.95 | 6.74 | 14.93 | -2.45 |
| Teacher followed most/all activity steps | 0.84 | -2.98 | 0.19 | -3.11 | 2.39 |
| *Observations* | *966* | *456* | *509* | *503* | *462* |
| **Listening comprehension** | | | | | |
| Treatment Effect | -9.33** | -8.57 | -8.66 | -11.47 | -13.76 |
| Teacher followed most/all activity steps | 9.19** | 4.88 | 10.66* | 12.75 | 10.34 |
| *Observations* | *1,699* | *774* | *923* | *954* | *743* |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows.
* p<.1, ** p <.05, *** p<.01

For ease of interpretation, we compute the overall treatment effect for the full sample as well as the gender and SES strata in terms of effect sizes, measured in standard deviations. We plot the computed effect sizes for each of the subtasks and plot them in Figure 16. We see that **the**

**average effect sizes for all subtasks except reading comprehension are large in magnitude and range between .20 and .37 SD**, for the full sample, with the largest estimated effect being in syllable identification. These are remarkable effect sizes given the short implementation period of three months. The average treatment effect on reading comprehension is about .1 SD and is negative for girls.

**Figure 16. Summary of effect sizes, by subtask and student subgroup**




## C3.2 Zero Scores

In this subsection, we replicate the analysis in Table 11 but looking at the probability of receiving a score of zero on each of the five timed subtasks. Table 12 is structured similarly to Table 11 without the inclusion of reading or listening comprehension. Listening comprehension is excluded from this analysis as the extensive margin includes under 13 percent of the total sample and reading comprehension zero scores are the same as oral reading fluency zero scores since students qualify to take the reading comprehension subtask if they can read at least one word. As such, analysis of reading comprehension zero scores adds no value over the analysis of ORF zero scores. Overall, we find that **students in the treatment group had a lower probability of receiving zero scores on all subtasks among both the group whose teachers did and did not follow most or all the lesson steps.**

**On letter sounds, syllable identification, and familiar word reading, we estimate that the program lowered the students' probability of receiving a score of zero by between -3.96 and -13.93 percentage points, with the largest effect on syllable identification zero scores**. However, in terms of letter sounds, we find that the group who did not follow all or most of the lesson steps perform similarly or better than the group that did. In terms of syllable identification and familiar word reading, students whose teacher completed most or all lesson steps significantly outperformed the control group and to a lesser extent the treatment group whose teachers did not complete the steps. Again, we find that **program impacts were not heterogeneous in terms of gender and that the impact estimates**

**were larger for the high SES group than the low SES group across the letter sound, syllable identification, and familiar word subtasks**.

The MMT program, however, does not show significant effects in terms of lower zero scores in oral reading fluency. We find that when the teacher does not follow most or all lesson steps, their students' probability of being able to read at least one word does not improve, and in some cases, it decreases. The group whose teacher follows at least most lesson steps, show more improvement in that regard than their control group counterpart, except for girls in the 1-minute administration of the ORF subtask where we find a slight increase in the probability of receiving a score of zero. From an equity standpoint we find that boys were more positively affected by the program in reducing their probability of receiving a score of zero (equivalently, the probability of being able to read at least 1 word) than girls. We find more mixed results in terms of SES, where the high SES group performs better than the low SES group in the 2-minute administration and the low SES group performs relatively better than the high SES group in the 1-minute administration.

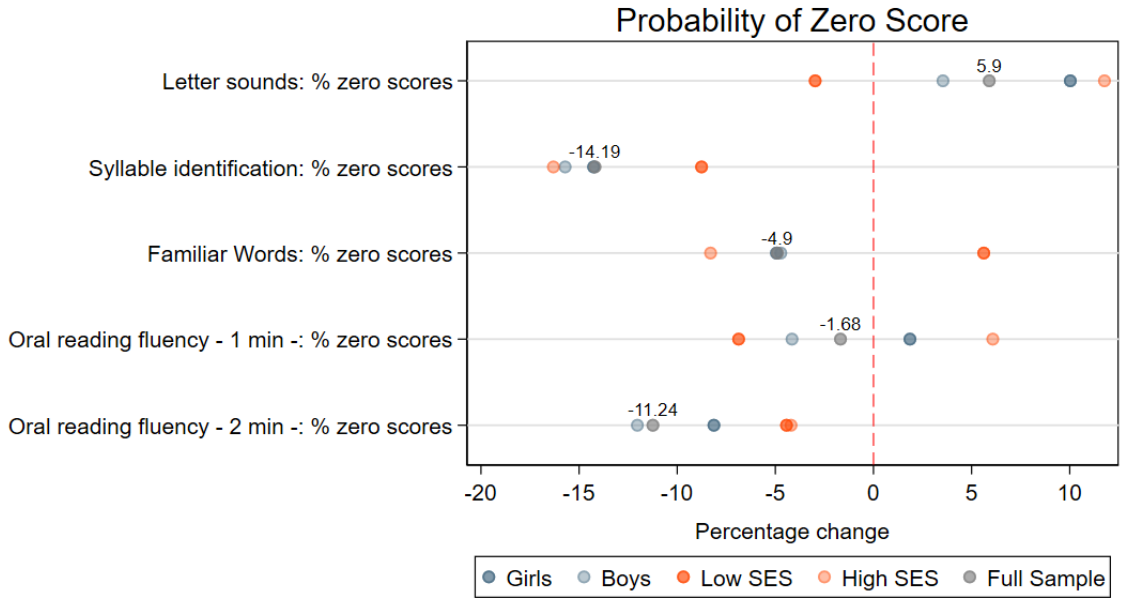
**Table 12. Regression results – MMT marginal effects on the probability of receiving a zero score**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Letter sounds: % zero scores** | | | | | |
| Treatment Effect | -14.64** | -14.09 | -15.48** | -1.43 | -22.17** |
| Teacher followed most/all activity steps | 10.68* | 15.63 | 7.96 | -3.33 | 19.66* |
| *Observations* | *1,776* | *805* | *968* | *1,017* | *756* |
| **Syllable identification: % zero scores** | | | | | |
| Treatment Effect | 4.63 | 3.00 | 6.16 | 4.91 | -3.38 |
| Teacher followed most/all activity steps | -18.56*** | -18.27** | -20.79*** | -11.93 | -19.31** |
| *Observations* | *1,777* | *806* | *968* | *1,018* | *756* |
| **Familiar Words: % zero scores** | | | | | |
| Treatment Effect | -4.90 | -3.81 | -5.42 | -13.01 | -10.59 |
| Teacher followed most/all activity steps | -4.89 | -5.20 | -4.56 | 9.95 | -7.77 |
| *Observations* | *1,772* | *801* | *968* | *1,015* | *754* |
| **Oral reading fluency - 1 min -: % zero scores** | | | | | |
| Treatment Effect | -0.25 | 1.45 | -1.78 | 10.95 | 0.24 |
| Teacher followed most/all activity steps | -2.02 | 1.96 | -4.70 | -11.01 | 7.44 |
| *Observations* | *1,767* | *800* | *964* | *1,011* | *753* |
| **Oral reading fluency - 2 min -: % zero scores** | | | | | |
| Treatment Effect | 7.15 | 8.57 | 3.92 | 12.96 | -2.94 |
| Teacher followed most/all activity steps | -15.52** | -12.02 | -15.73** | -8.48 | -4.49 |
| *Observations* | *1,734* | *776* | *955* | *991* | *740* |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows.
* p<.1, ** p <.05, *** p<.01

Figure 17 summarizes the findings from Table 12 by showing the overall average effect of the program for the full sample and student subgroups rather than showing the effect disaggregated by whether the teacher followed most or all steps in the lesson. We can see that the program was able to, on average, improve upon the probability of receiving a score of zero in syllable identification, familiar words, and the 2-minute administration of the oral reading fluency subtask, whereas we observe the proportion of zero scores worsening in terms of letter sounds.

**Figure 17. Summary of regression estimates, by subtask and student subgroup**




## C3.3 Minimal Proficiency

The final set of regressions assess the impact of the MMT program on the probability that a student achieves minimal proficiency as defined by the performance benchmarks presented in Table 10. Table 13 presents the regression coefficients of the base treatment effect and the incremental effect from teachers following most or all steps for the lesson activities. Like the previous subsection, we employ a linear probability model to estimate program impacts on the probability of achieving proficiency.

**Across all subtasks we observe a consistent pattern where, on average, the treatment group outperforms the control group except for syllable identification**. We also consistently find that within the treatment group, students whose teachers follow at least most of the lesson steps outperform their control group counterparts as well as the group whose teachers do not follow most steps. **We estimate that the MMT intervention increased students' probability of attaining minimal proficiency by 7, 4, and 11 percentage points on the letter sounds, syllable identification, and familiar word reading subtasks respectively when the teacher follows most lesson steps. When the teacher does not follow the steps, we do not find any statistically significant impact of the program.** Unlike the findings from the timed subtask regression analyses, we find that girls are slightly more affected at this margin than boys. Meaning that **girls are more likely to achieve minimal proficiency than boys because of the MMT intervention**. However,

the differences in program impacts between boys and girls are not statistically significant. Between low and high SES groups, we find that treated students in both groups are affected similarly by the intervention.

On oral reading fluency, students in the treatment group whose teachers follow most or all lesson steps have a higher probability of achieving minimal reading proficiency than similar students in the control group, and more so than students whose teachers did not follow the steps. Only in the low SES group we find that students in the treatment group, regardless of whether the teacher followed the steps, are almost equally likely to achieve proficiency. As in the case of the intermediate reading outcomes, we find that girls are more likely to achieve proficiency because of the program, but low and high SES students are equally likely to achieve proficiency.

**Table 13. Regression results – MMT marginal effects on the probability of receiving a zero score**

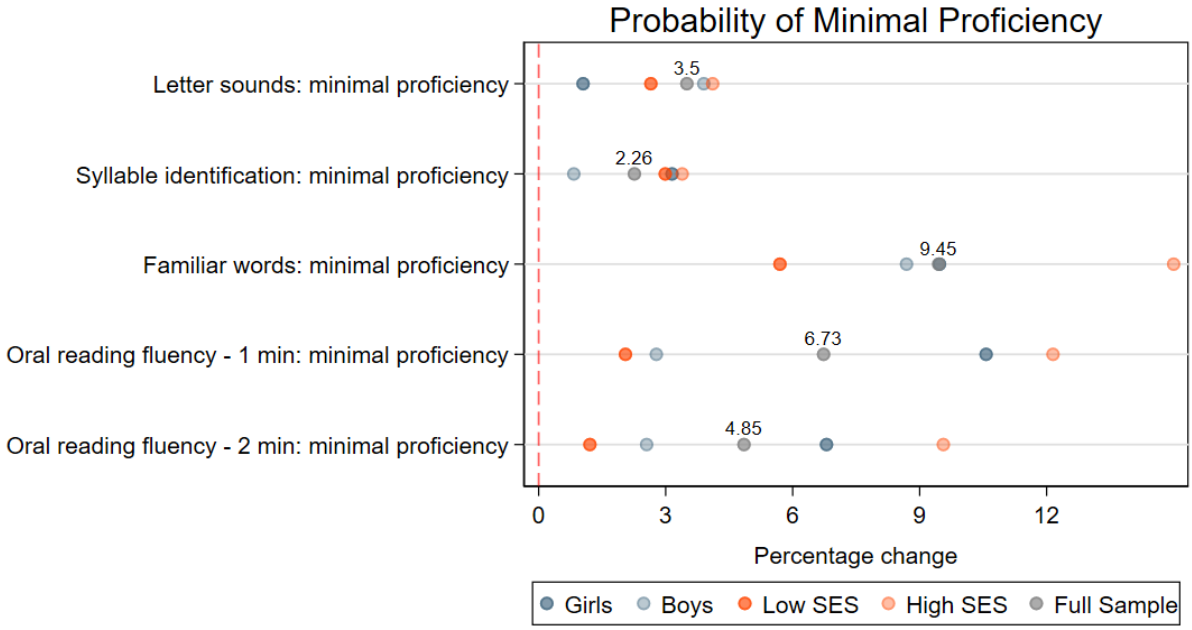
| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Letter sounds: minimal proficiency** | | | | | |
| Treatment Effect | 0.03 | 0.11* | 0.01 | 0.02 | 0.03 |
| Teacher followed most/all activity steps | 0.04 | -0.01 | 0.05 | 0.03 | 0.04 |
| *Observations* | *1,777* | *806* | *968* | *1,018* | *756* |
| **Syllable identification: minimal proficiency** | | | | | |
| Treatment Effect | 0.01 | 0.03 | 0.01 | -0.01 | -0.02 |
| Teacher followed most/all activity steps | 0.03 | 0.03 | 0.01 | 0.04* | 0.05 |
| *Observations* | *1,777* | *806* | *968* | *1,018* | *756* |
| **Familiar words: minimal proficiency** | | | | | |
| Treatment Effect | -0.01 | 0.02 | -0.02 | 0.05 | -0.10 |
| Teacher followed most/all activity steps | 0.12*** | 0.11* | 0.11*** | 0.06 | 0.21** |
| *Observations* | *1,777* | *806* | *968* | *1,018* | *756* |
| **Oral reading fluency - 1 min: minimal proficiency** | | | | | |
| Treatment Effect | -0.01 | -0.03 | 0.00 | 0.06 | -0.11 |
| Teacher followed most/all activity steps | 0.09*** | 0.14*** | 0.03 | 0.01 | 0.18** |
| *Observations* | *1,777* | *806* | *968* | *1,018* | *756* |
| **Oral reading fluency - 2 min: minimal proficiency** | | | | | |
| Treatment Effect | 0.01 | -0.00 | 0.02 | 0.06 | -0.07 |
| Teacher followed most/all activity steps | 0.06* | 0.08 | 0.03 | 0.00 | 0.13* |
| *Observations* | *1,766* | *801* | *962* | *1,008* | *755* |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows.
* p<.1, ** p <.05, *** p<.01

Lastly, we plot the overall average effect of the program on each EGRA subtask, stratified by student subgroup. Figure 18 shows that **all estimated effects, on average, for the full sample and the subgroups are positive and range between an effect of 2.6 and 9.6 percent, with the largest effect on familiar word reading and the smallest on syllable identification**. However, the figure also shows a high degree of variability in the subgroup level

estimates and we can observe clear heterogeneity in familiar word reading and both versions of the oral reading fluency subtask.

**Figure 18. Summary of regression estimates, by subtask and student subgroup**




# F. DISCUSSION AND FINAL COMMENTS

In this report we presented the main results of the impact evaluation of MMT's Grade 1 reading and writing pilot program on student reading outcomes. This pilot, conducted over the course of four months (April – July 2018) served to evaluate the impact of approximately one third of the entire reading and writing program (10-12 weeks of a 33-week program). Overall, the results presented in this report are encouraging in the context of the short pilot period. Large and statistically significant improvements on letter sound, syllable, and familiar word reading corroborate with the content and skills students were led to practice during the 10-12 weeks of the pilot (see program description for details on sequencing of skills). It is hoped that this impact evaluation will serve to help the Ministry of Education make an informed decision on which reading approach to scale-up. The discussion will summarize some of the key points that emerge from the data presented in this report and will draw upon formative evaluation data collected twice during the life of the pilot to explain certain findings.

***Fidelity of implementation of the pilot program***: It is encouraging to see the large effect size of the Grade 1 MMT program on all subtasks (except reading comprehension) ranging between .20 and .37 SD, for the full sample, with the largest estimated effect being in syllable identification. Further, our analysis shows that teachers who implemented most or all steps of the lesson plan

obtained better student learning outcomes than teachers who did not. This means there was significant and positive correlation between student performance on almost all subtasks and fidelity of implementation (number of steps of the lesson plan activities followed by teachers). Standard lesson plan activity steps reflect the "I do, we do, you do" approach and allocate at least 60% of the whole lesson to the "you do" step or for individual student practice reading letters, syllables, words and/or text or writing while the teacher circulates to monitor and assist. In Madagascar's daily timetable for Grade 1, there are two 20-minute reading lessons per day and one 30-minute writing lesson, this means students should have approximately 24 minutes per day to practice independent reading and 18 minutes to practice their writing. Students were also allowed to take their books home to practice reading and writing. This focus on increased time allocation aligns with the research showing that extended and supportive reading practice is critical to the development of reading proficiency.[9] At the launch of the program, formative evaluation data found that most teachers were not applying the "you do" step. In the refresher training, this was emphasized as an essential step for lesson success and teachers and mentors (school directors) where shown and practiced effective classroom management strategies to applying this step of the lesson. In the second round of formative evaluation, more teachers showed to effectively apply this step in their reading and writing lessons. This means that overall, most students in treatment schools were eventually afforded more time to practice their skills individually with support from the teacher. It is therefore no surprise that treatment school students showed significant improvements in the mechanics of reading (decoding).

Regarding the finding that the program had no significant effects on student's reading comprehension, this is unsurprising for two reasons. First, this result aligns with what students were exposed to in the MMT pilot program. According to monitoring data, most teachers had not yet reached week 24 of the program at the time of the endline EGRA administration. This means students were never exposed to lessons from sequence 3 where skill focus switches from a reading mechanics and decoding to building reading accuracy and fluency through reading leveled text and reading comprehension. Therefore, the only other consistent exposure students would have received to comprehension skill-building is through the daily oral language development lessons (fanazarana hiteny). However, in the first round of formative evaluation data collection, teachers noted having problems knowing how to effectively carry out their oral language development lessons. Even though the refresher training focused more on helping teachers effectively apply their oral language development lessons, it may have still been insufficient to translate into student learning outcomes regarding comprehension.

These reasons may also help to explain the lack of significant effect of the program on listening comprehension. Still, the low listening comprehension scores at both baseline and endline and in both control and treatment schools may point to a larger language of instruction comprehension

[9] Krashen, S. (1988). Do we learn to reading by reading? The relationship between free reading and reading ability. In Linguistics in context: Connecting observation and understanding, ed. D. Tannen. Norwood, NJ: Ablex, pp. 269-298; Cipielewski, J., & Stanovich, K. E. (1992). Predicting growth in reading ability from children's exposure to print. Journal of Experimental Child Psychology 54: 74-89

issue. In Madagascar, the language of instruction is official Malagasy. As part of its support program, MMT conducted separate research on two (of several) dialectical variants of Malagasy (Tsimihety and Sakalava) to better understand the similarities and differences between the dialectical variants and official Malagasy. It was found that Tsimihety and Sakalava had 68% and 70% basic vocabulary in common with Official Malagasy, by the same method which affirms French and Italian to have 89% of basic vocabulary in common. Half of the treatment school students came from a DRENs in a Sakalava-speaking region. Therefore, it could be posited that poor oral language comprehension is a reason for poor student performance in listening comprehension.

***Attaining minimum proficiency (benchmarks):*** Regarding students' likelihood of attaining the benchmarks set by the MOE, it is encouraging to see noticeable shifts in the treatment group toward emergent, proficient, and fluent categories (the latter two categories associated with hitting the benchmark) in almost all subtasks. It was anticipated that this movement from beginner to the other categories would be greatest in the letter sound, syllable identification, and word reading because that is what students spent the most time practicing in the 10-12 weeks of the pilot intervention. This was confirmed by the data which estimated that the MMT intervention increased students' probability of attaining minimal proficiency by 7, 4, and 11 percentage points on the letter sounds, syllable identification, and familiar word reading subtasks respectively when the teacher follows most lesson steps. During the 10-12-week pilot, students also practiced reading decodable texts but given the short timeframe of the program, they were not introduced to the higher levels texts aimed to increase their fluency. Still, the16 percentage point baseline to endline shift from beginner to emergent, and, proficient and fluent categories with an increase of 7% hitting the Ministry-set benchmark in oral reading fluency shows that decodable text reading practice seems to have contributed to improvement in fluency.

Knowing how students shifted across categories relative to the benchmarks is invaluable information for the MOE. Not only do benchmarks help the MOE track student progress relative to particular goals but it also allows them to see if the benchmarks set are realistic and attainable. It is recommended that all future EGRA reports, especially those for which baseline and endline span an entire school year include this comparative analysis. For example, if the Ministry scales up the Grade 1 program to be used over a full school year, it is strongly advised the EGRA is employed again to measure student progress relative to the benchmarks previously set.

***Socioeconomic status***: The results from this report point to the link between socioeconomic status and student's progress in reading that is, the larger effect of the program from students coming from higher SES is troubling, even more so is that the gap widens with the intervention (for letter sounds correct per minute). Though the gains from lower SES treatment school students are higher than those from the control school group, they did not improve as quickly. These findings could be anticipated when looking at the research trends of the achievement gap between low and high SES. Often, students in low SES don't have support for education at home. In the context of MMT, the parent sensitization component attempts to communicate and show parents or family members simple-to-apply activities they can do with their children at home to support reading. However, when talking with parents and family members in focus groups during the formative evaluation data collection, many stated they had no time to help their children at home because they were too busy in the fields or with other income generating activities. From these interviews, we concluded that many students may not have received the support they

needed at home. To face this issue, more robust community-based or extra-curricular activities should be considered. For example, the Japan International Cooperation Agency has shown good success in getting community to provide struggling students with extra-curricular tutoring support. Further research on the specific factors that lead to lower performance in low SES students in Madagascar would also be helpful to inform the parent sensitization strategy, among other points of intervention.

***Use of Collaborating, Learning, and Adapting (CLA):*** Though it cannot be scientifically confirmed, we believe employing the Collaborating, Learning, and Adapting (CLA)[10] approach was a key determinant of the program's overall success and impact on student outcomes despite the short timeframe. Using this approach, MOE enumerators visited all pilot schools twice, three weeks following each training. The intention of these visits were to:

1. Evaluate the implementation of the reading program in the classroom (fidelity of implementation);
2. Identify improvements to make to the materials, training, and community pieces (through questionnaires and focus groups);
3. Provide support and answer questions about the activities piloted at the end of each visit.

The results from these visits were entered, analyzed and immediately shared with the materials development and training teams so they could draw conclusions to make improvements before the next set of materials and trainings were delivered. This data was invaluable to ensuring the program was directly adapted to meet the needs and interests of all beneficiaries. For example, short focus groups with students found what types of stories and activities they liked and didn't like and what words they found difficult to read or understand. As a result, modifications were made to meet the interests of the students and to make the book more attractive to them. For teachers, we asked them questions on what they learned from the training, on what they thought of the teacher's guide, and what they would like more help with. This data was very useful to make appropriate modifications or additions to the teacher's guide and trainings. Other interviews with mentors (role of school directors), ZAP leaders, and parents also served to inform the training. Finally, the data from the fidelity of implementation of the lessons helped us adjust the timing of activities and to identify issues to further address through the training and for mentorship[11]. It is hoped that a CLA will continue to be used as the MOE continues to develop reading and writing programs for Grades 2 and 3.

### Recommendations for the way forward:

At the closing of the MMT program, a lessons learned conference was held so MOE technicians who had worked closely with the program could reflect the activities and process that the

[10] Collaborating, Learning, and Adapting (CLA) approach encompasses ongoing evaluation for the purpose of learning and adapting a program to better fit the needs and realities of the context and beneficiaries. For more information, see USAID CLA toolkit:. https://usaidlearninglab.org/node/14633

[11] Learn more about how CLA was implemented on the MMT program: https://researchforevidence.fhi360.org/how-to-successfully-apply-the-collaborating-learning-and-adapting-cla-approach-in-your-programs

program undertook[12]. Among the many discussions, one involved a reflection on the EGRA adaptation, administration, and analysis and on the use of formative evaluation data /the CLA approach to improve the program as it was being implemented and another on the findings from the research on the dialectical variants of Malagasy. The conclusions and recommendations drawn from these discussions include the following.

For the EGRA:

- Adapting reading assessments to the standards and benchmarks set by the MOE for Grade 1 – 3.
- Continue to employ assessments like the EGRA to assess impact of reading programs as they continue to be developed and piloted.
- Select statistically representative samples including treatment versus control samples so the MOE can make decisions surrounding the results of the EGRA

For use of CLA:

- Continue use of formative evaluation in all programs for reading and other subjects the MOE will develop and use the same approach to pilot the Grade 1 program over a full school year.

Research on the dialectical variants:

- Complete and make good use of the current linguistic analysis. To complete, adopt a similar approach to collecting authentic text corpuses as was done for Sakalava and Tsimhety.
- Create a linguistic map of local dialectical variants of Malagasy focusing especially on pupil and teacher language.
- Identify and analyze relevant research already completed. Involve the Academy and Universities in this process.
- Build effective linkage between available research data, pedagogical/didactic theory and findings from the research led by MMT.
- Expand and deepen research on the use of dialectal variants in Grade 1 classrooms.
- Plan small pilot projects which take a few local dialectical variants of Malagasy as the starting point for the acquisition of reading / writing skills.
- From the linguistic data on local dialectical variants of Malagasy, review the sequence for learning to read, especially the sequence for the teaching of graphemes.
- Identify a clearer language policy for learning official Malagasy (including transfer from local dialectical variants of Malagasy).
- Develop modules for local dialectical variants of Malagasy in teacher training schools

[12] See MMT final report for a comprehensive summary of the program activities, process, and lessons learned.

# ANNEX I. EGRA ADAPTATION

This document describes the process employed during the EGRA adaptation workshop during which stories and comprehension questions were developed. It also includes the psychometric analysis of the pre-test data. Members of the DPE, DTIC and World Bank participated in this workshop (approximately 10 to 12 participants depending on the day). The EGRA adaptation workshop was conducted between 5 March and 8 March 2018.

## I. PROCESS FOR DEVELOPING THE STORIES

As we were working from an existing EGRA tool which was administered to T2 students nationally in 2015, this adaptation workshop focused on story and reading comprehension development for T1.
**Criteria used:** Workshop participants used the text leveling criteria that had been developed during the standards and benchmark setting workshop to guide the development of stories.
Stories for oral reading: For T1, description for text levels 3 and 4 guided the development of the oral reading passages (read by students). This corresponds to the following:

| | |
|---|---|
| Level 3 | • Words have a common structure<br>• On average, words have 4 syllables with longest word at 6 syllables<br>• No more than 3 words with 5 syllabes, 1 word with 6 syllabes<br>• All words are decodable |
| Level 4 | • Words have a common structure<br>• On average, words have 4 syllables with longest word at 6 syllables<br>• No more than 4 words with 5 syllabes, 2 words with 6 syllabes |

The following criteria were also used :

- No compound words
- Proper nouns are capitalized
- The first letter of each sentence is capitalized
- No preference for a particular dialect (words should be understood by all dialects)
- An apostrophe determines the end of a word. For instance "amin' ny" is considered two words.
- One space must be inserted between two words (such that Tangerine recognizes them as two distinct words)

Each story includes 40 to 60 words. For each story, there are 5 comprehension questions: 4 literal questions and 1 inferential question. The inferential question is always the last one.
For this EGRA tool, there are two oral reading subtasks:

1. Oral reading timed at 1 minute without comprehension questions
2. Oral reading timed at 2 minutes with comprehension questions (the story remains in front of the student and the student can consult the story to answer questions)

Stories for listening comprehension : For T1, the text leveling criteria for levels 5 and 6 guided the development of the listening comprehension stories (for which the enumerator reads the story to the student and then asks comprehension questions).

| | |
|---|---|
| Level 5 | • Words have a common structure<br>• On average, words have 4 to 5 (4.5) syllables with the longest word at 7 syllables.<br>• No more than 4 words with 5 syllables, 3 words with 6 syllables and 1 word with 7 syllables. |
| Level 6 | • Some words may have less common syllabic structure but not many<br>• On average, words have 4 to 5 (4.5) syllables with the longest word at 8 syllables<br>• No more than 4 words with 5 syllables, 2 words with six syllables and 1 word with 7 syllables. |

Each story includes 40 to 60 words. For each story, there are 5 comprehension questions: 4 literal questions and 1 inferential question. The inferential question is always the last one.

**Process for developing the stories :**

- During the first two days of the workshop, participants were divided into groups of 3. Each group was tasked with developing 2 stories each along with comprehension questions. For each story, the group then reviewed it using the quality control checklist from the EGRA toolkit to ensure it met the criteria and was appropriate for the Malgache context.
- Of the two stories developed by each group, one was retained as the one for oral reading. Therefore three different stories were developed for oral reading along with 5 comprehension questions each. Then for each of these three stories, we created an "equivalent' form. The EGRA toolkit indicates that the best way to do this is to change the name of the characters, change action verbs or other nouns with others that are approximately of the same level of difficulty, while keeping the same grammatical syntax for all sentences. Using these guidelines, the workshop participants modified the original story to create an "equivalent" story that is very similar to the original.
- The other story developed by each group (not retained for oral reading) was used for listening comprehension. Because these stories were originally developed for oral reading using the criteria for text levels 3 and 4, workshop groups modified the stories slightly to ensure they were of slightly higher difficulty level (levels 5 and 6).
- In total, 9 stories were developed and pre-tested: 6 stories for oral reading (3 pairs of "equivalent" stories) and 3 stories for listening comprehension.
- The oral reading stories are numbered from 1 to 6 (and the "equivalent" pairs are 1-4, 2-5 and 3-6), and the listening comprehension stories are labeled A, B and C.
- The stories are all included in Section II.

**Pre-test process :**

- For the pre-test, workshop participants were the enumerators. They were divided into two groups, group A and B.
- Because of the high number of stories, it was not possible to administer all 9 stories to one student. Instead, each student read 4 stories (oral reading) and listened to 2 stories (listening comprehension.
- Group A and B enumerators therefore administered the stories as described in the table below, such that half of students were tested by Group A and the other half by Group B.

| Groupe A | Groupe B |
|---|---|
| Story 1 (anchor)<br>Story 2<br>Story 3<br>Story 4 (anchor) | Story 1 (anchor)<br>Story 5<br>Story 6<br>Story 4 (anchor) |
| Story A (anchor)<br>Story B | Story A (anchor)<br>Story C |

- In total, we pre-tested the stories with approximately 50 T2 students and 10 T1 students in two schools around Antananarivo (it is usually recommended to pre-test with students of older grades to avoid obtaining floor effects data, which are not informative psychometrically speaking).

**Qualitative results :**

Based on enumerator feedback, the following texts were retained:

- Oral reading timed at 1 min without comprehension : Story 1 and "equivalent" Story 4
- Oral reading timed at 2 min with comprehension : Story 3 and « equivalent » Story 6
  - Back-up for oral reading : Stories 2 and 5
- Listening comprehension : Story B
  - Back-up for listening comprehension : Stories C and A.

**Quantitative results from psychometric analysis :**

Psychometric analysis of the pre-test data was also performed. Overall it confirmed that (1) the pairs of stories for oral reading are indeed equivalent and can be used to create different forms of EGRA; scores can be compared without further need for equating, (3) stories A and B for listening comprehension are of similar level of difficulty, and (3) comprehension questions do not display quality issues in terms of difficulty and discrimination ability. As a result, one EGRA form with stories 1, 3 and B was used for the baseline data collection, and another EGRA form with stories 4, 6 and A for the endline data collection. More detailed results are presented in section III.

## II. STORIES AND COMPREHENSION QUESTIONS

### Stories retained for oral reading timed at 1 minute without comprehension questions

Note : We include the comprehension questions below since they were developed during the workshop, but they are not included in the final EGRA tool.

**STORY 1**

| | |
|---|---|
| Tia fahadiovana i Vero. Mifoha maraina izy. Mangatsiaka anefa amin' ny maraina. Mila misasa izy alohan' ny hianatra. Lasa ihany i Vero misasa amin' ny rano madio sy savony. Manadio ny tarehiny sy ny tanany ary ny tongony i Vero. Zaza madio i Vero. Falifaly i neny. | Vero aime être propre. Elle se lève tôt le matin. Il fait froid. Il faut se laver avant d'aller à l'école. Vero se lave avec de l'eau claire et du savon. Le visage, les mains et les pieds de Vero sont lavées. Vero est un enfant propre. Maman est contente. |

| | |
|---|---|
| Tia fahadiovana i Vero. (4) | Inona no tian' i Vero? [Fahadiovana] |

| | |
|---|---|
| | *Qu'est-ce que Vero aime ? [Etre propre]* |
| Mifoha maraina izy. Mangatsiaka (8) | Manao ahoana ny andro amin'ny maraina? [Mangatsiaka]<br>*Comment est le climat? [Froid]* |
| anefa amin' ny maraina. Mila misasa izy alohan' ny hianatra. (18) | Inona no ataon' i Vero alohan' ny hianatra? [Misasa]<br>*Que fait Vero avant d'aller à l'école ? [se laver]* |
| Lasa ihany i Vero misasa amin' ny rano madio sy savony. Manadio ny tarehiny (32) | Inona avy ireo faritra amin' ny vatana nosasan' i Vero? [tarehy/tanana/tongotra]<br>*Quelles parties de son corps Vero a-t-elle lavées ? [le visage/les mains/les pieds]* |
| sy ny tanany ary ny tongony i Vero. Zaza madio i Vero. Falifaly i neny. (47) | Nahoana i neny no falifaly? [Satria madio i Vero]<br>*Pourquoi la maman est-elle contente ? [parce que Vero s'est lavée]* |

## STORY 4 (EQUIVALENT TO STORY 1)

| | |
|---|---|
| Tia madio i Bema. Mifoha maraina izy. Mangatsiaka anefa amin' ny maraina. Mila midio izy alohan' ny hianatra. Lasa ihany i Bema midio amin'ny rano madio sy savony. Manasa ny tarehiny sy ny tanany ary ny tongony i Bema. Zaza madio i Bema. Ravoravo i Dada. | Bema aime être propre. Il se lève tôt le matin. Il fait froid. Il faut se laver avant d'aller à l'école. Bema se lave avec de l'eau claire et du savon. Le visage, les mains et les pieds de Bema sont lavés. Bema est un enfant propre. Papa est content. |

| | |
|---|---|
| Tia madio i Bema. (4) | Inona no tian' i Bema? [madio]<br>*Qu'est-ce que Bema aime ? [Etre propre]* |
| Mifoha maraina izy. Mangatsiaka (8) | Manahoana ny andro amin' ny maraina ? [Mamanala]<br>*Comment est le climat? [Froid]* |
| anefa amin' ny maraina. Mila midio izy alohan' ny hianatra. (18) | Inona no mila atao alohan' ny hianatra? [Midio]<br>*Que faut-il faire avant d'aller à l'école ? [se laver]* |
| Lasa ihany i Bema midio amin' ny rano madio sy savony. Manasa ny tarehiny (32) | Inona avy ireo faritra amin' ny vatana sasan' i Bema? [Tarehy, Tanana, tongotra] *Quelles parties de son corps Bema a-t-il lavées ? [le visage/les mains/les pieds]* |
| sy ny tanany ary ny tongony i Bema. Zaza madio i Bema. Ravoravo i Dada. (47) | Nahoana i Dada no ravoravo? [Satria madio i Bema]<br>*Pourquoi le papa est-il content ? [parce que Bema s'est lavé/parce qu'il est propre]* |

## Stories retained for oral reading timed at 2 minutes with comprehension questions

## STORY 3

| | |
|---|---|
| Asabotsy ny andro. Mankany anaty saha i Bao sy Zoky. Vokatra be ireo manga tao anaty saha. Sarotra alaina ireo manga masaka. Mitoraka manga i Zoky. Taitra ny fanenitra. Voakaikitra i Bao. Mitomany be izy. Asian' i Zoky ravina ny orony voakaikitra. | C'est samedi. Bao et sa grande sœur vont au champ. Les mangues sont mûres dans les champs. C'est difficile de cueillir les mangues mûres. La grande sœur lance des pierres pour attraper les mangues. Elle touche un nid de guêpe. Bao s'est fait piquer. Elle pleure beaucoup. Sa soeur met une feuille sur son nez. |

| | |
|---|---|
| Asabotsy ny andro. Mankany anaty saha i Bao (8) | Nankaiza i Bao sy i Zoky? [Nankany anaty saha] |
| sy Zoky. Vokatra be ireo manga (14) | Inona ny voankazo vokatra ao anaty saha? [manga] |
| tao anaty saha. Sarotra alaina ireo manga masaka. Mitoraka manga i Zoky. (26) | Inona no ataon' i Zoky ahazoana manga? [mitoraka] |
| Taitra ny fanenitra. Voakaikitra i Bao. Mitomany be izy. (35) | Maninona no mitomany i Bao? [voakaikitra] |

| | |
|---|---|
| Asian' i Zoky ravina ny orony voakaikitra. (42) | Nahoana no asian' i Zoky ravina ny oron' i Bao? [satria marary/mivonto/mba ho sitrana] |

## STORY 6 (EQUIVALENT TO STORY 3)

| | |
|---|---|
| Asabotsy ny andro. Mankany anaty saha i Levelo sy Dada. Vokatra ny voasary ao anaty saha. Sarotra alaina ireo voasary masaka. Mitoraka voasary i Dada. Taitra ny fanenitra. Voakaikitra i Levelo. Mitomany be i Levelo. Nasian' i Dada ravina ny sofiny voakaikitra. | |

| | |
|---|---|
| Asabotsy ny andro. Mankany anaty saha i Levelo (8) | Nankaiza i Levelo sy Dada? [Nankany an-tsaha] |
| sy Dada. Vokatra ny voasary (13) | Inona ny voankazo vokatra ao anaty saha? [Voasary] |
| ao anaty saha. Sarotra alaina ireo voasary masaka. Mitoraka voasary i dada. (25) | Inona no ataon' i Dada ahazoana voasary ? [Mitoraka] |
| Taitra ny fanenitra. Voakaikitra i Levelo. Mitomany be i Levelo. (35) | Maninona no mitomany Levelo ? [voakaikitra] |
| Nasian' i Dada ravina ny sofiny voakaikitra. (42) | Nahoana no asian' i Dada ravina ny sofin'i Levelo? [Marary/mivonto/mba ho sitrana] |

### Back-up for oral reading

## STORY 2

| | |
|---|---|
| Mirava ny sekoly. Maika hody i Miora. Mihazakazaka izy. Miandry azy i Jao anadahiny. Milalao kanety ny fanaony. Matetika mandresy i Miora. Tonga i Miora. Mbola mamafa tokontany i Jao. Manampy azy i miora. Faly ery izy mianadahy miaraka milalao. | La classe est finie. Miora a hâte de rentrer chez elle. Elle court. Son frère Jao l'attend. Ils ont l'habitude de jouer aux billes. Miora gagne souvent. Miora arrive chez elle. Jao balaie encore la cour. Miora l'aide. Enfin, la sœur et le frère sont contents de jouer ensemble |

| | |
|---|---|
| Mirava ny sekoly. Maika hody i Miora. (7) | 1. Iza no maika hody? / i Miora<br>*Qui a hate de rentrer [Miora]* |
| Mihazakazaka izy. Miandry azy i Jao (13) | 2. Iza no miandry an'i Miora? / i Jao/anadahiny<br>*Qui attend Miora? [Jao, son frere]* |
| anadahiny. Milalao kanety ny fanaony. (18) | 3. Inona ny kilalao fanaon'i Miora sy Jao? / kanety<br>*A quoi Miora et Jao ont-ils l'habitude de jouer ? [Aux billes]* |
| Matetika mandresy i Miora. Tonga i Miora. (25) | 4. Iza no mandresy matetika? / i Miora<br>*Qui gagne souvent? [Miora]* |
| Mbola mamafa tokontany i Jao. (30) | 5. Nahoana no tsy tonga dia milalao izy ireo? / satria mbola mamafa tokontany i Jao./manao raharaha<br>*Pourquoi ne peuvent-ils pas jouer tout de suite ? [Jao balaie encore la cour]* |
| Manampy azy i miora. Faly ery izy mianadahy miaraka milalao ihany. (41) | |

## STORY 5 (EQUIVALENT TO STORY 2)

| | |
|---|---|
| Mirava ny sekoly. Maika hody i Zaka. Mihazakazaka izy. Miandry azy i Zefa sakaizany. Milalao baolina ny fanaony. Matetika mandresy i Zaka. Tonga i Zaka. Mbola manondraka anana i Zefa. Manampy azy i Zaka. Faly ery izy roalahy miaraka milalao. | Le grand ménage est fini. Zaka a hâte de rentrer chez lui. Il court. Son camarade Zefa l'attend. Ils ont l'habitude de jouer au ballon. Zaka gagne souvent. Zaka arrive chez lui. Zefa arrose encore les légumes. Enfin les deux camarades sont contents de jouer ensemble. |

| | |
|---|---|
| Mirava ny sekoly. Maika hody i Zaka. (7) | 1. Iza no maika hody? / i Zaka<br>*Qui a hâte de rentrer? / C'est Zaka* |
| Mihazakazaka izy. Miandry azy i Zefa (13) | 2. Iza no miandry an'i Zaka? / i Zefa/sakaizany<br>*Qui attend Zaka? / Zefa* |
| sakaizany. Milalao baolina ny fanaony. (18) | 3. Inona ny kilalao fanaon'i Zaka sy Zefa? / baolina<br>*A quoi Zaka et Zefa ont -ils l'habitude de jouer ? / au ballon* |
| Matetika mandresy i Zaka. Tonga i Zaka. (25) | 4. Iza no mandresy matetika? / i Zaka<br>*Qui gagne souvent? / Zaka* |
| Mbola manondraka anana i Zefa. (30) | 5. Nahoana no tsy tonga dia milalao izy ireo ? / mbola manondraka anana/ manao raharaha<br>*Pourquoi ne peuvent-ils pas jouer tout de suite ?/ Zefa arrose encore les légumes.* |
| Manampy azy i Zaka. Faly ery izy roalahy miaraka milalao. (40) | |

### Stories retained for listening comprehension

## STORY B

| | |
|---|---|
| Maka aina eo ambodimanga Ragidro. Milatsaka eo ambony lohany ny manga. Marary be ny lohany. Miantsoantso Ragidro hoe mianjera ny lanitra. Maheno azy daholo ny biby rehetra. Anisan'izany Rasaka. Mihazakazaka Rasaka milaza amin' i Goaika. Tsy mianjera ny lanitra. Diso hevitra Ragidro. | 1° Aiza i Ragidro no maka aina? /eo ambodimanga<br>2° Inona no milatsaka eo ambony lohany? / manga<br>3° Inona no antsoantson' i Gidro?/ mianjera ny lanitra<br>4° Iza avy no nahare ny antsoantson' i Gidro? / ny biby rehetra<br>5°? Nahoana no diso hevitra I Gidro?/ satria tsy ny lanitra no latsaka fa ny manga |
| Le singe se repose sous un cocotier. Une noix de coco tombe sur sa tête. Sa tête lui fait très mal. Le singe crie : le ciel tombe. Tous les animaux l'entendent. Y compris le chat. Le chat court pour le dire au corbeau. Le ciel ne tombe pas. Le singe délire. | 1° Où se trouve le singe? /sous le cocotier<br>2° Qu'est-ce qui tombe sur sa tête/un noix de coco<br>3° Que crie le singe ?/ le ciel tombe<br>4° Qui entend le cri du singe? / tous les animaux<br>5° Pourquoi le singe délire-t-il ? / Il a très mal à la tête (suite à l'accident) |

## STORY A

| | |
|---|---|
| Ao amin' ny taona faharoa i Toky. Manomboka ny fampianarana mamaky teny i Ramose. Anjaran' i Toky | Ao amin' ny taona fahafiry i Toky ? *(taona faharoa)* |

| | |
|---|---|
| izao no mamaky teny. Variana mitabataba anefa i Toky ka tsy hitany akory Ramose manondro azy. Taitra i Toky. Soa ihany fa mamerina ny teniny i Ramose. Miezaka i Toky. Afaka mamaky teny amin' ny feo mafy ihany i Toky. | Inona no ampianarin-dramose ? *(mamaky teny)*<br>Inona no mahavariana an' i Toky ? *(mitabataba)*<br>Inona no nataon-dramose rehefa taitra i Toky ? (namerina ny teniny)<br>Ahoana no nahatonga an' i Toky afaka mamaky teny ihany ? *(Satria namerina ny teniny Ramose , niezaka izy nanatanteraka ny tenin-dramose)* |

## Back-up for listening comprehension

### HISTOIRE C

| | |
|---|---|
| Tonga ny taombaovao tamin' izay. Manomana akoho atao fanomezana ho an' i dadabe i Toto sy neny. Tampoka teo, lasa nandositra ilay akoho. Fihazakazahana manerana ny tokontany ny an' i Toto manenjika ilay akoho. Nasain' i dadabe vonoina atao sakafo ilay akoho. Miara-mikorana ny mpianakavy. | 1) Oviana no miseho ny tantara ? [taombaovao]<br>2) Ho an'iza ny fanomezana omanin' i neny sy Toto ? [Ho an' i dadabe]<br>3) Iza no nanenjika ny akoho? [Toto]<br>4) Atao inona ilay akoho, hoy dadabe? [Sakafo]<br>5) Nahoana no mikorana ny mpianakavy? [Miaraka misakafo/ nahita an' i dadabe/fety ny andro] |

| | |
|---|---|
| Toky est en classe de T2. Le maître commence la séance de lecture. C'est au tour de Toky de lire maintenant. Toky est distrait, Il bavarde et ne remarque pas que le maître le désigne. Toky sursaute. Heureusement, le maître répète la consigne. Toky fait des efforts. Finalement, Toky peut lire à haute voix. | Toky est dans quelle classe ? *(Classe T2)*<br>Quelle séance d'enseignement le maître commence-t-elle ? *(Lecture)*<br>Pourquoi Toky est-il distrait ? *(Il bavarde)*<br>Qu'est-ce que le maitre a fait en voyant Toky sursauté ? ()<br>Pourquoi Toky arrive à lire à la fin ?*(Le maitre répète la consigne, il fait un effort))* |

## III. PSYCHOMETRIC ANALYSIS OF PRE-TEST DATA

As explained earlier, the stories were pre-tested in two groups, Group A and Group B. Group A included 48 students and Group B 35 students, for a total of 83 students. Below we present the psychometric analysis investigating the equivalence of the stories and comprehension questions.

### 1. Oral reading and comprehension

***Descriptive results***

**Means and proportions of zero scores in oral reading fluency by story and by group**

| | Group A | | Group B | |
|---|---|---|---|---|
| | **Mean (SD)** | **% zero** | **Mean (SD)** | **% zero** |
| **Story 1** | 25.4 (14.3) | 4.2% | 19.5 (13.6) | 2.9% |
| **Story 4** | 27.4 (16.5) | 0% | 19.9 (13.2) | 2.9% |
| **Story 2** | 23.7 (17.6) | 2.1% | - | - |
| **Story 5** | - | - | 17.2 (12.3) | 5.7% |
| **Story 3** | 20.4 (11.3) | 2.1% | - | - |
| **Story 6** | | | 13.8 (9.7) | 2.9% |

Stories 1 and 4 seem to be of similar difficulty while stories 2, 3, 5 and 6 are more difficult. We observe that Group A students are more competent than Group B students since oral reading fluency mean scores on stories 1 and 4 (common to both groups) are higher for Group A.

**Correlation between oral reading fluency scores for Group A**

| | **Story 1** | **Story 4** | **Story 2** | **Story 3** |
|---|---|---|---|---|
| **Story 1** | 1.00 | | | |
| **Story 4** | 0.93 | 1.00 | | |
| **Story 2** | 0.95 | 0.93 | 1.00 | |
| **Story 3** | 0.94 | 0.94 | 0.94 | 1.00 |

**Correlation between oral reading fluency scores for Group B**

| | **Story 1** | **Story 4** | **Story 5** | **Story 6** |
|---|---|---|---|---|
| **Story 1** | 1.00 | | | |
| **Story 4** | 0.96 | 1.00 | | |
| **Story 5** | 0.95 | 0.94 | 1.00 | |
| **Story 6** | 0.94 | 0.95 | 0.96 | 1.00 |

Correlations between stories are high for all stories and both groups.

**Means and proportions of zero scores in reading comprehension by story and by group**

| | Group A | | Group B | |
|---|---|---|---|---|
| | **Mean (SD)** | **% zero** | **Mean (SD)** | **% zero** |
| **Story 1** | 2.8 (1.7) | 18.8% | 2.2 (1.7) | 28.6% |
| **Story 4** | 2.8 (1.9) | 20.8% | 2.3 (1.7) | 22.9% |
| **Story 2** | 2.9 (1.7) | 18.8% | - | - |
| **Story 5** | - | - | 2.0 (1.8) | 32.3% |
| **Story 3** | 2.0 (1.7) | 31.3% | - | - |
| **Story 6** | - | - | 1.6 (1.6) | 37.1% |

The level of difficulty for Story 1 and Story 4 reading comprehension questions are similar while they are more difficult for Stories 3, 5 and 6. Comprehension questions for Story 2 are as easy as Story 1 and 4 for Group A.

**Correlations between reading comprehension scores for Group A**

| | Story 1 | Story 4 | Story 2 | Story 3 |
|---|---|---|---|---|
| **Story 1** | 1.00 | | | |
| **Story 4** | 0.64 | 1.00 | | |
| **Story 2** | 0.70 | 0.67 | 1.00 | |
| **Story 3** | 0.62 | 0.63 | 0.64 | 1.00 |

**Correlations between reading comprehension scores for Group B**

| | Story 1 | Story 4 | Story 5 | Story 6 |
|---|---|---|---|---|
| **Story 1** | 1.00 | | | |
| **Story 4** | 0.69 | 1.00 | | |
| **Story 5** | 0.76 | 0.56 | 1.00 | |
| **Story 6** | 0.65 | 0.72 | 0.62 | 1.00 |

While lower than for oral reading fluency scores, correlations between reading comprehension scores remain high. It is important to consider that comprehension scores range from 0 to 5 while oral reading fluency scores range from 0 to 100 or higher.

***Combining results from both groups***

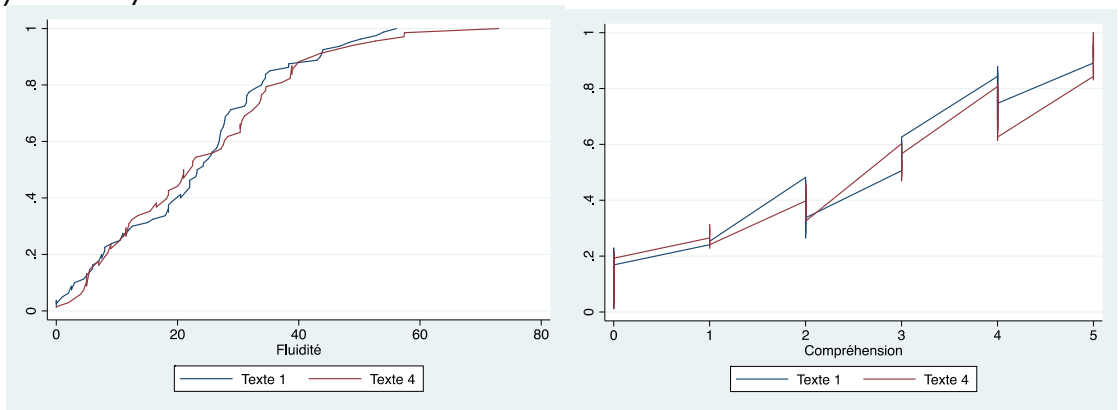
In this section we present graphs where the y-axis represents the percentile of the respondent while the x-axis represents the oral reading fluency score (left graph) or comprehension score (right graph). The blue and red lines present the relationship between the percentile and the oral reading fluency/comprehension score for each story. The more the blue and red lines overlap, the more similar the level of difficult of the two stories.

In the graphs below, we see that:

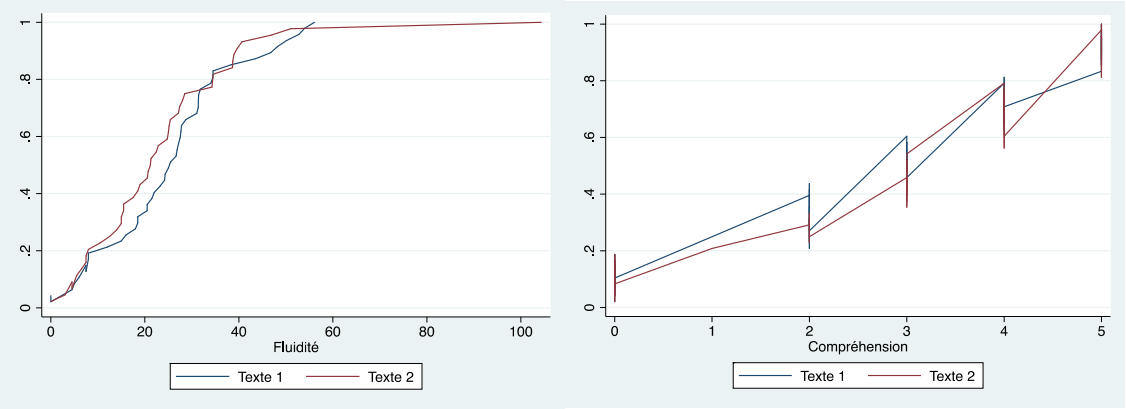
(1) In terms of oral reading fluency, stories 1, 4, 2 and 5 are very similar. Stories 3 and 6 are similar to each other but more difficult than the other four stories.

(2) In terms of comprehension questions, the trend is similar except that the questions for story 5 are more difficult than for story 2.

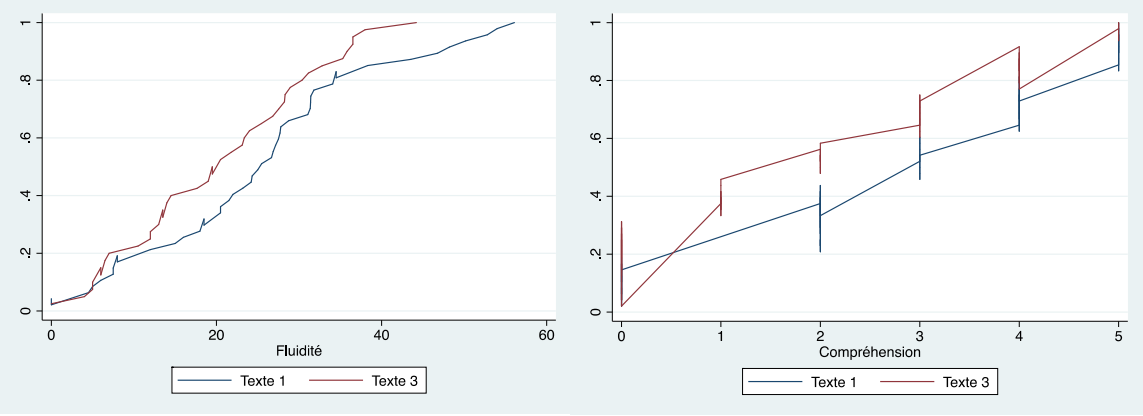
*Story1 and Story 4*



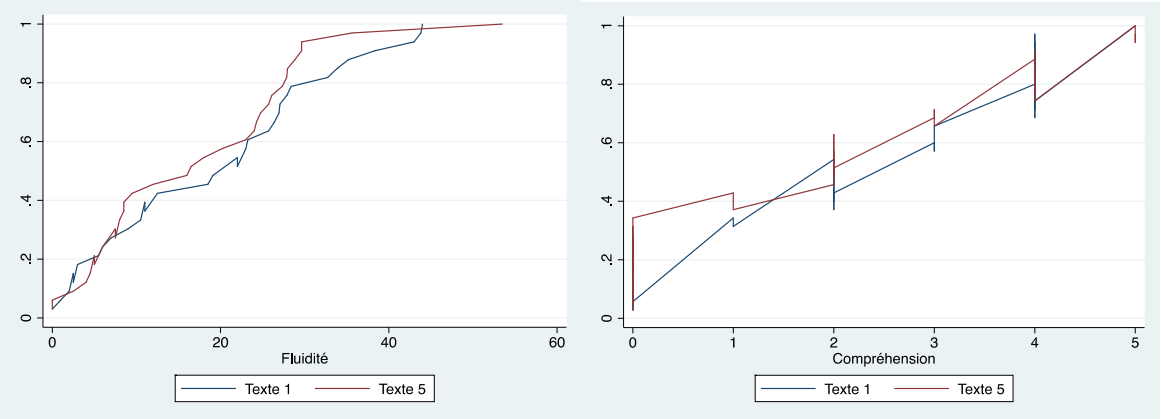
*Story 1 and Story 2*



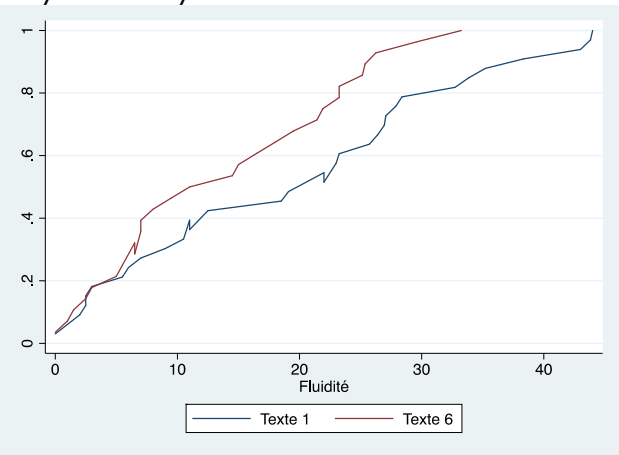

*Story 1 and Story 3*

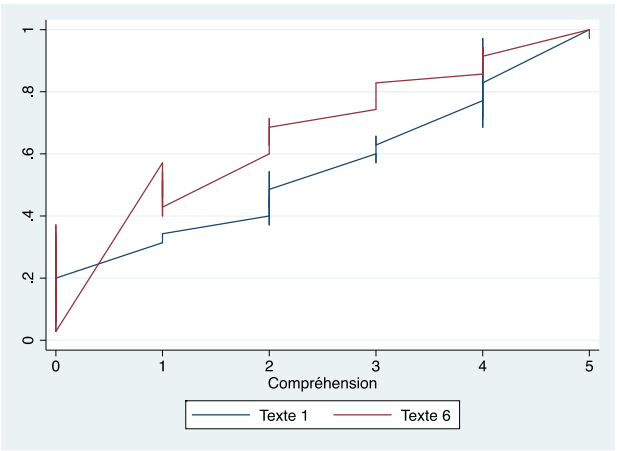



*Story 1 and Story 5*

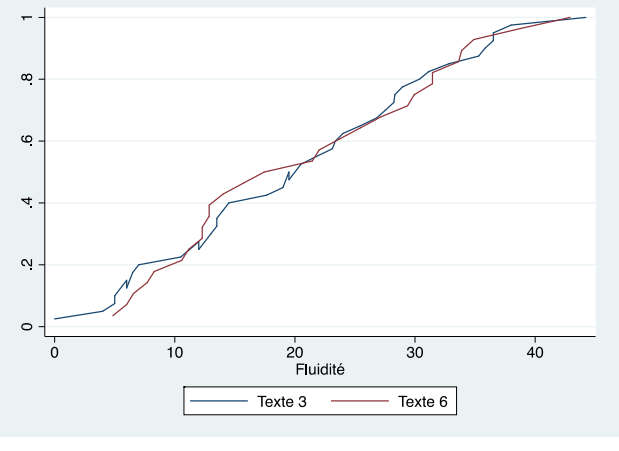


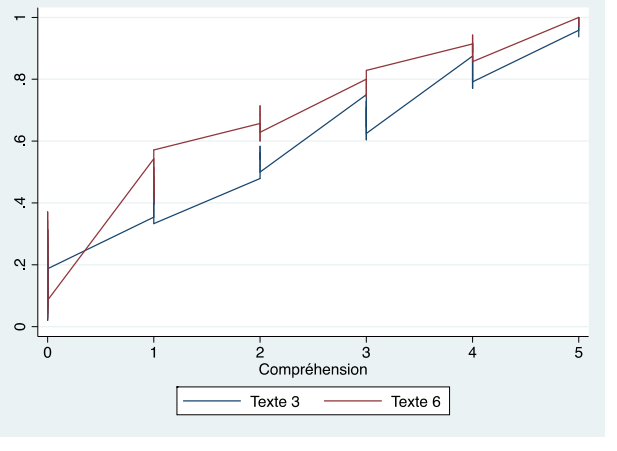
*Story 1 and Story 6*



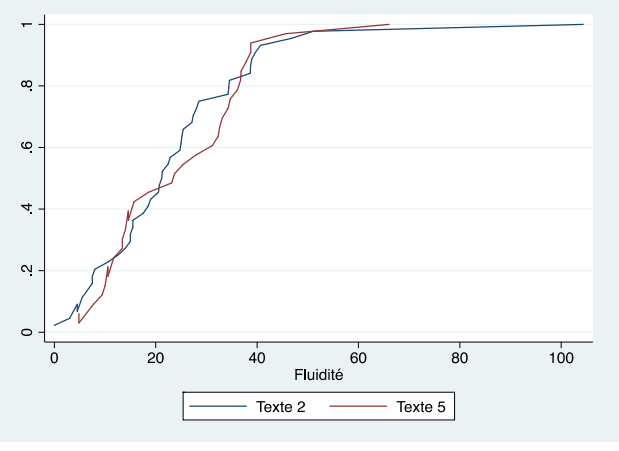



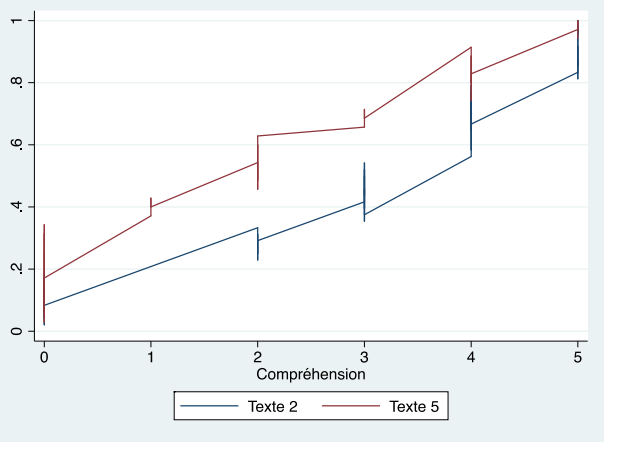

*Story 3 and Story 6*




*Story 2 and Story 5*



## 2. Listening comprehension

### *Descriptive results*

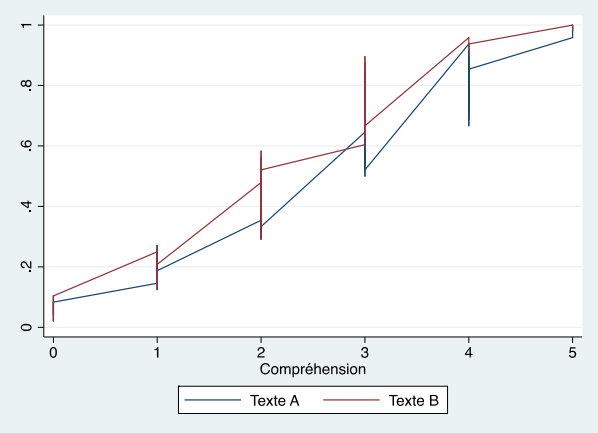
**Means and proportions of zero scores in listening comprehension by story and by group**

| | Group A | | Group B | |
|---|---|---|---|---|
| | **Mean (SD)** | **% zero** | **Mean (SD)** | **% zero** |
| **Story A** | 2.56 (1.49) | 10.4% | 2.14 (1.61) | 22.9% |
| **Story B** | 2.19 (1.23) | 10.4% | - | - |
| **Story C** | - | - | 2.46 (1.38) | 11.4% |

There does not seem to be important differences in terms of difficulty level between the listening comprehension questions of each story. In Group A, story B has a lower mean than Story A, while in Group B, Story C has a higher mean than story A.

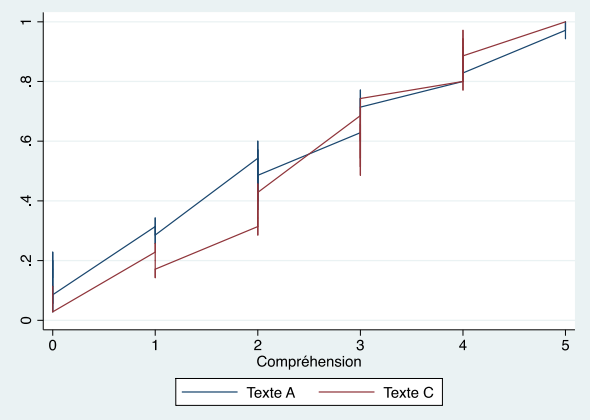
### *Combining results from both groups*

*Story A and Story B*



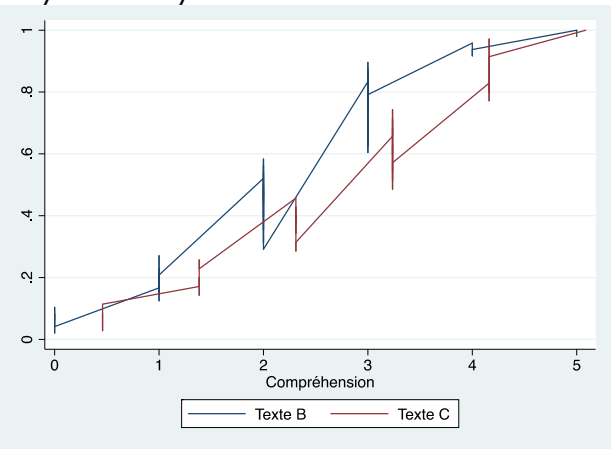

We observe few differences in difficulty level between story A and B.

*Story A and Story C*




Similarly, Stories A and C seem to be of similar difficulty level. De même, le niveau de difficulté de l'histoire A et C semble aussi similaire.

*Story B and Story C*




Comparing Stories B and C, we observe a somewhat bigger difference than with Story A. This difference is more noticeable with students that obtained a score of 3 or 4.

***Score on each question item***

| | Histoire A | | Histoire B | Histoire C |
|---|---|---|---|---|
| | **Groupe A** | **Groupe B** | | |
| **Question 1** | 68.1% | 51.4% | 34.0% | 40% |
| **Question 2** | 68.1% | 62.9% | 42.6% | 62.9% |
| **Question 3** | 59.6% | 51.4% | 48.9% | 57.1% |
| **Question 4** | 36.2% | 17.1% | 68.1% | 60% |
| **Question 5** | 29.8% | 31.4% | 29.8% | 25.7% |

## 3. Analysis of reliability of comprehension questions

A reliability analysis of comprehension questions allows us to verify the quality of the comprehension measure from the questions associated with each story. However, tt is important to remember that the small sample may lead to somewhat unstable results and that findings must be interpreted as general trends.

For each set of comprehension questions, certain indicators of quality are calculated, namely difficulty and discrimination. The difficulty indicator represents the proportion of respondents that answered the question correctly. The higher the value, the easier the question. The discrimination indicator represents the capacity of a question to distinguish between students with the lowest score from those with the highest score on the set of five questions. The higher the value, the better its discrimination capacity.
Finally, Cronbach's alpha coefficients were also calculated. This indicator represents the quality of the score obtained out of the five questions as a measure of students' reading or listening comprehension ability.

***Reading comprehension***

Overall for the reading comprehension questions, Cronbach's alpha varies between 0.654 (Story 1) and 0.798 (Story 2). These values are acceptable given the small sample size and the small number of questions comprising the comprehension score. In summary, there are no major problems in terms of the quality of the comprehension questions for all 6 stories.

**Story 1 : Difficulty and discrimination indicators for comprehension questions**

| Item | Difficulté | Discrimination |
|---|---|---|
| **Question 1** | 0.41 | 0.47 |
| **Question 2** | 0.42 | 0.13 |
| **Question 3** | 0.69 | 0.51 |
| **Question 4** | 0.64 | 0.50 |
| **Question 5** | 0.56 | 0.62 |
| | alpha de Cronbach | 0.654 |

**Story 2 : Difficulty and discrimination indicators for comprehension questions**

| Item | Difficulté | Discrimination |
|---|---|---|
| **Question 1** | 0.79 | 0.64 |
| **Question 2** | 0.60 | 0.60 |
| **Question 3** | 0.66 | 0.61 |
| **Question 4** | 0.70 | 0.66 |
| **Question 5** | 0.25 | 0.40 |
| | alpha de Cronbach | 0.798 |

**Story 3 : Difficulty and discrimination indicators for comprehension questions**

| Item | Difficulté | Discrimination |
|---|---|---|
| **Question 1** | 0.38 | 0.63 |
| **Question 2** | 0.58 | 0.73 |
| **Question 3** | 0.39 | 0.47 |
| **Question 4** | 0.31 | 0.53 |
| **Question 5** | 0.31 | 0.36 |
| | alpha de Cronbach | 0.769 |

**Story 4 : Difficulty and discrimination indicators for comprehension questions**

| Item | Difficulté | Discrimination |
|---|---|---|
| **Question 1** | 0.56 | 0.64 |
| **Question 2** | 0.27 | 0.43 |
| **Question 3** | 0.61 | 0.55 |
| **Question 4** | 0.61 | 0.67 |
| **Question 5** | 0.57 | 0.61 |
| | alpha de Cronbach | 0.797 |

**Story 5 : Difficulty and discrimination indicators for comprehension questions**

| Item | Difficulté | Discrimination |
|---|---|---|
| **Question 1** | 0.67 | 0.58 |
| **Question 2** | 0.48 | 0.56 |
| **Question 3** | 0.33 | 0.47 |
| **Question 4** | 0.77 | 0.48 |
| **Question 5** | 0.29 | 0.45 |
| | alpha de Cronbach | 0.692 |

**Story 6 : Difficulty and discrimination indicators for comprehension questions**

| Item | Difficulté | Discrimination |
|---|---|---|
| **Question 1** | 0.27 | 0.08 |
| **Question 2** | 0.61 | 0.26 |
| **Question 3** | 0.59 | 0.68 |
| **Question 4** | 0.33 | 0.55 |
| **Question 5** | 0.50 | 0.52 |

| | | |
|---|---|---|
| | alpha de Cronbach | 0.680 |

***Listening comprehension***

For listening comprehension, Cronback's alpha coefficients are lower than those observed for reading comprehension. Story B has a relatively low Cronbach's alpha, although we caution against any conclusions given the small sample sizes.

**Story A : Difficulty and discrimination indicators for comprehension questions**

| Item | Difficulté | Discrimination |
|---|---|---|
| **Question 1** | 0.61 | 0.46 |
| **Question 2** | 0.66 | 0.43 |
| **Question 3** | 0.56 | 0.41 |
| **Question 4** | 0.28 | 0.38 |
| **Question 5** | 0.30 | 0.31 |
| | alpha de Cronbach | 0.644 |

**Story B : Difficulty and discrimination indicators for comprehension questions**

| Item | Difficulté | Discrimination |
|---|---|---|
| **Question 1** | 0.34 | 0.13 |
| **Question 2** | 0.43 | 0.21 |
| **Question 3** | 0.49 | 0.21 |
| **Question 4** | 0.68 | 0.19 |
| **Question 5** | 0.30 | 0.08 |
| | alpha de Cronbach | 0.237 |

**Story C : Difficulty and discrimination indicators for comprehension questions**

| Item | Difficulté | Discrimination |
|---|---|---|
| **Question 1** | 0.40 | 0.26 |
| **Question 2** | 0.63 | 0.30 |
| **Question 3** | 0.57 | 0.33 |
| **Question 4** | 0.60 | 0.15 |
| **Question 5** | 0.26 | 0.23 |
| | alpha de Cronbach | 0.473 |

# ANNEX II. INSTRUMENTS

**Mada Egra April Baseline**

**Last Updated: 1522757575000**

Enumerator Name

---

*Sekoly*

DREN

CISCO

ZAP

Ecole

EcoleID

---

*id*

id

---

*Consentement - Faneken'ny mpianatra ho adinina*

**Hazavaiko aminao ny antony ahatongavako eto. Miara-miasa amin'ny mpampianatrao aho, ary te hahafantatra ny fomba fianaran'ny ankizy mamaky teny. Voafidy ho amin'izany ny sekolinao ka voafantina ho isan'ireo mandray anjara amin'izany ianao**.

**Alohan'ny hanapahanao hevitra dia tiako ho azonao tsara ny hataontsika eto.**

**Hamaky teny isika ka tahaka ny milalao no hanaovantsika izany.  Aorian'izay dia hametraka fanontaniana vitsivitsy aminao aho. Hampiasa ity tablette ity aho handraisana ny valinteninao.**

**Raha mbola tianao dia hamerina ity vakiteny ity ianao  amin'ny faran'ny  taom-pianarana. Afaka miova hevitra hatrany anefa ianao.**

**Manana fanontaniana ve ianao ?**

**Tsy fanadinana akory ity an! Tsy lazaiko an'iza na iza ny valinteninao na ny mpampianatrao na ny ray aman-dreninao. Ho raisiko ny anaranao mba hahamora ny fandraisanao anjara amin'ny faran'ny taom-pianarana fa tsy hisy hahafantatra izany.**

**Zava-dehibe tokoa ny fandraisanao anjara nefa na izany aza afaka mandà ianao raha toa ka tsy sitrakao izany.Afaka tsy mamaly fanontaniana ianao.Azo hajanona rehefa tsy te hanohy intsony ianao.**

**Manana fanontaniana ve ianao?**

Manaiky sy vonona ny handray anjara ve ianao ? Afaka manomboka ve isika ? ☐

---

*Lettres - Famantarana litera*

**Inty misy taratasy ahitana litera sy litera mifampibaby. Omeo ahy ny feo hanononana ireo litera ireo fa tsy ny anarany.**

**Ohatra, ity litera ity** [Atoro azy ny ''a''] dia **vakiana hoe "a" toy ny amin'ny teny hoe ''dada''.**

**Andao handramantsika. Vakio io litera io.** [Atoro azy ny ''s'']

[Raha marina ny valiny, dia tenenina hoe Tsara be]: **Vakiana hoe "s" toy ny amin'ny teny hoe ''sa'' io litera io.**

[Raha diso ny valiny,dia lazaina hoe]: **Vakiana hoe "s" toy ny amin'ny teny hoe ''sa'' io litera io.**

**Hijery ohatra hafa indray isika. Vakio ity litera mifampibaby ity.** [atoro azy ny ''tr'']

[Raha marina ny valiny, dia tenenina hoe]: **Tsara be, vakiana hoe "tr" toy ny amin'ny teny hoe ''trano'' io litera mifampibaby io.**

[Raha diso ny valiny, lazaina hoe]: **Vakiana hoe “tr” toy ny amin’ny teny hoe ‘’trano’’ io litera mifampibaby io.**

**Rehefa miteny aho hoe "Atombohy", dia manomboka mamaky eto ianao** [tondroy amin’ny fanondro ny litera voalohany], **dia kendreo tsara ny famakianao azy isan’andalana ka avy any ankavia miankavanana.**

**Vakio tsara sy haingana ary ataovy mafy.**

**Tondroy amin’ny fanondronao ireo litera rehefa mamaky ianao. Vakio tsara sy haingana. Raha misy tsy hainao, dia dingano dia tohizo ny vakiteninao.**

**Tondroy amin’ny fanondronao ny litera voalohany.**

**Vonona ianao?**

**Atombohy àry.**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| E | p | o | ts | dr | ntr | f | v | d | R |
| ng | nj | S | y | z | H | nts | a | mb | N |
| m | K | ndr | i | j | b | nd | g | nt | T |
| nk | L | tr | mp | t | e | Y | R | k | I |
| TR | A | s | F | h | M | n | O | V | M |
| h | E | A | i | n | O | s | r | y | v |
| nd | R | O | z | a | i | K | n | f | ts |
| i | O | t | I | tr | g | a | o | nt | A |
| p | m | L | Y | k | N | b | e | H | j |
| mp | nj | N | d | mb | o | ng | I | a | N |

Time Remaining

Autostop?

---

*Syllabes - Famantarana vaninteny*

**Inty misy taratasy ahitana vaninteny. Omeo ahy ny feo hanononana ireo vaninteny ireo .**

**Ohatra, ity vaninteny ity** [Atoro azy ny ‘’hi’’] **vakiana hoe “hi”.**

**Andao handramantsika. Vakio ity vaninteny ity.** [Atoro azy ny ‘’la’’]

[Raha marina ny valiny, dia tenenina hoe Tsara be]: **Vakiana hoe “la”.**

[Raha diso ny valiny,dia lazaina hoe]: **Vakiana hoe “la”.**

**Hijery ohatra hafa indray isika. Vakio ity vaninteny ity.** [atoro azy ny ‘’se’’]

[Raha marina ny valiny, dia tenenina hoe]: **Tsara be, vakiana hoe “se”.**

[Raha diso ny valiny,dia lazaina hoe]: **Vakiana hoe “se”.**

**Rehefa miteny aho hoe “Atombohy”,dia manomboka mamaky eto ianao** [tondroy amin’ny fanondro ny vaninteny voalohany], **dia kendreo tsara ny famakianao azy isan’andalana ka avy any ankavia no miankavanana. Vakio tsara sy haingana ary ataovy mafy.**

**Tondroy amin’ny fanondronao ireo vaninteny rehefa mamaky ianao. Vakio tsara sy haingana. Raha misy tsy hainao, dia dingano dia tohizo ny vakiteninao.**

**Tondroy amin’ny fanondronao ny vaninteny voalohany.**

**Vonona ianao?**

**Atombohy àry.**

| | | | | |
|---|---|---|---|---|
| dy | mi | do | mpo | tro |
| ri | pa | si | fi | so |
| ndro | ki | da | hy | li |
| ma | ky | tso | bo | za |
| zi | ly | vo | ndra | zy |
| ke | nka | mpa | di | ngo |
| po | ni | vi | tsa | ndri |
| ne | nga | nta | me | ti |
| tra | le | nda | mpi | nja |
| mbo | tsi | ntsi | mo | ha |

Time Remaining

Autostop?

---

*Mots familiers - Famakiana teny mahazatra*

**Inty misy taratasy ahitana teny maromaro. Ezaho vakiana araka izay vitanao ireo teny ireo.**

**Tsy tononina fotsiny ny litera fa vakiana mihitsy ny teny.**

**Ohatra : vakiana hoe ‘’nono’’ fa tsy hoe ‘’n’’, ‘’o’’, ‘’n’’, ‘’o’’**

**Andao handramantsika. Vakio ity teny ity** [Atoro azy ny ‘’faly’’]

[Raha marina ny valiny, dia tenenina hoe]: **Tsara be, vakiana hoe “faly”’io teny io.**

[Raha diso ny valiny, dia lazaina hoe]: **Vakiana hoe “faly”’io teny io fa tsy izany.**

**Hijery ohatra hafa indray isika. Vakio ity teny ity** [atoro azy ny teny hoe ‘’ireo’’].

[Raha marina ny valiny, dia tenenina hoe:] **Tsara be, vakiana hoe “ireo’io teny io.**

[Raha diso ny valiny, dia lazaina hoe]: **Vakiana hoe “ireo” io teny io fa tsy izany.**

**Ao tsara ve (na mety!) ? Afaka manohy isika?**

**Raha vao miteny aho hoe:" atombohy", dia tondroy amin'ny fanondro ny teny tsirairay rehefa mamaky azy ianao.**

**Kendreo tsara ny famakiana azy isan'andalana ka avy any ankavia no miankavanana. Tsy hiteny na inona na inona aho amin'ity indray mitoraka ity. Ianao irery no mamaky misesy ireo teny ireo.**

**Azonao tsara ve? Apetraho eo amin'ny teny voalohany ny fanondronao.**

**Vonona ve ianao? Vakio tsara sy mafy.**

**Atombohy àry.**

| | | | | |
|---|---|---|---|---|
| mody | bota | zaza | loha | mena |
| salama | vary | maso | fohy | hena |
| saka | lay | may | mofo | jiro |
| roa | kibo | vato | kapa | reraka |
| neny | valo | lakana | bika | tanana |
| boky | seza | lava | sary | lela |
| nify | biby | hoho | sira | folo |
| mazoto | mamy | mavo | marary | izy |
| harona | tady | vy | volo | orana |
| mafy | kisoa | sofina | lova | doda |

Time Remaining

Autostop?

---

*Histoire 1 min*

**Izao an! Inty misy tantara. Vakio tsara, mafy sy haingana araka izay vitanao. Rehefa miteny aho hoe : 'Atombohy  ny vakiteny', dia manomboka ianao.**

**Raha misy teny tsy hainao vakiana dia dingano ary tohizo ny manaraka.**

**Apetraho eo amin'ny teny voalohany ny fanondronao.**

**Vonona ve ianao?**

**Atombohy àry**.

| | | | | | | |
|---|---|---|---|---|---|---|
| Tia | fahadiovana | i | Vero. | Mifoha | maraina | izy. |
| Mangatsiaka | anefa | amin' | ny | maraina. | Mila | misasa |
| izy | alohan' | ny | hianatra. | Lasa | ihany | i |
| Vero | misasa | amin' | ny | rano | madio | sy |
| savony. | Manadio | ny | tarehiny | sy | ny | tanany |
| ary | ny | tongony | i | Vero. | Zaza | madio |
| i | Vero. | Falifaly | i | neny. | | |

Time Remaining

Autostop?

---

*Histoire 2 min*

**Izao an! Inty misy tantara.**

**Vakio tsara, mafy sy haingana araka izay vitanao. Hametraka fanontaniana vitsivitsy aminao aho rehefa avy eo.**

**Rehefa miteny aho hoe 'Atombohy' ny vakiteny dia manomboka ianao.**

**Raha misy teny tsy hainao vakina dia dingano ary tohizo ny manaraka.**

**Apetraho eo amin'ny teny voalohany ny fanondronao.**

**Vonona ve ianao?**

**Atombohy àry**.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Asabotsy | ny | andro. | Mankany | anaty | saha | i | Bao |
| sy | i | Zoky. | Vokatra | be | ireo | manga | tao |
| anaty | saha. | Sarotra | alaina | ireo | manga | masaka. | Mitoraka |
| manga | i | Zoky. | Taitra | ny | fanenitra. | Voakaikitra | i |
| Bao. | Mitomany | be | izy. | Asian’ | i | Zoky | ravina |
| ny | orony | voakaikitra. | | | | | |

Time Remaining

Autostop?

---

*Comprehension de l'histoire*

**Izao an!**

**Hametraka fanontaniana vitsivitsy aminao aho.**

**Valio araka izay ahaizanao azy ny fanontaniana.**

**Azonao jerena ao anatin’ny tantara ny valiny.**

Nankaiza i Bao sy i Zoky? [Nankany anaty saha]

Marina Diso Tsy namaly

Inona ny voankazo vokatra ao anaty saha? [manga]

Marina Diso Tsy namaly

Inona no ataon’ i Zoky ahazoana manga? [mitoraka]

Marina Diso Tsy namaly

Maninona no mitomany i Bao? [voakaikitra]

Marina Diso Tsy namaly

Nahoana no asian’ i Zoky ravina ny oron’ i Bao? [satria marary/mivonto/mba ho sitrana]

Marina Diso Tsy namaly

---

*Comprehension Orale*

Izao dia izaho indray no hamaky tantara aminao. Hovakiko aminao indray mandeha ihany ny tantara koa mihainoa tsara. Avy eo dia valio araka izay ahaizanao azy ny fanontaniana.

**Maka aina eo am-bodimanga Ragidro. Milatsaka eo ambony lohany ny manga.**

**Marary be ny lohany. Miantsoantso Ragidro hoe mianjera ny lanitra.**

**Maheno azy daholo ny biby rehetra. Anisan’izany Rasaka.**

**Mihazakazaka Rasaka milaza amin’ i Goaika.**

## Tsy mianjera ny lanitra.  Diso hevitra Ragidro.

Ireto àry ny fanontaniana.

Aiza i Ragidro no maka aina? (eo ambodimanga)

☐ Marina ☐ Diso ☐ Tsy namaly

Inona no milatsaka eo ambony lohany? [manga]

☐ Marina ☐ Diso ☐ Tsy namaly

Inona no antsoantson‘ i Gidro? (mianjera ny lanitra)

☐ Marina ☐ Diso ☐ Tsy namaly

Iza avy no nahare ny antsoantson’ i Gidro? [ny biby rehetra na Rasaka na goaika]

☐ Marina ☐ Diso ☐ Tsy namaly

Nahoana no diso hevitra i Gidro? [satria tsy ny lanitra no latsaka fa ny manga]

☐ Marina ☐ Diso ☐ Tsy namaly

---

*Questionnaire élève - Mombamomba ny mpianatra*

0. Iza ny anaranao ?

☐

1. Lahy sa Vavy ?

☐ Lahy ☐ Vavy

2.a. Kilasy fahafiry ianao izao?

☐ T1 ☐ Tsy namaly

2.b. Kilasy fizarana misy anao

☐ Section A ☐ section B ☐ Section C ☐ Section D ☐ Section E ☐ Section F

3. Firy taona ianao ?

☐

4.a. Kilasy faha-firy ianao tamin’ny taona lasa?

☐ Tsy nianatra ☐ Préscolaire/Maternelle ☐ T1 ☐ Tsy namaly

4.b. Efa nanao maternelle ve ianao talohan’ ny nidirana T1? ([Raha “T1” ny valiny dia apetraka ny fanontaniana b] )

☐ Eny ☐ Tsia ☐ Tsy namaly

5.a. Manana boky famakian-teny ve ianao ao an-dakilasy?

☐ Eny ☐ Tsia

5.b. Azonao entina mody ve?

☐ Eny ☐ Tsia ☐ Tsy namaly

6. Ankoatra ny boky fianarana, misy boky hafa na gazety azonao vakiana ve any an-trano?

☐ Eny ☐ Tsia ☐ TSy namaly

7. Misy olona mahay mamaky teny ve ao an-tranonareo?

☐ Eny ☐ Tsia ☐ Tsy namaly

8. Manome asa entimody ve ny mpampianatra anao?

☐ Eny ☐ Tsia ☐ Tsy namaly

9. Misy olon-kafa ve manampy anao amin'ny fanaovana ny asa entimody?

☐ Eny ☐ Tsia ☐ Tsy namaly

10. Impiry misakafo ianao isan’ andro?

☐ Indray ☐ Indroa ☐ Intelo

11. Misy an’ireto ve any an-tranonareo?

☐ Radio ☐ Tele ☐ Telefaonina ☐ Bisikileta ☐ Moto ☐ Fiarakodia ☐ Rano amin’ ny paompy ao an-trano

☐ Jiro avy amin’ ny herinaratra ☐ Kabone ☐ Tsy manana ☐ Tsy namaly

12. Sekoly miankina sa sekoly tsy miankina

☐ Sekoly miankina ☐ Sekoly tsy miankina

13. Sekoly ambonivohitra sa ambanivohitra?

☐ Ambonivohitra ☐ Ambanivohitra

14. Fandaharana maraina sa ny tolakandro ?

☐ Maraina ☐ Tolakandro

*Coordonnées GPS*

**Mada Egra Juil Endline**

**Last Updated: 1530022884000**

Enumerator Name

---

*Sekoly*

DREN

CISCO

ZAP

Ecole

EcoleID

---

*id*

id

---

*Consentement - Faneken'ny mpianatra ho adinina*

Hazavaiko aminao ny antony ahatongavako eto. Miara-miasa amin'ny mpampianatrao aho, ary mbola te hahafantatra ny fomba fianaran'ny ankizy mamaky teny. Voafidy ho amin'izany ny sekolinao ka voafantina ho isan'ireo nandray anjara ianao. Hamerina ny vakiteny indray ianao izao nefa alohan'izay dia haveriko hazavaina tsara ny hataontsika eto.

Hamaky teny isika ka tahaka ny milalao no hanaovantsika izany.  Aorian'izay dia hametraka fanontaniana vitsivitsy aminao aho Hampiasa ity tablette ity aho handraisana ny valinteninao.

Tsy fanadinana akory ity an tsy lazaiko an'iza na iza ny valinteninao na amin'ny mpampianatrao na amin'ny ray aman-dreninao. Ho raisiko ny anaranao mba hahamora ny fandraisana ny valinteninao.

Manana fanontaniana ve ianao ?

Zava-dehibe tokoa ny handraisanao anjara indray nefa na izany aza afaka mandà ianao raha toa ka tsy sitrakao izany.Afaka tsy mamaly fanontaniana ianao.Azo hajanona rehefa tsy te hanohy intsony ianao.

Manaiky sy vonona ny handray anjara ve ianao ? Afaka manomboka ve isika ? ☐

---

*Lettres - Famantarana litera*

**Inty misy taratasy ahitana litera sy litera mifampibaby. Omeo ahy ny feo hanononana ireo litera ireo fa tsy ny anarany.**

**Ohatra, ity litera ity** [Atoro azy ny ‘’a’’] dia **vakiana hoe “a” toy ny amin’ny teny hoe ‘’dada’’.**

**Andao handramantsika. Vakio io litera io.** [Atoro azy ny ‘’s’’]

[Raha marina ny valiny, dia tenenina hoe Tsara be]: **Vakiana hoe “s” toy ny amin’ny teny hoe ‘’sa’’ io litera io.**

[Raha diso ny valiny,dia lazaina hoe]: **Vakiana hoe “s” toy ny amin’ny teny hoe ‘’sa’’ io litera io.**

**Hijery ohatra hafa indray isika. Vakio ity litera mifampibaby ity.** [atoro azy ny ‘’tr’’]

[Raha marina ny valiny, dia tenenina hoe]: **Tsara be, vakiana hoe “tr” toy ny amin’ny teny hoe ‘’trano’’ io litera mifampibaby io.**

[Raha diso ny valiny, lazaina hoe]: **Vakiana hoe “tr” toy ny amin’ny teny hoe ‘’trano’’ io litera mifampibaby io.**

**Rehefa miteny aho hoe "Atombohy", dia manomboka mamaky eto ianao** [tondroy amin'ny fanondro ny litera voalohany], **dia kendreo tsara ny famakianao azy isan'andalana ka avy any ankavia miankavanana.**

**Vakio tsara sy haingana ary ataovy mafy.**

**Tondroy amin'ny fanondronao ireo litera rehefa mamaky ianao. Vakio tsara sy haingana. Raha misy tsy hainao, dia dingano dia tohizo ny vakiteninao.**

**Tondroy amin'ny fanondronao ny litera voalohany.**

**Vonona ianao?**

**Atombohy àry.**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| p | v | E | dr | o | d | R | f | ts | ntr |
| i | A | F | nts | t | mp | K | a | tr | M |
| m | y | ndr | ng | j | e | mb | g | nj | I |
| nk | L | nd | H | z | b | Y | R | k | T |
| TR | nt | z | S | h | M | n | E | O | N |
| h | O | A | e | m | O | t | r | i | v |
| nd | R | V | S | a | i | K | j | n | mb |
| i | O | s | I | tr | b | y | o | nt | L |
| p | f | A | Y | k | N | g | a | H | n |
| ts | ng | N | a | mp | o | nj | I | d | N |

Time Remaining

Autostop?

---

*Syllabes - Famantarana vaninteny*

**Inty misy taratasy ahitana vaninteny. Omeo ahy ny feo hanononana ireo vaninteny ireo .**

**Ohatra, ity vaninteny ity** [Atoro azy ny ''hi''] **vakiana hoe "hi".**

**Andao handramantsika. Vakio ity vaninteny ity.** [Atoro azy ny ''la'']

[Raha marina ny valiny, dia tenenina hoe Tsara be]: **Vakiana hoe "la".**

[Raha diso ny valiny,dia lazaina hoe]: **Vakiana hoe "la".**

**Hijery ohatra hafa indray isika. Vakio ity vaninteny ity.** [atoro azy ny ''se'']

[Raha marina ny valiny, dia tenenina hoe]: **Tsara be, vakiana hoe "se".**

[Raha diso ny valiny,dia lazaina hoe]: **Vakiana hoe "se".**

**Rehefa miteny aho hoe "Atombohy",dia manomboka mamaky eto ianao** [tondroy amin'ny fanondro ny vaninteny voalohany], **dia kendreo tsara ny famakianao azy isan'andalana ka avy any ankavia no miankavanana. Vakio tsara sy haingana ary ataovy mafy.**

**Tondroy amin'ny fanondronao ireo vaninteny rehefa mamaky ianao. Vakio tsara sy haingana. Raha misy tsy hainao, dia dingano dia tohizo ny vakiteninao.**

**Tondroy amin'ny fanondronao ny vaninteny voalohany.**

**Vonona ianao?**

**Atombohy àry.**

| | | | | |
|---|---|---|---|---|
| do | mi | dy | tro | mpo |
| ki | da | zi | li | vo |
| ndra | ky | pa | hy | fi |
| ha | ri | tsa | po | za |
| si | ly | so | ndro | vi |
| nta | le | mpi | ni | nga |
| bo | ti | zy | tso | ndri |
| ne | ngo | nka | ke | di |
| tra | me | nja | mpa | nda |
| tsi | mbo | ntsi | mo | ma |

Time Remaining

Autostop?

---

*Mots familiers - Famakiana teny mahazatra*

**Inty misy taratasy ahitana teny maromaro. Ezaho vakiana araka izay vitanao ireo teny ireo.**

**Tsy tononina fotsiny ny litera fa vakiana mihitsy ny teny.**

**Ohatra : vakiana hoe ‘’nono’’ fa tsy hoe ‘’n’’, ‘’o’’, ‘’n’’, ‘’o’’**

**Andao handramantsika. Vakio ity teny ity** [Atoro azy ny ‘’faly’’]

[Raha marina ny valiny, dia tenenina hoe]: **Tsara be, vakiana hoe “faly”’io teny io.**

[Raha diso ny valiny, dia lazaina hoe]: **Vakiana hoe “faly”’io teny io fa tsy izany.**

**Hijery ohatra hafa indray isika. Vakio ity teny ity** [atoro azy ny teny hoe ‘’ireo’’].

[Raha marina ny valiny, dia tenenina hoe:] **Tsara be, vakiana hoe “ireo’io teny io.**

[Raha diso ny valiny, dia lazaina hoe]: **Vakiana hoe “ireo” io teny io fa tsy izany.**

**Ao tsara ve (na mety!) ? Afaka manohy isika?**

**Raha vao miteny aho hoe:" atombohy", dia tondroy amin'ny fanondro ny teny tsirairay rehefa mamaky azy ianao.**

**Kendreo tsara ny famakiana azy isan'andalana ka avy any ankavia no miankavanana. Tsy hiteny na inona na inona aho amin'ity indray mitoraka ity. Ianao irery no mamaky misesy ireo teny ireo.**

**Azonao tsara ve? Apetraho eo amin'ny teny voalohany ny fanondronao.**

**Vonona ve ianao? Vakio tsara sy mafy.**

**Atombohy àry.**

| | | | | |
|---|---|---|---|---|
| bota | zaza | mena | mody | loha |
| salama | sary | valo | boky | hena |
| lava | lay | may | folo | jiro |
| roa | kibo | vato | kapa | tanana |
| nify | mavo | lakana | sira | reraka |
| fohy | lova | saka | vary | lela |
| neny | izy | volo | bika | mofo |

Time Remaining

Autostop?

---

*Histoire 1 min*

**Izao an! Inty misy tantara. Vakio tsara, mafy sy haingana araka izay vitanao. Rehefa miteny aho hoe : ‘Atombohy  ny vakiteny’, dia manomboka ianao.**

**Raha misy teny tsy hainao vakiana dia dingano ary tohizo ny manaraka.**

**Apetraho eo amin’ny teny voalohany ny fanondronao.**

**Vonona ve ianao?**

**Atombohy àry**.

| Tia | madio | i | Bema. | Mifoha | maraina | izy. |
|---|---|---|---|---|---|---|
| Mangatsiaka | anefa | amin’ | ny | maraina. | Mila | midio |
| izy | alohan’ | ny | hianatra. | Lasa | ihany | i |
| Bema | midio | amin’ny | rano | madio | sy | savony. |
| Manasa | ny | tarehiny | sy | ny | tanany | ary |
| ny | tongony | i | Bema. | Zaza | madio | i |
| Bema. | Ravoravo | i | Dada. | | | |

Time Remaining

Autostop?

---

*Histoire 2 min*

**Izao an! Inty misy tantara.**

**Vakio tsara, mafy sy haingana araka izay vitanao. Hametraka fanontaniana vitsivitsy aminao aho rehefa avy eo.**

**Rehefa miteny aho hoe ‘Atombohy’ ny vakiteny dia manomboka ianao.**

**Raha misy teny tsy hainao vakina dia dingano ary tohizo ny manaraka.**

**Apetraho eo amin’ny teny voalohany ny fanondronao.**

**Vonona ve ianao?**

**Atombohy àry**.

| Asabotsy | ny | andro. | Mankany | anaty | saha | i | Levelo |
|---|---|---|---|---|---|---|---|
| sy | Dada. | Vokatra | ny | voasary | ao | anaty | saha. |
| Sarotra | alaina | ireo | voasary | masaka. | Mitoraka | voasary | i |
| Dada. | Taitra | ny | fanenitra. | Voakaikitra | i | Levelo. | Mitomany |
| be | i | Levelo. | Nasian’ | i | Dada | ravina | ny |
| sofiny | voakaikitra. | | | | | | |

Time Remaining

Autostop?

---

*Comprehension de l'histoire*

**Izao an!**

**Hametraka fanontaniana vitsivitsy aminao aho.**

**Valio araka izay ahaizanao azy ny fanontaniana.**

**Azonao jerena ao anatin’ny tantara ny valiny.**

Nankaiza i Levelo sy Dada? [Nankany an-tsaha]

☐ Marina ☐ Diso ☐ Tsy namaly

Inona ny voankazo vokatra ao anaty saha? [Voasary]

☐ Marina ☐ Diso ☐ Tsy namaly

Inona no ataon’ i Dada ahazoana voasary ? [Mitoraka]

☐ Marina ☐ Diso ☐ Tsy namaly

Maninona no mitomany Levelo ? [voakaikitra]

☐ Marina ☐ Diso ☐ Tsy namaly

Nahoana no asian’ i Dada ravina ny sofin’ i Levelo? [Marary/mivonto/mba ho sitrana]

☐ Marina ☐ Diso ☐ Tsy namaly

---

*Comprehension Orale*

Izao dia izaho indray no hamaky tantara aminao. Hovakiko aminao indray mandeha ihany ny tantara koa mihainoa tsara. Avy eo dia valio araka izay ahaizanao azy ny fanontaniana.

---

## Ao amin’ ny taona faharoa i Toky.

## Manomboka ny fampianarana mamaky teny i Ramose.

## Anjaran’ i Toky izao no mamaky teny.

## Variana mitabataba anefa i Toky ka tsy hitany akory Ramose manondro azy.

## Taitra i Toky. Soa ihany fa mamerina ny teniny i Ramose.

## Miezaka i Toky. Afaka mamaky teny amin’ ny feo mafy ihany i Toky.

---

Ireto àry ny fanontaniana.

Ao amin’ ny taona fahafiry i Toky ? [taona faharoa]

☐ Marina ☐ Diso ☐ Tsy namaly

Inona no ampianarin-dramose ? [ mamaky teny ]

☐ Marina ☐ Diso ☐ Tsy namaly

Inona no mahavariana an' i Toky ? [ mitabataba ]

☐ Marina ☐ Diso ☐ Tsy namaly

Inona no nataon-dramose rehefa taitra i Toky ? [ namerina ny teniny ]

☐ Marina ☐ Diso ☐ Tsy namaly

Ahoana no nahatonga an' i Toky afaka mamaky teny ihany ? [ satria namerina ny teniny Ramose / niezaka izy nanatanteraka ny tenin-dramose ]

☐ Marina ☐ Diso ☐ Tsy namaly

---

*Questionnaire élève - Mombamomba ny mpianatra*

0. Iza ny anaranao ?

_____

1. Lahy sa Vavy ?

☐ Lahy ☐ Vavy

2.a. Kilasy fahafiry ianao izao?

☐ T1 ☐ Tsy namaly

2.b. Kilasy fizarana misy anao

☐ Section A ☐ section B ☐ Section C ☐ Section D ☐ Section E ☐ Section F

3. Firy taona ianao ?

_____

4.a. Kilasy faha-firy ianao tamin'ny taona lasa?

☐ Tsy nianatra ☐ Préscolaire/Maternelle ☐ T1 ☐ Tsy namaly

4.b. Efa nanao maternelle ve ianao talohan' ny nidirana T1? ([Raha "T1" ny valiny dia apetraka ny fanontaniana b] )

☐ Eny ☐ Tsia ☐ Tsy namaly

5.a. Manana boky famakian-teny ve ianao ao an-dakilasy?

☐ Eny ☐ Tsia

5.b. Azonao entina mody ve?

☐ Eny ☐ Tsia ☐ Tsy namaly

6. Ankoatra ny boky fianarana, misy boky hafa na gazety azonao vakiana ve any an-trano?

☐ Eny ☐ Tsia ☐ TSy namaly

7. Misy olona mahay mamaky teny ve ao an-tranonareo?

☐ Eny ☐ Tsia ☐ Tsy namaly

8. Manome asa entimody ve ny mpampianatra anao?

☐ Eny ☐ Tsia ☐ Tsy namaly

9. Misy olon-kafa ve manampy anao amin'ny fanaovana ny asa entimody?

☐ Eny ☐ Tsia ☐ Tsy namaly

10. Impiry misakafo ianao isan’ andro?

☐ Indray ☐ Indroa ☐ Intelo

11. Misy an’ireto ve any an-tranonareo?

☐ Radio ☐ Tele ☐ Telefaonina ☐ Bisikileta ☐ Moto ☐ Fiarakodia ☐ Rano amin’ ny paompy ao an-trano

☐ Jiro avy amin’ ny herinaratra ☐ Kabone ☐ Tsy manana ☐ Tsy namaly

12. Sekoly miankina sa sekoly tsy miankina

☐ Sekoly miankina ☐ Sekoly tsy miankina

13. Sekoly ambonivohitra sa ambanivohitra?

☐ Ambonivohitra ☐ Ambanivohitra

14. Fandaharana maraina sa ny tolakandro ?

☐ Maraina ☐ Tolakandro

---

*Coordonnées GPS*

---

# ANNEX III. STANDARDS AND BENCHMARKS

Table 5. Performance benchmarks for reading (Malagasy), T1

| Composante | | Comment la mesurer | Catégorie de performance | | | | Seuil minimal |
|---|---|---|---|---|---|---|---|
| | | | Débutant | Émergent | Compétent | Performant | |
| Compréhension | Compréhension à l'oral/à l'audition | % d'informations qu'un enfant retire d'un texte à son niveau qui lui est lu à l'oral L'évaluation se fait à la base des réponses aux questions posé par l'évaluateur sur le texte. | 0 à 20% | 20 à 60% | 80% | 100% | 80% |
| | Compréhension du texte écrit | % d'informations un enfant retire d'un texte à son niveau (T1) d'environs 60 mots qu'il lit lui-même pour la première fois. | 0 à 20% | 20 à 60% | 80% | 100% | 80% |
| Lecture, texte continu | Fluidité de lecture d'un texte continu | Nombre de mots correctement lus – à haute voix - d'un texte continu d'environs 60 mots et à la portée de l'élève, dans un délai de 1 minute. | 0 à 5 mclm | 6 à 14 mclm | 15 à 30 mclm | 31+ mclm | 15 mclm |
| Lecture des mots familiers | Fluidité de lecture des mots familiers | Nombre de mots familiers correctement lus – à haute voix - dans un délai de 1 minute. | 0 à 5 mclm | 6 à 14 mclm | 15 à 30 mlcm | 31 + mclm | 15 mlcm |
| Conscience alphabétique | Fluidité de lecture des lettres et des syllabes | Nombre de lettres, combinaisons de lettres ou syllabes correctement lus[13] – à haute voix - dans un délai de 1 minute | 0 à 14 lclm | 15 à 39 lclm | 30 à 44 lclm | 45+lclm | 30 lclm |

[13] Les lettres et syllabes présentés à l'élève se limitent à ceux étudiés au cours de l'année en question. Il y a lieu de préciser les lettres et combinaisons de lettres ciblées pour le T1 et le T2.

# ANNEX IV. FULL REGRESSION RESULTS

**Table A1. Marginal effects - Letter sounds**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | 2.73*** | 2.72* | 3.13** | 2.90 | 1.64 |
| | (0.95) | (1.46) | (1.27) | (2.26) | (2.09) |
| Teacher followed most/all steps | 3.36*** | 3.49** | 2.92** | 1.19 | 5.08** |
| | (0.95) | (1.54) | (1.26) | (2.19) | (2.07) |
| **Student characteristics:** | | | | | |
| Age | 0.16 | 0.24 | 0.15 | -0.05 | 0.33 |
| | (0.25) | (0.40) | (0.32) | (0.37) | (1.12) |
| Books available in class | -0.92 | -0.68 | -1.85 | -0.70 | -4.01 |
| | (1.93) | (2.83) | (2.00) | (2.59) | (3.15) |
| Books can be taken home | -0.30 | 0.22 | -0.95 | 0.57 | -0.07 |
| | (0.63) | (1.10) | (0.78) | (1.02) | (2.02) |
| Other books at home | 1.07** | 0.92 | 1.07* | 0.32 | 2.84** |
| | (0.45) | (0.75) | (0.56) | (0.70) | (1.25) |
| Other family members can read | 0.41 | 0.03 | 1.01 | 0.58 | 0.11 |
| | (0.62) | (0.89) | (0.95) | (1.08) | (2.00) |
| Teacher assigns homework | -0.82 | -1.61 | -0.16 | -0.70 | 1.09 |
| | (0.62) | (1.07) | (0.73) | (1.11) | (1.59) |
| Gets help with homework at home | 0.90* | 1.16 | 0.70 | 1.13 | 0.46 |
| | (0.54) | (1.04) | (0.55) | (0.89) | (1.54) |
| Number of meals per day | -0.53 | -0.51 | -0.69 | -0.24 | -1.83 |
| | (0.54) | (1.09) | (0.53) | (0.77) | (1.88) |
| Wealth index | 0.06 | 0.32 | -0.14 | 0.39 | -0.26 |
| | (0.28) | (0.51) | (0.34) | (0.89) | (1.13) |
| Observations | 1776 | 805 | 968 | 1017 | 756 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A2. Marginal effects - Syllable identification**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | 0.91 | 0.81 | 1.15 | 1.23 | 0.28 |
| | (0.73) | (1.25) | (0.91) | (1.55) | (1.73) |
| Teacher followed most/all steps | 3.74*** | 3.92*** | 3.57*** | 1.50 | 5.97*** |
| | (0.75) | (1.30) | (0.98) | (1.57) | (1.81) |
| **Student characteristics:** | | | | | |
| Age | -0.01 | -0.03 | 0.03 | 0.04 | -0.39 |
| | (0.19) | (0.30) | (0.27) | (0.31) | (0.65) |
| Books available in class | -1.09 | 0.46 | -1.54 | -0.72 | -3.85 |
| | (1.27) | (2.45) | (1.40) | (2.12) | (2.46) |
| Books can be taken home | 0.50 | 0.92 | 0.00 | 1.43* | 0.17 |
| | (0.48) | (0.86) | (0.59) | (0.83) | (1.32) |
| Other books at home | 1.41*** | 0.92 | 1.79*** | 1.02* | 2.24** |
| | (0.39) | (0.68) | (0.44) | (0.60) | (0.95) |
| Other family members can read | -0.22 | -0.20 | -0.12 | -0.42 | 0.33 |
| | (0.54) | (0.82) | (0.75) | (0.83) | (2.03) |
| Teacher assigns homework | -0.92* | -1.47 | -0.49 | -0.48 | -0.29 |
| | (0.51) | (0.90) | (0.62) | (0.91) | (1.35) |
| Gets help with homework at home | 0.55 | 0.68 | 0.43 | 0.70 | 0.39 |
| | (0.47) | (0.87) | (0.50) | (0.76) | (1.15) |
| Number of meals per day | -0.36 | -0.26 | -0.50 | -0.12 | -1.04 |
| | (0.42) | (0.77) | (0.51) | (0.73) | (1.19) |
| Wealth index | 0.00 | 0.26 | -0.20 | 0.06 | 1.04 |
| | (0.24) | (0.46) | (0.27) | (0.74) | (0.75) |
| Observations | 1777 | 806 | 968 | 1018 | 756 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A3. Marginal effects - Familiar words**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | 0.81 | 0.67 | 0.95 | 1.60 | -0.16 |
| | (0.57) | (1.06) | (0.68) | (1.21) | (1.33) |
| Teacher followed most/all steps | 1.56*** | 1.47 | 1.57** | -0.17 | 3.00** |
| | (0.59) | (1.11) | (0.74) | (1.23) | (1.31) |
| **Student characteristics:** | | | | | |
| Age | 0.14 | 0.22 | 0.12 | 0.09 | -0.20 |
| | (0.15) | (0.23) | (0.21) | (0.26) | (0.47) |
| Books available in class | -1.18 | 2.69 | -1.86 | -1.04 | -4.33** |
| | (1.14) | (2.95) | (1.24) | (1.93) | (1.87) |
| Books can be taken home | 0.48 | 0.48 | 0.39 | 1.20* | 0.81 |
| | (0.41) | (0.69) | (0.53) | (0.65) | (1.08) |
| Other books at home | 1.07*** | 0.82* | 1.21*** | 0.38 | 1.57* |
| | (0.29) | (0.45) | (0.38) | (0.50) | (0.80) |
| Other family members can read | -0.35 | -0.51 | -0.26 | -0.50 | 0.15 |
| | (0.42) | (0.62) | (0.59) | (0.66) | (1.62) |
| Teacher assigns homework | -0.67 | -0.77 | -0.51 | -0.19 | -0.76 |
| | (0.42) | (0.72) | (0.51) | (0.77) | (1.06) |
| Gets help with homework at home | 0.34 | 0.68 | 0.08 | 0.54 | 0.31 |
| | (0.34) | (0.61) | (0.38) | (0.64) | (0.88) |
| Number of meals per day | 0.03 | 0.07 | -0.16 | 0.15 | -0.67 |
| | (0.35) | (0.69) | (0.38) | (0.49) | (0.89) |
| Wealth index | -0.04 | -0.09 | -0.03 | -0.04 | 0.22 |
| | (0.19) | (0.36) | (0.22) | (0.61) | (0.56) |
| Observations | 1776 | 805 | 968 | 1017 | 756 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A4. Marginal effects - Oral reading fluency - 1 min**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | 0.24 | 0.30 | 0.34 | 0.45 | -1.10 |
| | (0.57) | (1.16) | (0.62) | (1.19) | (1.45) |
| Teacher followed most/all steps | 2.41*** | 2.35* | 2.13*** | 1.34 | 3.98*** |
| | (0.60) | (1.24) | (0.63) | (1.24) | (1.48) |
| **Student characteristics:** | | | | | |
| Age | -0.18 | -0.30 | -0.02 | -0.41 | -0.15 |
| | (0.20) | (0.38) | (0.17) | (0.43) | (0.55) |
| Books available in class | -0.31 | -0.71 | -0.46 | -1.63 | -2.24 |
| | (0.86) | (2.33) | (0.86) | (1.35) | (2.29) |
| Books can be taken home | 0.37 | 0.67 | 0.01 | 0.94 | 0.56 |
| | (0.44) | (0.85) | (0.48) | (0.74) | (1.18) |
| Other books at home | 1.25*** | 1.53** | 0.99*** | 1.15 | 1.32 |
| | (0.34) | (0.61) | (0.37) | (0.76) | (0.90) |
| Other family members can read | -0.15 | -0.01 | -0.27 | -0.35 | -0.36 |
| | (0.44) | (0.73) | (0.53) | (0.64) | (1.91) |
| Teacher assigns homework | -1.06** | -1.26 | -0.85* | -0.62 | -0.97 |
| | (0.46) | (0.89) | (0.44) | (0.87) | (1.27) |
| Gets help with homework at home | 0.28 | 0.48 | 0.08 | 0.60 | 0.89 |
| | (0.40) | (0.82) | (0.37) | (0.84) | (1.05) |
| Number of meals per day | -0.03 | 0.43 | -0.39 | 0.08 | 0.22 |
| | (0.41) | (0.91) | (0.36) | (0.56) | (1.30) |
| Wealth index | -0.24 | -0.41 | -0.08 | -0.30 | 0.42 |
| | (0.19) | (0.36) | (0.20) | (0.68) | (0.61) |
| Observations | 1767 | 800 | 964 | 1011 | 753 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A5. Marginal effects - Oral reading fluency - 2 min**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | 1.07** | -0.17 | 1.83*** | 1.91* | 0.44 |
| | (0.54) | (0.96) | (0.68) | (1.15) | (1.47) |
| Teacher followed most/all steps | 1.43** | 2.57** | 0.84 | -0.48 | 2.85* |
| | (0.56) | (1.04) | (0.71) | (1.17) | (1.49) |
| **Student characteristics:** | | | | | |
| Age | 0.00 | 0.02 | 0.01 | 0.02 | -0.41 |
| | (0.12) | (0.20) | (0.15) | (0.21) | (0.39) |
| Books available in class | -1.15 | -3.17* | -1.11 | -1.26 | -2.42 |
| | (0.80) | (1.62) | (0.90) | (1.42) | (1.98) |
| Books can be taken home | 0.11 | 0.11 | -0.04 | 0.95 | -0.09 |
| | (0.35) | (0.61) | (0.44) | (0.61) | (0.89) |
| Other books at home | 0.83*** | 0.98** | 0.64* | 0.57 | 1.07 |
| | (0.25) | (0.39) | (0.33) | (0.47) | (0.72) |
| Other family members can read | -0.04 | 0.13 | -0.09 | -0.06 | -0.91 |
| | (0.39) | (0.59) | (0.52) | (0.65) | (1.53) |
| Teacher assigns homework | -0.58* | -0.64 | -0.54 | -0.18 | -0.33 |
| | (0.32) | (0.57) | (0.37) | (0.55) | (0.85) |
| Gets help with homework at home | 0.22 | 0.03 | 0.32 | 0.35 | 0.60 |
| | (0.26) | (0.45) | (0.30) | (0.44) | (0.77) |
| Number of meals per day | -0.15 | 0.42 | -0.47 | 0.11 | -0.03 |
| | (0.29) | (0.59) | (0.32) | (0.52) | (0.71) |
| Wealth index | 0.09 | 0.14 | 0.05 | 0.23 | 0.48 |
| | (0.14) | (0.25) | (0.17) | (0.51) | (0.45) |
| Observations | 1734 | 776 | 955 | 991 | 740 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A6. Marginal effects - Reading comprehension**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | 7.50 | 9.95 | 6.74 | 14.93 | -2.45 |
| | (6.16) | (10.33) | (7.88) | (16.75) | (11.54) |
| Teacher followed most/all steps | 0.84 | -2.98 | 0.19 | -3.11 | 2.39 |
| | (6.22) | (9.98) | (7.48) | (16.42) | (10.48) |
| **Student characteristics:** | | | | | |
| Age | -0.91 | -0.26 | -1.28 | -2.68 | -1.42 |
| | (1.62) | (2.23) | (2.18) | (4.51) | (3.72) |
| Books available in class | 2.36 | 4.72 | -4.09 | -7.96 | -3.26 |
| | (8.57) | (7.79) | (9.04) | (20.39) | (21.89) |
| Books can be taken home | 0.36 | 4.58 | -2.96 | 1.01 | 7.37 |
| | (3.79) | (6.11) | (4.80) | (9.15) | (7.82) |
| Other books at home | 6.24** | 7.54** | 5.57 | 9.02 | 6.11 |
| | (2.52) | (3.43) | (3.60) | (6.63) | (5.83) |
| Other family members can read | 1.77 | 5.70 | -1.29 | -3.57 | 13.34 |
| | (5.74) | (7.42) | (6.30) | (6.79) | (22.34) |
| Teacher assigns homework | 0.27 | -2.40 | -1.40 | 7.16 | 0.43 |
| | (4.36) | (6.46) | (6.22) | (9.72) | (8.81) |
| Gets help with homework at home | -0.59 | -0.09 | 0.93 | -0.49 | -2.43 |
| | (3.47) | (5.19) | (5.18) | (7.29) | (8.16) |
| Number of meals per day | 0.91 | 8.65 | -5.49 | 0.13 | 3.20 |
| | (3.88) | (5.28) | (4.75) | (8.41) | (10.08) |
| Wealth index | -2.45 | -4.69 | 0.01 | 0.10 | -5.06 |
| | (2.20) | (3.45) | (3.29) | (9.48) | (4.80) |
| Observations | 966 | 456 | 509 | 503 | 462 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A7. Marginal effects - Listening comprehension**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | -9.33** | -8.57 | -8.66 | -11.47 | -13.76 |
| | (4.52) | (7.01) | (6.00) | (8.29) | (9.90) |
| Teacher followed most/all steps | 9.19** | 4.88 | 10.66* | 12.75 | 10.34 |
| | (4.44) | (7.26) | (5.83) | (8.09) | (9.64) |
| **Student characteristics:** | | | | | |
| Age | -0.22 | -2.41 | 1.70 | 0.12 | 0.31 |
| | (1.20) | (1.71) | (1.73) | (2.28) | (2.68) |
| Books available in class | -16.62 | 10.10* | -19.69 | -18.59 | 0.13 |
| | (10.89) | (6.03) | (12.13) | (23.39) | (13.49) |
| Books can be taken home | -0.62 | 0.33 | -0.89 | -2.68 | 3.12 |
| | (2.65) | (4.45) | (3.47) | (4.56) | (6.12) |
| Other books at home | 5.49*** | 5.68** | 5.29* | 6.34* | 6.93 |
| | (1.97) | (2.86) | (2.84) | (3.79) | (4.40) |
| Other family members can read | 0.48 | -0.92 | 3.35 | 0.12 | -2.87 |
| | (3.97) | (5.06) | (6.49) | (6.19) | (12.12) |
| Teacher assigns homework | 1.22 | -2.90 | 3.66 | 4.75 | -0.25 |
| | (2.76) | (4.18) | (3.79) | (4.66) | (7.18) |
| Gets help with homework at home | -3.18 | -1.14 | -5.05 | 0.17 | -3.60 |
| | (2.27) | (3.18) | (3.25) | (4.08) | (5.71) |
| Number of meals per day | 0.57 | 6.15 | -3.40 | 1.44 | 0.44 |
| | (2.52) | (3.78) | (3.26) | (4.55) | (6.67) |
| Wealth index | 0.86 | 2.00 | 0.96 | 0.84 | 2.18 |
| | (1.37) | (2.33) | (1.75) | (4.77) | (3.72) |
| Observations | 1699 | 774 | 923 | 954 | 743 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A8. Marginal effects - Letter sounds: minimal proficiency**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | 0.03 | 0.11* | 0.01 | 0.02 | 0.03 |
| | (0.03) | (0.06) | (0.02) | (0.05) | (0.08) |
| Teacher followed most/all steps | 0.04 | -0.01 | 0.05 | 0.03 | 0.04 |
| | (0.03) | (0.06) | (0.03) | (0.05) | (0.08) |
| **Student characteristics:** | | | | | |
| Age | 0.00 | 0.00 | 0.01 | -0.00 | 0.00 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.03) |
| Books available in class | 0.02 | -0.07 | 0.01 | 0.01 | 0.08 |
| | (0.02) | (0.06) | (0.01) | (0.02) | (0.10) |
| Books can be taken home | 0.01 | 0.01 | -0.00 | -0.01 | 0.02 |
| | (0.02) | (0.04) | (0.02) | (0.03) | (0.07) |
| Other books at home | 0.00 | 0.01 | -0.00 | 0.01 | 0.00 |
| | (0.01) | (0.02) | (0.01) | (0.02) | (0.04) |
| Other family members can read | -0.01 | -0.06** | 0.05 | 0.01 | 0.03 |
| | (0.02) | (0.03) | (0.03) | (0.02) | (0.11) |
| Teacher assigns homework | -0.03* | -0.05 | -0.03* | -0.02 | -0.05 |
| | (0.02) | (0.03) | (0.02) | (0.03) | (0.06) |
| Gets help with homework at home | 0.02 | 0.03 | 0.03 | 0.02 | 0.02 |
| | (0.02) | (0.03) | (0.02) | (0.02) | (0.06) |
| Number of meals per day | -0.02 | -0.04 | -0.01 | -0.01 | -0.04 |
| | (0.01) | (0.03) | (0.01) | (0.01) | (0.06) |
| Wealth index | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.03) |
| Observations | 1777 | 806 | 968 | 1018 | 756 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A9 Marginal effects - Syllable identification: minimal proficiency**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | 0.01 | 0.03 | 0.01 | -0.01 | -0.02 |
| | (0.03) | (0.06) | (0.02) | (0.01) | (0.09) |
| Teacher followed most/all steps | 0.03 | 0.03 | 0.01 | 0.04* | 0.05 |
| | (0.03) | (0.07) | (0.02) | (0.02) | (0.09) |
| **Student characteristics:** | | | | | |
| Age | 0.01 | 0.01 | 0.01** | 0.01 | 0.02 |
| | (0.00) | (0.01) | (0.00) | (0.01) | (0.02) |
| Books available in class | 0.01 | -0.08* | 0.01 | -0.00 | 0.00 |
| | (0.01) | (0.05) | (0.01) | (0.01) | (0.10) |
| Books can be taken home | 0.03* | 0.05* | 0.00 | 0.01 | 0.04 |
| | (0.01) | (0.03) | (0.01) | (0.02) | (0.05) |
| Other books at home | 0.00 | -0.01 | 0.01 | -0.02 | 0.01 |
| | (0.01) | (0.02) | (0.01) | (0.02) | (0.03) |
| Other family members can read | -0.01* | -0.01 | -0.00 | -0.00 | -0.03 |
| | (0.01) | (0.01) | (0.00) | (0.01) | (0.03) |
| Teacher assigns homework | -0.03** | -0.05** | -0.02 | -0.02 | -0.05 |
| | (0.01) | (0.03) | (0.01) | (0.02) | (0.05) |
| Gets help with homework at home | 0.03** | 0.06*** | 0.01 | 0.03* | 0.05 |
| | (0.01) | (0.02) | (0.02) | (0.02) | (0.05) |
| Number of meals per day | -0.01 | -0.00 | -0.01 | 0.00 | -0.01 |
| | (0.01) | (0.03) | (0.01) | (0.02) | (0.05) |
| Wealth index | -0.00 | -0.01 | -0.00 | -0.01 | -0.01 |
| | (0.00) | (0.01) | (0.01) | (0.01) | (0.02) |
| Observations | 1777 | 806 | 968 | 1018 | 756 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A10. Marginal effects - Familiar words: minimal proficiency**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | -0.01 | 0.02 | -0.02 | 0.05 | -0.10 |
| | (0.03) | (0.06) | (0.04) | (0.06) | (0.08) |
| Teacher followed most/all steps | 0.12*** | 0.11* | 0.11*** | 0.06 | 0.21** |
| | (0.03) | (0.06) | (0.04) | (0.07) | (0.08) |
| **Student characteristics:** | | | | | |
| Age | 0.01 | 0.00 | 0.01 | 0.01 | 0.00 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.03) |
| Books available in class | -0.07 | 0.20 | -0.11 | 0.01 | -0.10 |
| | (0.09) | (0.17) | (0.10) | (0.08) | (0.11) |
| Books can be taken home | 0.00 | -0.04 | 0.03 | 0.04 | 0.06 |
| | (0.03) | (0.04) | (0.03) | (0.04) | (0.07) |
| Other books at home | 0.03* | 0.05 | 0.02 | 0.03 | 0.06 |
| | (0.02) | (0.03) | (0.02) | (0.03) | (0.05) |
| Other family members can read | 0.01 | -0.01 | 0.03 | 0.04 | 0.06 |
| | (0.03) | (0.04) | (0.03) | (0.04) | (0.09) |
| Teacher assigns homework | -0.04* | -0.01 | -0.06** | -0.00 | -0.10 |
| | (0.03) | (0.05) | (0.03) | (0.04) | (0.08) |
| Gets help with homework at home | 0.05** | 0.06 | 0.05* | 0.06** | 0.06 |
| | (0.02) | (0.04) | (0.03) | (0.03) | (0.07) |
| Number of meals per day | -0.01 | -0.04 | -0.01 | -0.01 | -0.09* |
| | (0.02) | (0.03) | (0.02) | (0.02) | (0.05) |
| Wealth index | 0.00 | -0.00 | 0.00 | -0.03 | -0.02 |
| | (0.01) | (0.02) | (0.01) | (0.03) | (0.05) |
| Observations | 1777 | 806 | 968 | 1018 | 756 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A11. Marginal effects - Oral reading fluency - 1 min: minimal proficiency**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | -0.01 | -0.03 | 0.00 | 0.06 | -0.11 |
| | (0.03) | (0.05) | (0.04) | (0.06) | (0.08) |
| Teacher followed most/all steps | 0.09*** | 0.14*** | 0.03 | 0.01 | 0.18** |
| | (0.03) | (0.05) | (0.03) | (0.06) | (0.08) |
| **Student characteristics:** | | | | | |
| Age | 0.00 | -0.00 | 0.00 | -0.01 | 0.01 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.03) |
| Books available in class | 0.02 | 0.11 | 0.02 | 0.01 | -0.06 |
| | (0.02) | (0.15) | (0.02) | (0.04) | (0.11) |
| Books can be taken home | 0.03 | 0.05 | 0.01 | 0.02 | 0.06 |
| | (0.02) | (0.04) | (0.03) | (0.04) | (0.07) |
| Other books at home | 0.02 | 0.03 | 0.01 | 0.04 | 0.04 |
| | (0.02) | (0.03) | (0.02) | (0.03) | (0.06) |
| Other family members can read | 0.01 | 0.04 | -0.00 | -0.01 | -0.00 |
| | (0.02) | (0.03) | (0.02) | (0.03) | (0.07) |
| Teacher assigns homework | -0.02 | 0.00 | -0.03 | 0.04 | -0.07 |
| | (0.02) | (0.04) | (0.02) | (0.03) | (0.07) |
| Gets help with homework at home | 0.02 | 0.02 | 0.02 | 0.02 | 0.06 |
| | (0.02) | (0.03) | (0.02) | (0.02) | (0.07) |
| Number of meals per day | 0.01 | 0.03 | -0.00 | 0.02 | -0.02 |
| | (0.02) | (0.04) | (0.02) | (0.02) | (0.06) |
| Wealth index | -0.00 | -0.01 | 0.01 | 0.01 | -0.02 |
| | (0.01) | (0.02) | (0.01) | (0.03) | (0.04) |
| Observations | 1777 | 806 | 968 | 1018 | 756 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A12. Marginal effects - Oral reading fluency - 2 min: minimal proficiency**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | 0.01 | -0.00 | 0.02 | 0.06 | -0.07 |
| | (0.03) | (0.06) | (0.03) | (0.07) | (0.08) |
| Teacher followed most/all steps | 0.06* | 0.08 | 0.03 | 0.00 | 0.13* |
| | (0.03) | (0.06) | (0.04) | (0.07) | (0.08) |
| **Student characteristics:** | | | | | |
| Age | 0.00 | 0.01 | -0.00 | 0.01 | 0.01 |
| | (0.01) | (0.01) | (0.01) | (0.01) | (0.03) |
| Books available in class | 0.01 | -0.05 | 0.01 | -0.01 | -0.06 |
| | (0.02) | (0.07) | (0.02) | (0.03) | (0.11) |
| Books can be taken home | 0.02 | 0.02 | 0.00 | 0.02 | 0.06 |
| | (0.02) | (0.04) | (0.03) | (0.04) | (0.06) |
| Other books at home | 0.02 | 0.03 | 0.01 | 0.02 | 0.01 |
| | (0.02) | (0.03) | (0.02) | (0.03) | (0.05) |
| Other family members can read | 0.01 | 0.02 | 0.02 | 0.04 | 0.01 |
| | (0.02) | (0.04) | (0.03) | (0.04) | (0.07) |
| Teacher assigns homework | -0.00 | 0.01 | -0.02 | 0.01 | -0.01 |
| | (0.02) | (0.04) | (0.02) | (0.03) | (0.07) |
| Gets help with homework at home | 0.02 | 0.01 | 0.03 | 0.02 | 0.05 |
| | (0.02) | (0.03) | (0.02) | (0.03) | (0.06) |
| Number of meals per day | 0.00 | 0.03 | -0.02 | 0.02 | -0.02 |
| | (0.02) | (0.04) | (0.01) | (0.03) | (0.05) |
| Wealth index | 0.00 | -0.01 | 0.01 | -0.01 | -0.03 |
| | (0.01) | (0.02) | (0.01) | (0.04) | (0.04) |
| Observations | 1766 | 801 | 962 | 1008 | 755 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A13. Marginal effects - Letter sounds: % zero scores**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | -14.64** | -14.09 | -15.48** | -1.43 | -22.17** |
| | (5.73) | (10.61) | (6.95) | (10.63) | (9.51) |
| Teacher followed most/all steps | 10.68* | 15.63 | 7.96 | -3.33 | 19.66* |
| | (5.71) | (10.69) | (6.88) | (10.59) | (10.08) |
| **Student characteristics:** | | | | | |
| Age | 1.27 | 0.23 | 2.38 | 2.48 | 6.78* |
| | (1.53) | (2.21) | (2.08) | (2.34) | (3.74) |
| Books available in class | -15.13 | -54.32** | -10.31 | 14.86 | -45.43*** |
| | (17.16) | (22.52) | (17.75) | (22.91) | (16.40) |
| Books can be taken home | -3.89 | -5.04 | -4.38 | -9.69* | -3.04 |
| | (3.12) | (4.96) | (4.23) | (5.62) | (6.22) |
| Other books at home | 1.30 | -1.43 | 3.02 | -0.44 | -4.87 |
| | (2.46) | (3.53) | (3.59) | (4.94) | (5.04) |
| Other family members can read | -3.32 | -9.65 | 3.69 | -1.53 | -1.75 |
| | (4.31) | (6.42) | (6.02) | (7.72) | (6.95) |
| Teacher assigns homework | 2.64 | 7.51 | -2.29 | 2.33 | 5.84 |
| | (3.30) | (4.61) | (4.84) | (6.18) | (5.92) |
| Gets help with homework at home | -1.14 | -2.93 | 0.09 | 3.02 | -6.81* |
| | (2.51) | (3.61) | (3.67) | (4.60) | (3.91) |
| Number of meals per day | -0.69 | -1.92 | -0.13 | -0.07 | -7.72 |
| | (3.01) | (4.50) | (4.23) | (7.24) | (6.45) |
| Wealth index | 2.67 | 3.19 | 2.58 | -2.53 | 5.72 |
| | (1.93) | (3.22) | (2.54) | (6.85) | (3.98) |
| Observations | 1776 | 805 | 968 | 1017 | 756 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A14. Marginal effects - Syllable identification: % zero scores**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | 4.63 | 3.00 | 6.16 | 4.91 | -3.38 |
| | (5.19) | (7.46) | (7.27) | (10.10) | (9.66) |
| Teacher followed most/all steps | -18.56*** | -18.27** | -20.79*** | -11.93 | -19.31** |
| | (5.54) | (7.60) | (7.98) | (10.89) | (9.61) |
| **Student characteristics:** | | | | | |
| Age | 1.71 | 4.25* | -0.08 | 1.87 | 4.66 |
| | (1.68) | (2.53) | (2.45) | (3.10) | (3.67) |
| Books available in class | -3.06 | -7.24 | -5.88 | 11.01 | 5.03 |
| | (11.69) | (16.22) | (14.15) | (19.39) | (16.38) |
| Books can be taken home | -1.92 | -4.13 | -0.53 | -6.94 | -5.02 |
| | (3.56) | (5.25) | (4.84) | (6.92) | (7.68) |
| Other books at home | -5.22* | -2.33 | -8.04* | -9.34* | 0.17 |
| | (2.86) | (4.04) | (4.14) | (5.49) | (6.76) |
| Other family members can read | -0.25 | -3.12 | -0.93 | 10.60 | 1.33 |
| | (5.10) | (7.60) | (6.87) | (8.44) | (13.62) |
| Teacher assigns homework | 0.42 | 4.64 | -4.61 | -4.56 | -10.38 |
| | (4.05) | (5.91) | (5.77) | (7.00) | (9.78) |
| Gets help with homework at home | -0.72 | -2.52 | 2.85 | 1.19 | 8.36 |
| | (3.19) | (4.72) | (4.54) | (5.51) | (7.36) |
| Number of meals per day | 3.57 | 0.61 | 6.16 | -2.05 | 3.77 |
| | (3.53) | (5.40) | (4.73) | (6.97) | (8.27) |
| Wealth index | 3.32* | 2.01 | 5.30** | 0.41 | -8.31 |
| | (1.92) | (3.13) | (2.43) | (6.67) | (5.09) |
| Observations | 1777 | 806 | 968 | 1018 | 756 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A15. Marginal effects - Familiar Words: % zero scores**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | -4.90 | -3.81 | -5.42 | -13.01 | -10.59 |
| | (5.62) | (9.61) | (7.31) | (11.40) | (11.68) |
| Teacher followed most/all steps | -4.89 | -5.20 | -4.56 | 9.95 | -7.77 |
| | (5.75) | (9.62) | (7.54) | (11.86) | (11.04) |
| **Student characteristics:** | | | | | |
| Age | 0.34 | -0.32 | 0.97 | 0.31 | 3.95 |
| | (1.77) | (2.47) | (2.56) | (2.74) | (5.81) |
| Books available in class | 2.60 | -15.65 | 2.33 | 5.02 | -42.60** |
| | (12.58) | (19.80) | (14.21) | (18.52) | (17.95) |
| Books can be taken home | -0.04 | 0.55 | -1.10 | 0.39 | 2.06 |
| | (3.58) | (5.40) | (4.94) | (6.29) | (8.91) |
| Other books at home | -5.63** | -3.17 | -8.48** | 3.04 | -9.41 |
| | (2.78) | (4.26) | (3.80) | (5.36) | (7.23) |
| Other family members can read | -2.82 | -5.48 | 2.60 | 7.97 | -17.08 |
| | (4.61) | (6.64) | (6.61) | (6.89) | (14.86) |
| Teacher assigns homework | 1.89 | 6.76 | -1.75 | -1.09 | -2.60 |
| | (3.88) | (6.26) | (5.14) | (6.42) | (9.04) |
| Gets help with homework at home | 1.59 | -2.98 | 5.51 | 1.07 | 12.64* |
| | (3.10) | (4.82) | (4.28) | (5.28) | (7.55) |
| Number of meals per day | -2.74 | -4.00 | -1.60 | -9.35 | 8.89 |
| | (3.39) | (6.11) | (3.94) | (6.12) | (6.85) |
| Wealth index | 1.22 | 0.26 | 1.70 | -5.41 | 2.78 |
| | (1.73) | (2.81) | (2.36) | (5.78) | (5.33) |
| Observations | 1772 | 801 | 968 | 1015 | 754 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A16. Marginal effects - Oral reading fluency - 1 min -: % zero scores**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | -0.25 | 1.45 | -1.78 | 10.95 | 0.24 |
| | (6.06) | (10.22) | (7.64) | (11.56) | (12.10) |
| Teacher followed most/all steps | -2.02 | 1.96 | -4.70 | -11.01 | 7.44 |
| | (6.02) | (10.36) | (7.50) | (11.56) | (12.10) |
| **Student characteristics:** | | | | | |
| Age | 0.60 | -0.15 | 2.07 | 0.77 | 6.16 |
| | (2.00) | (3.03) | (2.61) | (3.70) | (4.69) |
| Books available in class | 31.38** | 39.52* | 24.42* | 57.33*** | -21.85 |
| | (15.30) | (23.83) | (14.62) | (21.02) | (19.61) |
| Books can be taken home | -1.01 | -1.71 | 0.26 | -2.05 | -0.73 |
| | (3.35) | (5.71) | (4.12) | (6.82) | (7.44) |
| Other books at home | -6.31** | -3.31 | -11.02** | -5.76 | -1.68 |
| | (3.04) | (4.34) | (4.39) | (6.87) | (6.38) |
| Other family members can read | 5.71 | 5.45 | 8.73 | 12.22 | 2.38 |
| | (5.79) | (7.96) | (8.91) | (10.02) | (16.55) |
| Teacher assigns homework | 1.20 | 0.43 | 0.32 | 2.15 | -11.19 |
| | (4.30) | (6.56) | (5.99) | (7.82) | (11.13) |
| Gets help with homework at home | -3.60 | -1.98 | -5.70 | -3.47 | -3.04 |
| | (3.25) | (4.86) | (4.55) | (6.73) | (7.61) |
| Number of meals per day | -1.75 | -2.93 | -1.26 | -1.89 | -1.34 |
| | (3.69) | (5.43) | (5.20) | (9.27) | (8.14) |
| Wealth index | 1.29 | 3.31 | -0.76 | 7.27 | -2.07 |
| | (2.10) | (3.69) | (2.66) | (7.91) | (6.07) |
| Observations | 1767 | 800 | 964 | 1011 | 753 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01

**Table A17. Marginal effects - Oral reading fluency - 2 min -: % zero scores**

| | Full Sample | Girls | Boys | Low SES | High SES |
|---|---|---|---|---|---|
| **Treatment Effect:** | | | | | |
| Treatment Effect | 7.15 | 8.57 | 3.92 | 12.96 | -2.94 |
| | (5.94) | (12.61) | (6.92) | (11.28) | (12.16) |
| Teacher followed most/all steps | -15.52** | -12.02 | -15.73** | -8.48 | -4.49 |
| | (6.04) | (13.26) | (6.81) | (11.53) | (11.51) |
| **Student characteristics:** | | | | | |
| Age | -1.21 | -1.69 | -0.45 | -1.41 | 6.47 |
| | (1.96) | (3.59) | (2.11) | (3.61) | (4.37) |
| Books available in class | 12.00 | 65.45** | 6.76 | 13.07 | -15.17 |
| | (13.75) | (32.80) | (14.79) | (25.12) | (22.09) |
| Books can be taken home | -0.34 | -0.94 | 0.27 | -10.36 | -0.66 |
| | (3.77) | (6.24) | (4.99) | (7.59) | (7.86) |
| Other books at home | -6.20** | -5.93 | -6.67 | -9.13 | -3.62 |
| | (3.01) | (4.19) | (4.50) | (6.12) | (7.62) |
| Other family members can read | 2.27 | 6.22 | 1.56 | 10.48 | 10.12 |
| | (5.61) | (8.15) | (8.00) | (9.38) | (19.78) |
| Teacher assigns homework | 7.43 | 8.17 | 7.58 | 1.98 | 3.91 |
| | (4.53) | (6.51) | (6.59) | (8.34) | (12.02) |
| Gets help with homework at home | -2.64 | -5.20 | -1.90 | -4.00 | 3.28 |
| | (3.55) | (5.36) | (4.98) | (7.01) | (7.97) |
| Number of meals per day | -2.81 | -0.73 | -4.73 | -8.07 | -10.20 |
| | (3.47) | (6.02) | (4.51) | (7.29) | (7.70) |
| Wealth index | 0.61 | -2.57 | 2.16 | -3.05 | -1.69 |
| | (2.25) | (3.96) | (2.87) | (7.02) | (5.57) |
| Observations | 1734 | 776 | 955 | 991 | 740 |

Notes: Standard errors in parentheses, clustered at the school level. All regression specifications include individual student and school fixed effects. Asterisks denote statistical significance as follows. * p<.1, ** p <.05, *** p<.01