Harnessing Youth Volunteers as Literacy Leaders (HYVALL)

Endline Student Assessment Report

November, 2014

Submitted by Education Development Center, Inc. to the YMCA
Agreement # AID-OAA-F-12-00021

Table of Contents

Table of Figures

Executive summary

Harnessing Youth Volunteers as Literacy Leaders (HYVALL) is a 2 year education program financed by the United States Agency for International Development (USAID) under the *All Children Reading Grand Challenge* and implemented by YMCA Senegal with technical assistance from Education Development Center, Inc. It aims to improve literacy rates in 6,000 grade 1 to 6 students using an out of school reading program led by community volunteers. These volunteers are each assigned a small group of students with reading difficulties to mentor and tutor twice weekly after school and during summer camps over the course of 1.5 years. The endline assessment conducted in November 2014 allows for an evaluation of the program results with respect to student reading performance in comparison to the baseline assessment undertaken in March 2013. These assessments were conducted in randomly selected "intervention" schools and control schools in the regions of Dakar, Thies, Kaolack, St. Louis and Ziguinchor where HYVALL intervenes. It aims to assess whether or not and how much students in the HYVALL intervention groups may have progressed since the intervention began in April 2013. The study evaluated student reading fluency and comprehension of a text adapted to the student's level (independent or instructional-level text). The study is designed as a longitudinal study: students who participated in the baseline assessment were tracked wherever possible for this endline assessment. The assessment was implemented by Senegalese YMCA staff trained by EDC assessment advisor. This report summarizes the results of the endline assessment and compares these results against the results of the baseline study.

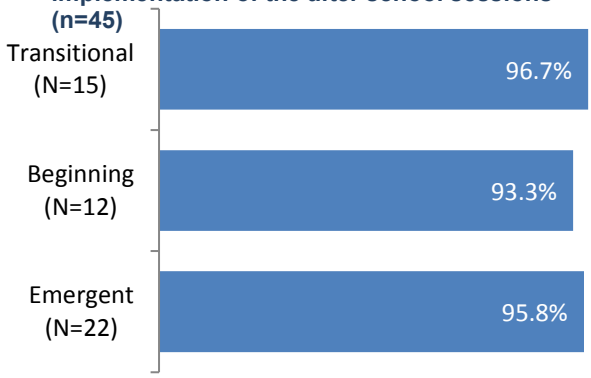
Summary results show that students who participated in the HYVALL intervention exhibited larger gains in their fluency and reading comprehension from baseline to endline, compared to their control group counterparts ($p<.001$). These results suggest that an intervention like HYVALL that gives students opportunities for reading instruction outside school setting (increased time on task), provides regular one-on-one tutoring and mentoring and encourages parental involvement in the reading development of their children can lead to dramatic student improvement in reading.

Overview of the HYVALL intervention

To better understand HYVALL intervention results, data on the fidelity of implementation as well contextual information was collected throughout the project implementation. This section of the report presents a summary of the fidelity of implementation data including volunteer application of the program, student attendance to activities and their engagement with home literacy activities, and their opinions on reading. Finally, data from interviews with parents of intervention students is also presented. This data provides a context from which to interpret the student reading results.

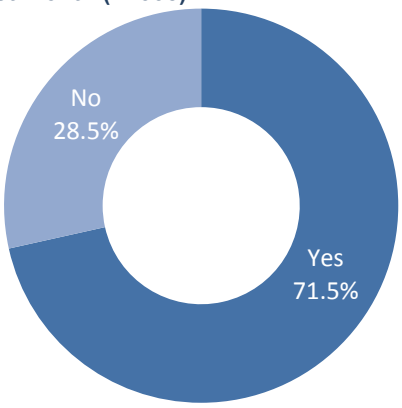
Volunteers' fidelity of implementation. The design of the HYVALL intervention called for "at-risk" students to be tutored twice weekly by a local volunteer in reading and writing for a period of 1.5 years. Volunteers were provided with leveled reading materials, lesson plan templates for each level and scripted activity guides. Depending on the reading levels of their students (emergent, beginning, or transitional), volunteers selected activities listed in their lesson plan templates to carry out in their 45-minute sessions. Volunteers received a four day training in planning for and applying these activities. Part of this training involved a shadowing experience where they would accompany the center supervisor in a tutoring session. Thereafter, volunteers were assigned a group of students to tutor. The median ratio of volunteer to student was 1 to 10 ranging from a minimum of 8 students to a maximum of 20 per one volunteer[1]. In order to group students by level, short screening placement assessments were individually administered consisting of letter reading, high frequency word reading, and passage reading. Students were placed in a given reading group according to their scores on this screening. Once these groups were determined, volunteers were assigned to groups of students. In cases where many students were placed into the same level, centers would assign two volunteers to the same level. For example, some centers had two beginning reading groups, one medium group but no advanced group.

Figure 1. Literacy volunteers' mean fidelity of implementation of the after school sessions (n=45)



Given this design, how did HYVALL volunteers apply these lessons plans? Using an observation tool that was based on the lesson plan templates, center supervisors conducted observations twice a quarter to track fidelity of implementation as well as provide feedback to those volunteers with regard to the implementation issues. Figure 1 shows that average fidelity of implementation of the reading activities was very high. These results show that the program was implemented as designed: volunteers followed the session plans as modeled in their training and as stated in their reading activity guide.

Figure 2. Volunteer home visits in the past month (n=358)



Volunteer home visits. As part of its design, HYVALL also aimed to promote a culture of reading and to encourage home environments to be more conducive to students' literacy development. To stimulate this, volunteers were asked to visit their tutees' homes once a quarter. At the endline assessment, students were asked to report whether or not their YMCA volunteer had visited their home in the last month. The large majority of students

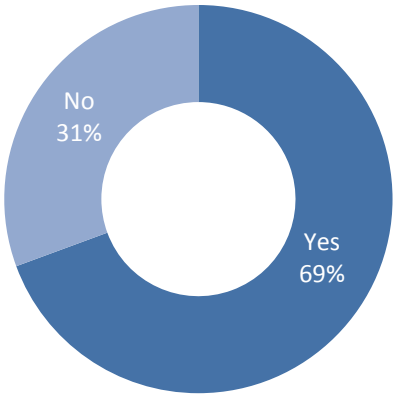
[1] Number of students per volunteer was determined by student reading levels. Because students were divided into leveled reading groups according to their ability levels, in cases where the majority of students were at a certain level, volunteers would sometimes have to take into their group more students than was originally planned.

(71.5%) reported that a YMCA volunteer had visited their home to talk about literacy with his or her parents Figure 2).

The intention of these visits was to talk to parents about the importance of reading, how their child was progressing in reading, and how they as a family could help support their child in his or her literacy development.

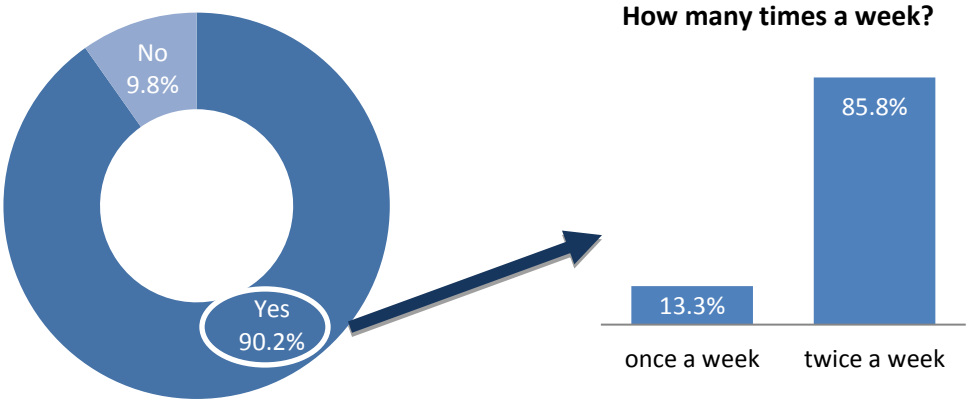
Student attendance in project activities. Student participation in the HYVALL intervention is another essential element of program implementation. The two main interventions offered to students through HYVALL were two 45-minute reading sessions per week for 1.5 years and a month-long summer camp offered during the 2014 summer vacation. For the duration of these summer camps, 45-minute reading sessions were held every day of the week for one month. For struggling readers, long vacations or breaks are shown to be detrimental to their progress (Entwisle et al., 2000). The goal of these camps was to ensure that students would start school not having lost what they had attained over the past school year and to give them an extra boost of confidence in reading prior to re-entering school. At the endline assessment, tested students were asked whether or not they had participated in these camps. As Figure 3 shows, 69% of students reported having attended the summer camp.

Figure 3. Student attendance at reading summer camps (n = 358)



At endline, students were also asked whether or not they had participated in the weekly after school reading session, and if so, how frequently they participated. The figures below show that 90% of sampled students reported regularly attending the after-school reading sessions. From those who participated, 86% of students reported attending twice weekly, as intended by the intervention.

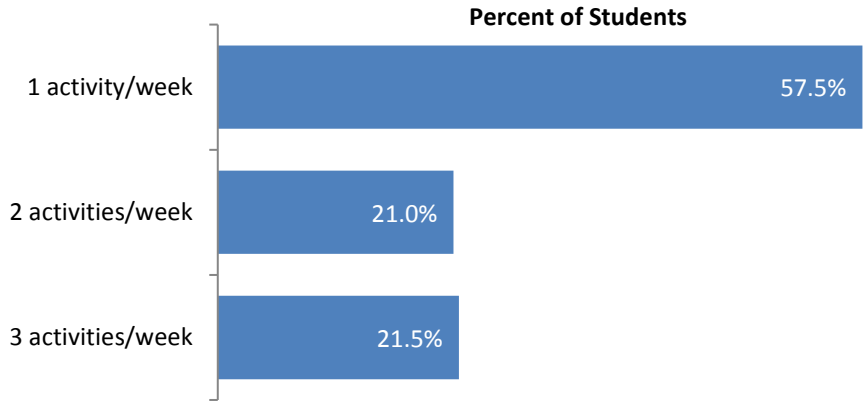
Figure 4. Student attendance at afterschool reading session: do you attend the after school reading sessions?



These participation rates were confirmed by students' weekly submission of the "Reading Passport" tool. This tool was given to each student participating in sessions. The Reading Passport traveled between the student's home and the center once a week. This Reading Passport was intended for students to record when they engaged in literacy activities outside of the center. The Reading Passport was also presented to parents by volunteers during home visits as examples of the types of activities they could encourage their children to do at home as literacy practice.

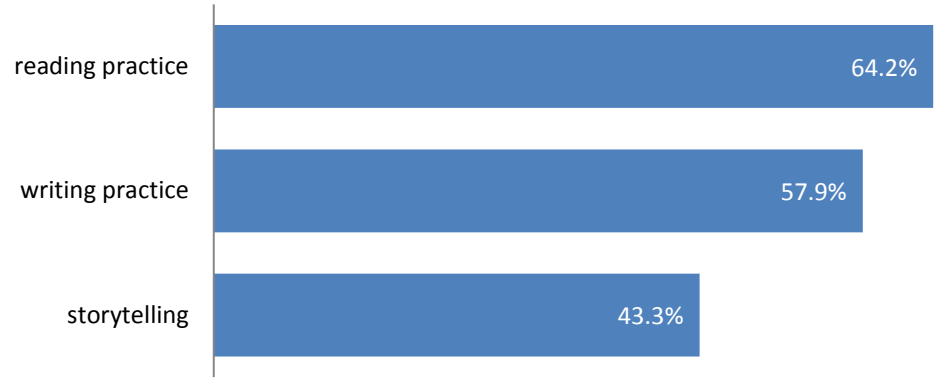
From these Reading Passports, volunteers recorded their students' home activities weekly in a summary sheet. Each week, students could have checked a total of three boxes if they had engaged in reading practice, writing practice, or in storytelling. For the purpose of the study, data for a sample of intervention students was entered for the last 7 weeks of the HYVALL intervention. Over the period of 7 weeks, students on average performed 1.65 home literacy activities. The figure below shows percent of students who performed one, two or all three activities on a weekly basis. More than half of students performed one out of the three activities. Less than a quarter of students (21.5%) completed all three activities every week.

Figure 5. Number of Reading Activities Students Perform Outside of the Center, per week (n=214)



Out of the three activities students could do at home, the most popular reported activity was reading practice (64.2%) followed by writing practice (57.9%) and then storytelling (43.3%). This is shown in the figure below.

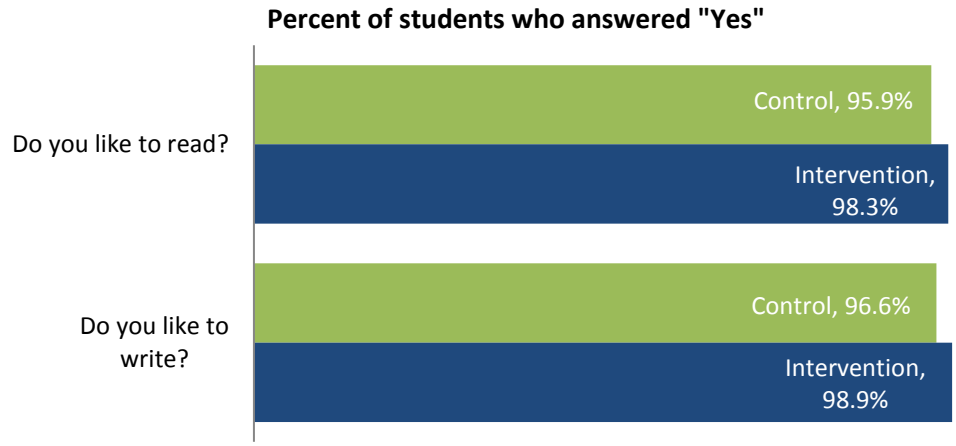
Figure 6. Student's Home Literacy Activities (n=214, multiple response)



These results are important as research shows home literacy practice to be just as important for reading development as school-based reading (Save the Children, 2013). This is confirmed by this study which shows that when pairing student reading scores with their application of home activities, 10.6% of the variation in students' reading comprehension scores could be explained by their engagement in home-based literacy activities (p=.000, d=2.2, ES=.074). It is therefore suggested that there is significant added value in the home-literacy component of this program.

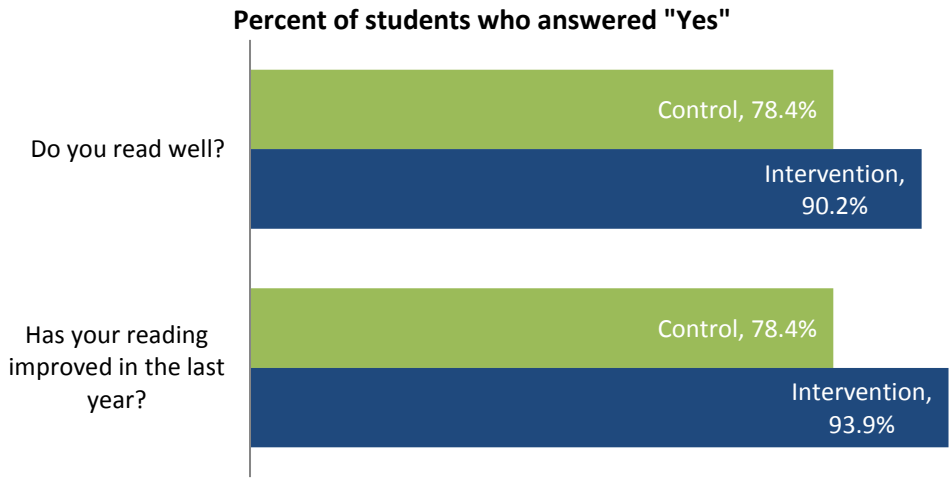
Student opinions of their literacy. At the endline assessment, students from both the intervention and control groups were also asked about their opinions on literacy. First, students were asked if they liked to read and to write. Answers to these questions did not differ significantly between intervention and control students.

Figure 7. Students Like to Read and Write (n=628)



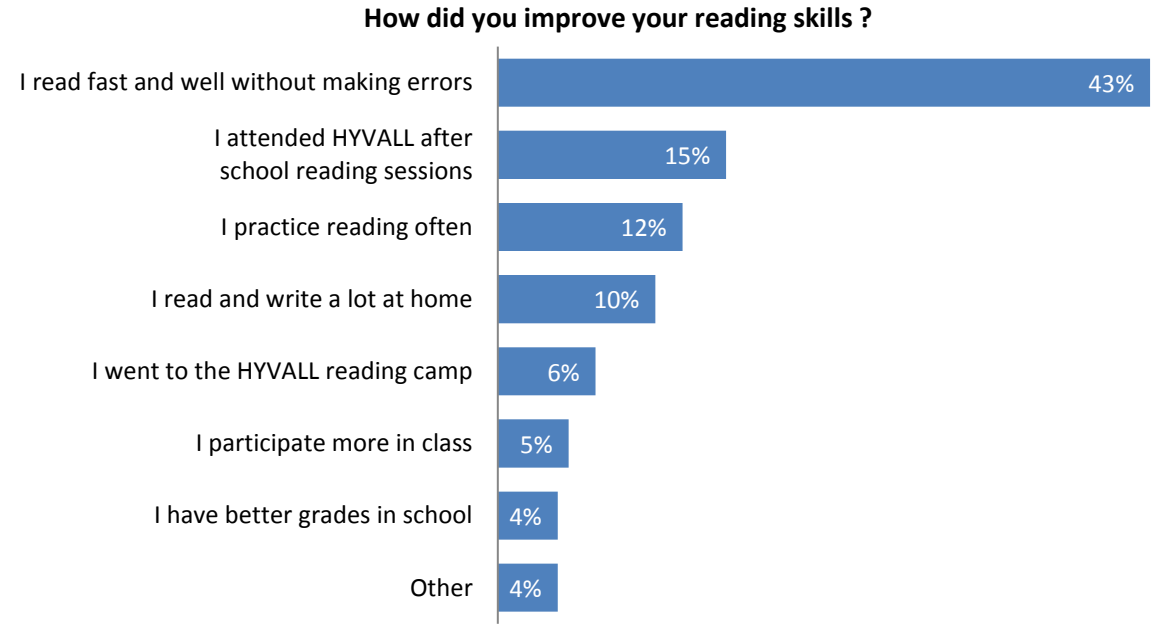
Students were then asked to think about whether or not they thought they were good readers. Significantly more intervention students indicated that they thought of themselves as good readers in comparison to their control counterparts. A larger percentage of intervention students (94%) also stated that they had improved in their reading over the last year.

Figure 8. Students Opinions of their own Reading (n=626)



How did students think they became better readers? Those intervention students who answered they had improved gave several answers to this question. Just over 40% noticed they had simply gotten to be more accurate and faster readers. Around 15% said they were better readers because they attended the HYVALL after school reading sessions. Around 13% said they improved because they practiced reading often, and 11% said they practiced at home. These responses suggest that HYVALL had an impact on students' opinions of their reading through encouraging reading and writing practice outside of school.

Figure 9. Intervention Students' Explanations for Why Their Reading Skills Improved (n=198)



Parents' opinions of their children's literacy. During one of the volunteer's home visits, parents of HYVALL students were interviewed to surmise whether or not the community was literate, valued literacy, and to gauge what role they thought the community could play in their child's literacy

development. To gauge literacy rates in the family, parents were asked how many people in their household could read. Of the 114 parents sampled, 50% reported that three or more people in their household could read, while less than 3% reported that no one in the household could read.

Figure 10. Distribution of number of literate household members (n=114)

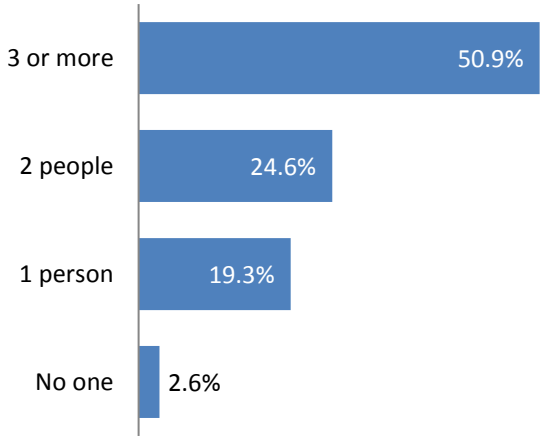
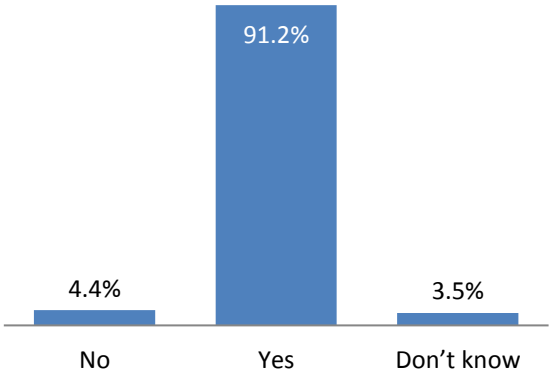


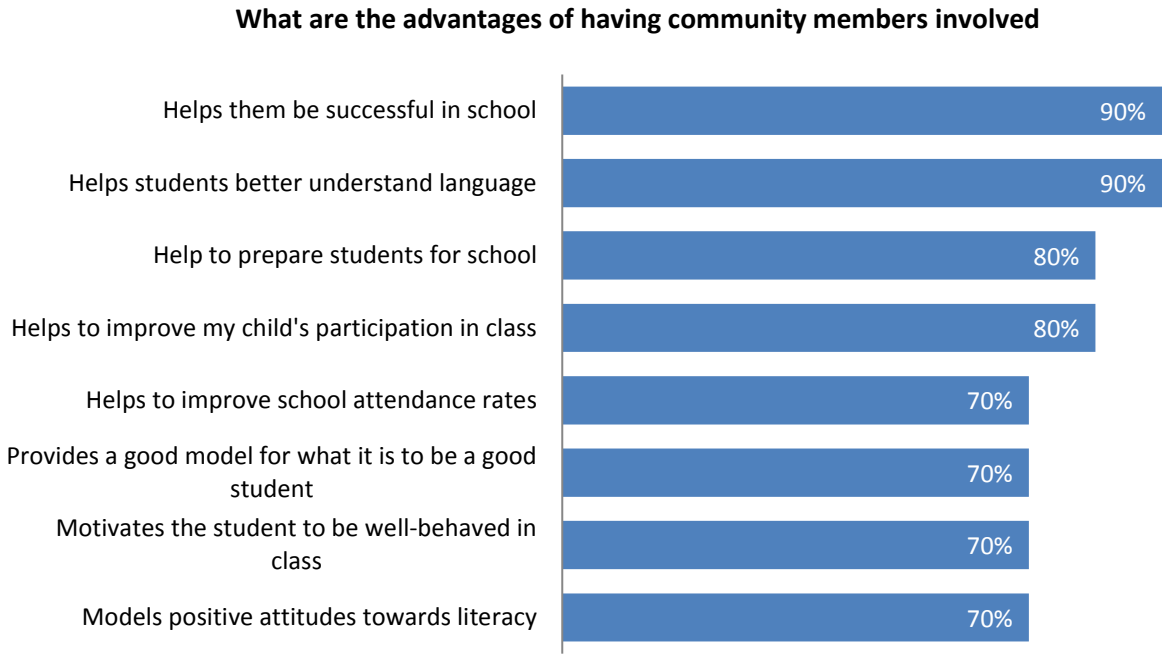
Figure 11. Do you think members of your community can help your child become a better reader? (n=114)



Overall over 90% of parents responded that the community could help a child become a better reader. This response suggests a positive reflection of HYVALL's community component.

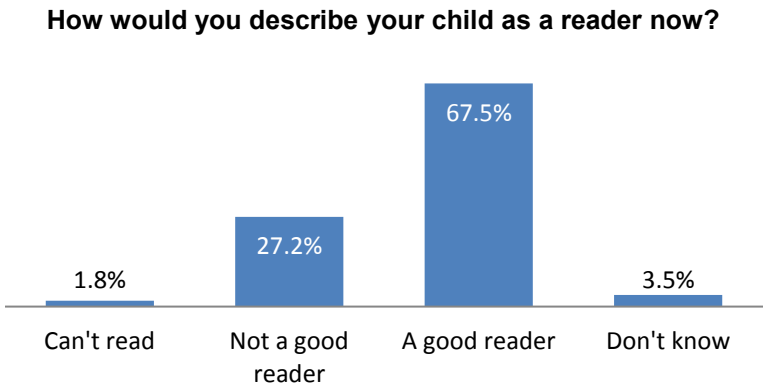
When parents were asked to elaborate on the advantages of having community members involved in their child's literacy development, parents had several responses. The figure below recapitulates parents' explanations. In the context of HYVALL, a community volunteer-led intervention, these responses imply that parents recognize how community members can contribute to helping their children be successful in school, with language, as well as serve as models for positive attitudes towards literacy.

Figure 12. Parent's Opinion on Advantages of Community Involvement in Their Child's Literacy Development (n=114)



Finally, parents were asked to gauge their child's reading level. Over 65% of parents stated their children were good readers while just fewer than 30% claimed their children were not good readers.

Figure 13. Parents Opinion of their Child's Reading Level (n=114)



Interestingly, the number of reported literate family members in the household correlated significantly with parents' responses to this question, with 10% of the variation in parents' responses being explained by the number of literate family members (p=.001, d= 0.66, ES=0.31). That is to say that the fewer literate people are in the household, the less parents thought their children were good readers. Though students from these less literate households may have indeed started out in the intervention as "poor readers", the data below shows that they dramatically improved in their reading skills. In the context of HYVALL which selected "at-risk" students for reading, this is important to bear in mind when looking at the student reading results.

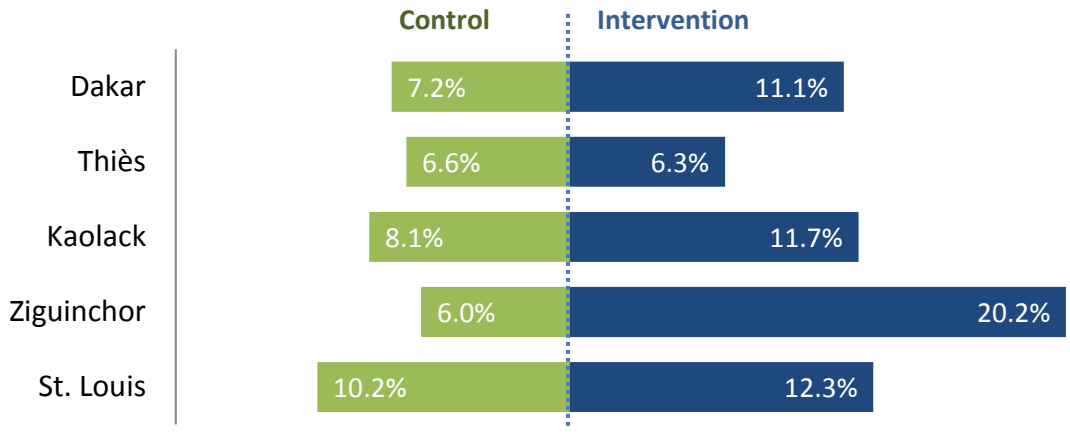
Overview of the Endline Assessment Instrument

The instrument administered at baseline and endline was designed to measure students' level of oral reading fluency (reading accuracy and speed) and reading comprehension at their "independent" or "instructional" level. This level is identified as the test is administered. If a student cannot read the first grade-level text presented to them with 75% accuracy, he or she will be asked to read a lower level text. This exercise is undertaken at the first reading of the text which is limited to 60 seconds. After the first reading of the "independent-level" or "instructional level" text is completed, the student is asked a series of comprehension questions on the content of the text he/she managed to read without text support. Then, the same text is read by the student a second time. The second reading is timed but is not limited to a specific time frame. Then the text is left in front of the student and questions are asked about the content of the entire text adding an inferential question which requires the student to use the information from the text to provide a "plausible" answer. Analysis of each instrument can be found in Appendix A.

Sample Demographics

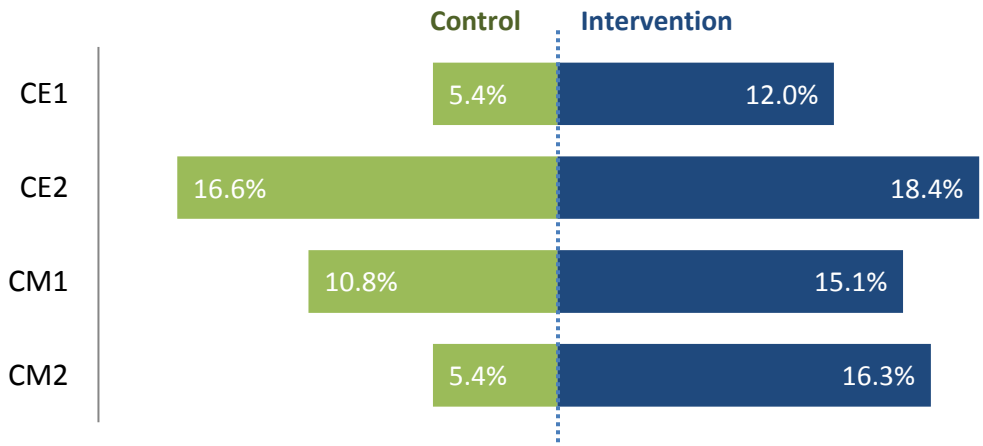
The figures below show the distribution of the sampled students by region, level, sex and language(s) spoken. The percentage of students in the sample varies greatly by region. Ziguinchor is most represented but seems to have experienced a dramatic change in study participants between baseline and endline for the control group. Thies is the least represented but maintains equal distribution across groups. This is explained by the significant initial selection of students in CM2 (6th grade) at baseline who at the endline assessment have moved onto secondary school.

Figure 14. Regional Distribution of Students Tested (n=332)



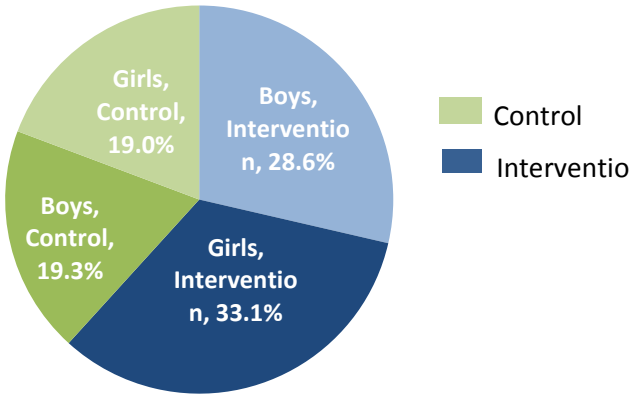
The percentage of students per level also varies demonstrating a strong student participation in CE2 (grade 3), in comparison with other levels. However, because the assessment allows the student to step down to a lower-level text it should be noted that over 10% of intervention CE students and over 50% of control CE students who started with the CE-level text actually ended up taking the CP-level text.

Figure 15. Distribution of Students by Level (n=332)



The distribution by sex remains fairly equal between intervention and control groups. The results presented in this report are disaggregated by sex.

Figure 16. Distribution of Students, by Sex



The distribution by sex and grade level also remains equal between intervention and control groups.

Table 2. Distribution by Level and Sex

Grade level		Girls	Boys
CE1 – 3rd grade	*Number*	31	27
	% of grade	53.5%	46.5%
CE2 – 4th grade	*Number*	57	59
	% of grade	49.2%	50.8%
CM1 – 5th grade	*Number*	47	39
	% of grade	54.7%	45.3%
CM2 – 6th grade	*Number*	38	34
	% of grade	52.8%	47.2%

It is also important to be aware of the language(s) spoken by the students. The language in which one is versed in can play a substantial role in student's learning to read and comprehend a text. At endline, almost all of the study students reported speaking Wolof and French, and 100% reported speaking another language (mother tongue), as well. Due to the lack of variation in languages spoken, it is not possible to identify potential influences of language on students' gains in reading.

Figure 17. Distribution of Students Tested by Language Spoken (n=332)

NB. Many students speak several languages.

Distribution of independent texts read by students, by grade-level

The assessment was designed to capture students' oral reading fluency levels when reading a text at their "independent" or "instructional" level. For example, a tested student in CE2 was first given the CE text to read. If the student could not read more than 75% of the words correctly, he or she was asked to read the CP-level text to give him or her the opportunity to demonstrate his or her real reading fluency capabilities. This is important because assessing a student on a text that is too difficult limits the student's ability to demonstrate his or her true reading skills. Given this evaluation design, it is imperative to consider not only the education level of the student but also their established level of independent reading as it pertains to this assessment. For the intervention group, almost 15% of CE1 students and 5% of CE2 students were asked to read the CP-level text because the CE text was too difficult for them. For the control group students, almost 30% of students in CE1 and 28% of CE2 students had to go down one level. This shows that a larger proportion of students in CE1 and CE2 in the intervention group could read a grade-level text in comparison to their control counterparts.

The majority of CM1 and CM2 intervention students managed to read the CM text but about 15% of students finished by reading the CE text and 2% read the CP text. A higher proportion of CM1 and CM2 control students could not read on grade level than their intervention counterparts: 50% of CM2 students were asked to drop down to a CE-level text and 18% of CM1 students were asked to read a CP-level text.

Figure 18. Distribution of Texts Read By Intervention (HYVALL) Students, By Level at Endline (N=205)

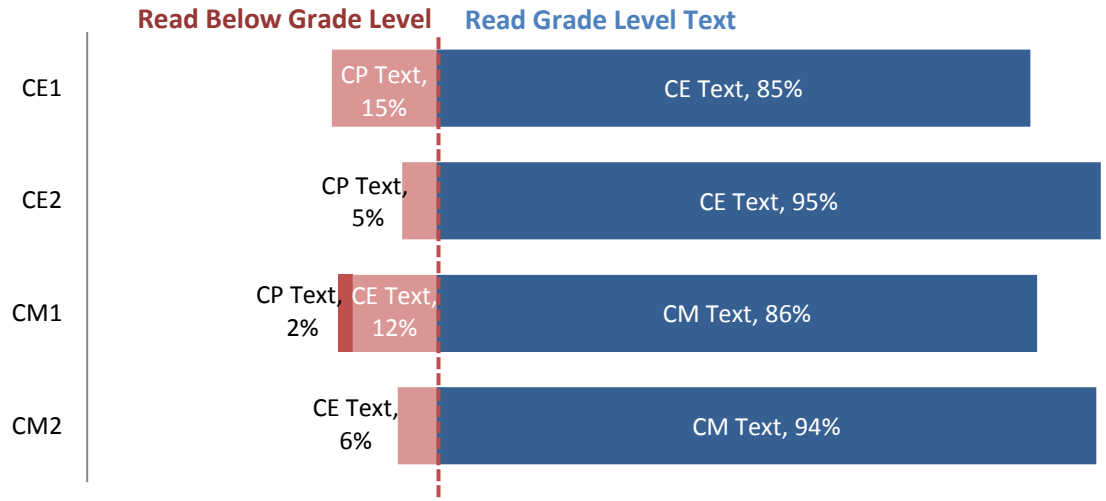
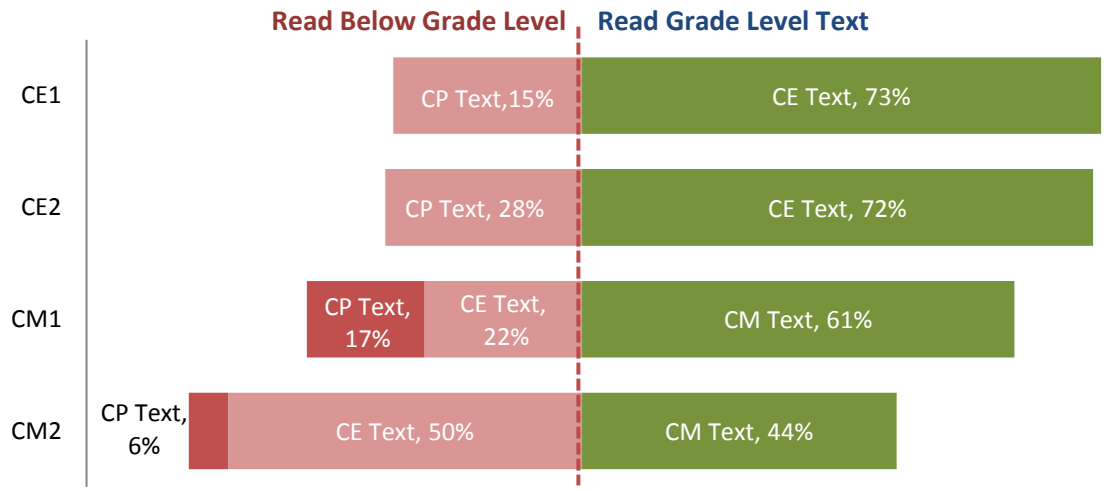


Figure 19. Distribution of Texts Read By Control Students, By Level at Endline (N=123)



This distribution of levels of text read is important to keep in mind when looking at the results presented below because oral reading fluency and comprehension results are judged at students' determined instructional or independent levels.

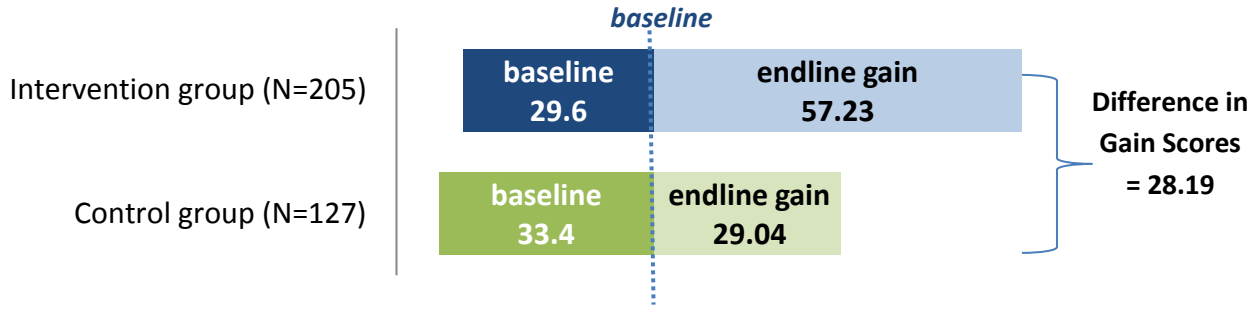
Student Results in Reading

Following the above context of the HYVALL intervention, this section presents the reading results derived from the endline assessment conducted in November 2014 at the conclusion of the HYVALL project. The results demonstrate gains made since the baseline in March 2013 for those students who were retained from baseline to endline.

Summary Reading Results

Fluency. As shown in the figure below, students' overall fluency results for both control and intervention groups increased significantly (p=.000) between baseline and endline. Intervention students were found to read significantly faster and with greater accurately at endline in comparison to their control counterparts (p=.000, d=0.85, ES=0.4). Accounting for the high participation rates in intervention activities and similar starting level, this difference can be attributed to the HYVALL activities.

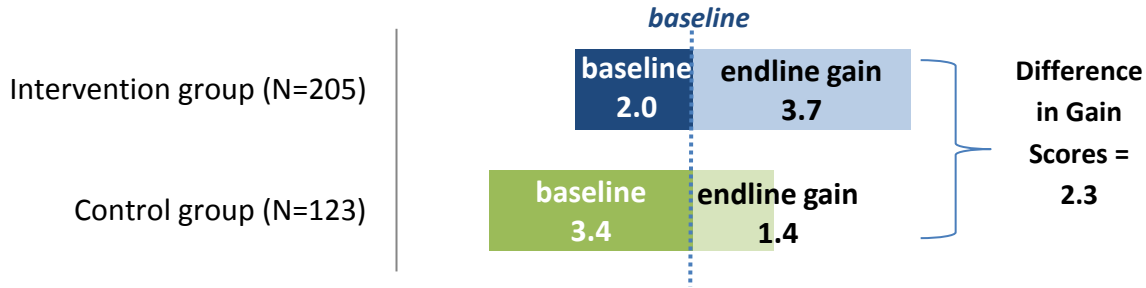
Figure 20 Overall of gains in Words Correct Per Minute (WCPM) across baseline and endline



Reading Comprehension. Following the first reading of the passage, students were asked five comprehension questions. Overall, students reading comprehension after the first reading for both control and intervention groups significantly increased over baseline and endline. However, differences gain scores significantly favored the intervention group (p=.000, d=0.84, ES=0.39). For the six comprehension questions asked after the second reading of the passage, students' average reading comprehension scores for both control and intervention groups increased significantly (p=.000) between baseline and endline. As shown in the figure below, intervention students, on average, were able to answer 3.67 more questions correctly (out of six) than they could at baseline. This gain for intervention students showed to be significantly greater than their control counterparts endline (p=.000, d=1.01, ES=0.45). This suggests that students who participated in 1.5 years of the intervention developed reading comprehension skills. In addition, reading comprehension gains for intervention students were found to be significantly correlated with their participation in home reading activities. This is confirmed by this study which shows that when pairing student reading scores with their application of home activities, 10.6% of the variation in students' reading comprehension scores could be explained by their

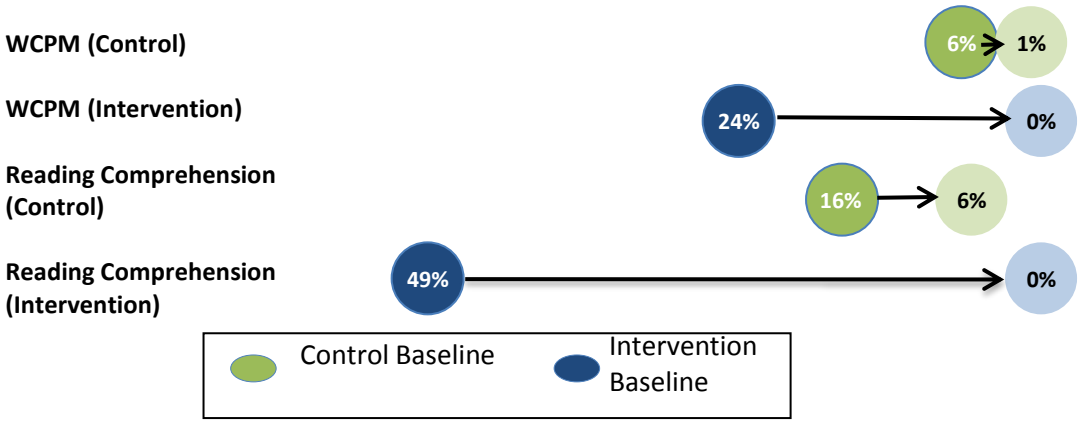
engagement in home-based literacy activities (p=.000, d=2.2, ES=.074). It is therefore suggested that there is significant added value in the home-literacy component of this program.

Figure 21. Overall Summary of Gains in Reading Comprehension (out of 6) Across Baseline and Endline



Zero scores. At baseline, a number of students had "zero" scores in oral reading fluency and / or reading comprehension. A "zero" in oral reading fluency means a student could not read a single word correctly or did not make an effort to read at all. A "zero" score in reading comprehension is defined by a student not answering any questions correctly, a student who did not attempt to respond or a student who failed to read the text and, consequently, was not asked any comprehension questions on this text. It is important to understand how the zero scores in oral reading fluency and reading comprehension between the baseline and midline may have changed. The figure below shows changed in “zero” scores between baseline and endline.

Figure 22. Change in Zero Scores in Fluency and Comprehension, Between Baseline and Endline



Assessment results showed that the proportion of students with zero scores decreased in both intervention and control groups. At baseline, about 5% of students from the control group could not read a single word whereas 24% of the intervention group students could not read a word. However, at endline, no intervention students had zero scores whereas 1% of control students still had zero scores. For reading comprehension, zero scores for intervention students fell from 49% to 0%. This is a notable improvement for intervention students as every student showed to be able to answer at least 1

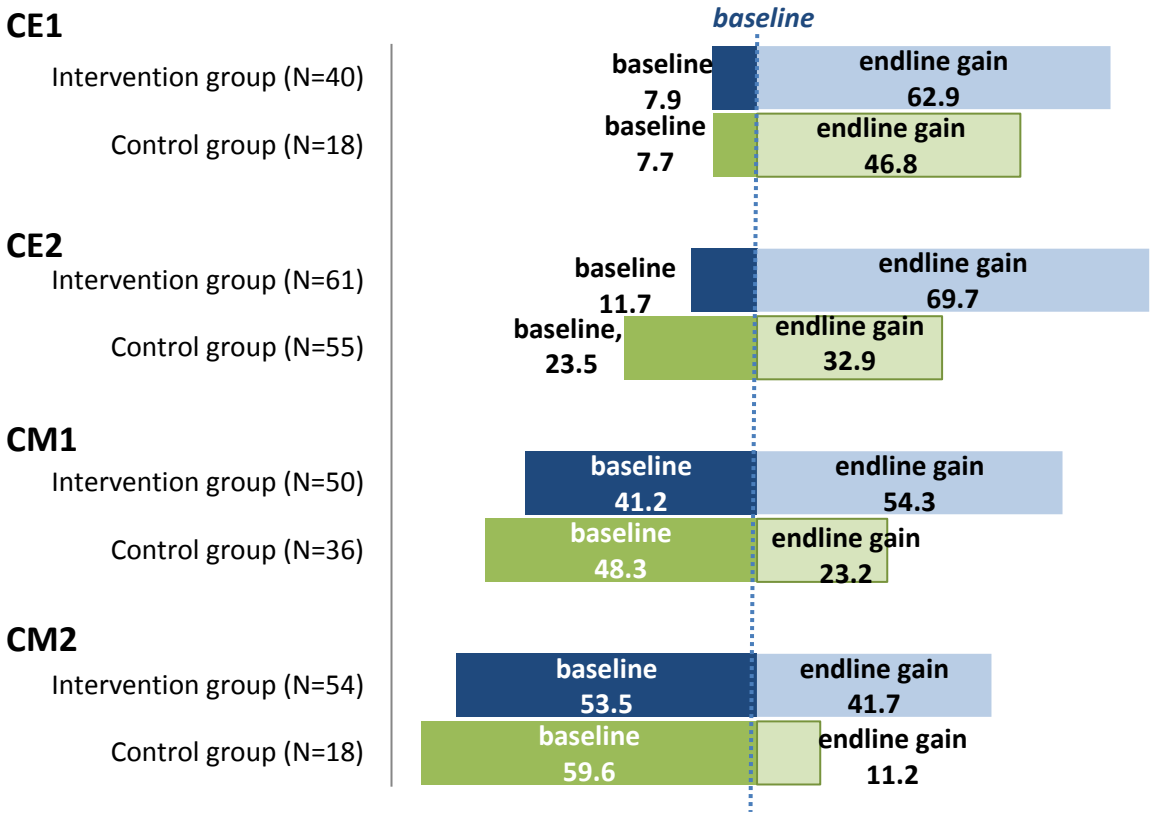
comprehension question correctly. Zero scores in reading comprehension for control students decreased by 10% between baseline and endline.

Oral reading fluency results

The oral reading fluency portion of the test was divided into two segments. First, students were asked to read a passage at their grade-level. For example, students in CE2 were asked to read a 56 word text in French for a given period of 60 seconds. If the student could not manage to read 75% of the text correctly in that time, the student was then given a test at a grade-level lower (ex. CP) to read in another period of 60 seconds. This section begins by presenting the results of students' performance on the timed fluency portion of the assessment across baseline and endline.

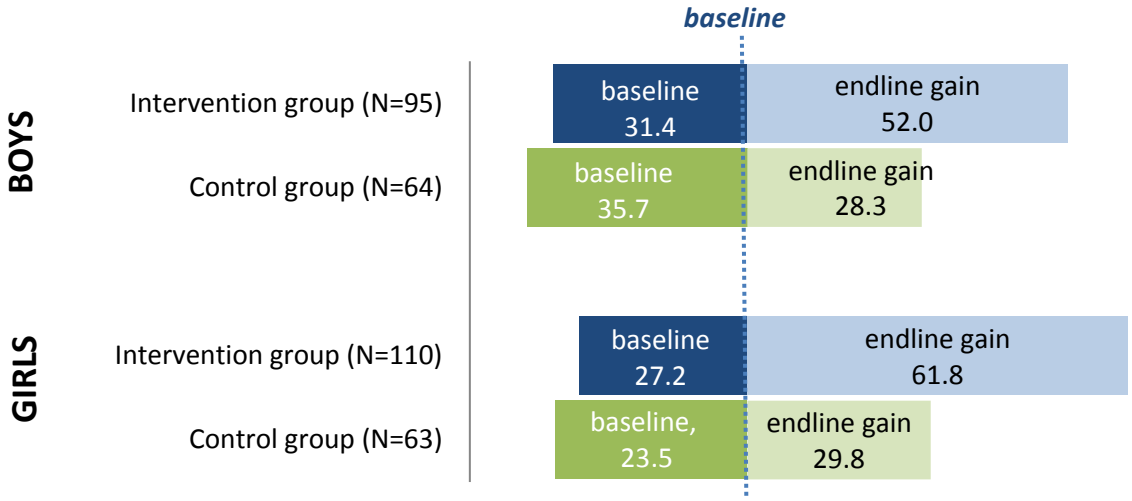
Oral reading fluency results by grade level. Looking at the results disaggregated by grade, student gains in words correct per minute (WCPM) were found to be more dramatic in CE1 and CE2 than in CM. Intervention students in these grades gained more significantly as compared to their control counterparts. The students in these grades started with the intervention program at the CI and CP levels. This suggests that intervening earlier can have a more substantial impact on reading improvement than for those students who entered the program in later grades. It is also interesting to note that students in CE1 from control and intervention groups started with similar mean scores whereas in the other grade levels, control students showed to start out as better readers at baseline. For these grade levels, however, students from the intervention group not only gained in their reading fluency but significantly surpassed their control counterparts at endline.

Figure 23. Summary of Gains in Fluency Performance (WCPM) Across Baseline and Endline by Grade Level



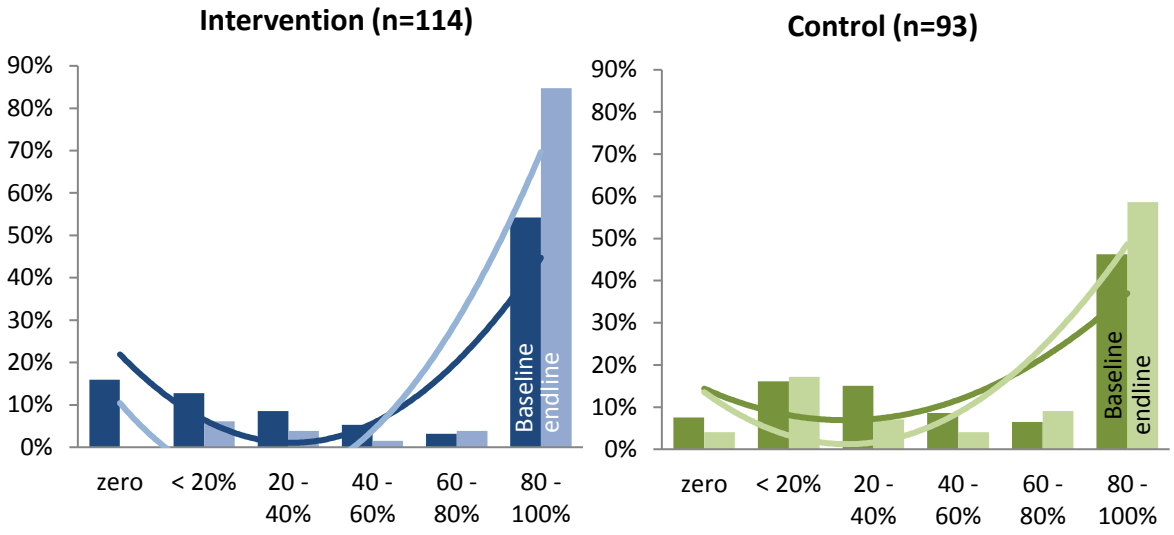
Oral reading fluency results by sex. Girls were found to make greater gains in terms of accuracy and reading speed across baseline and endline for both intervention and control groups. Girls in both intervention and control groups also started out with lower mean fluency scores at baseline in comparison to boys. However, at endline, the intervention girls showed to gain more than their male counterparts but also surpassed boys' reading fluency rates, on average at endline. On the other hand, control girls gained more than the control boys. Still, they showed to have lower mean fluency rates at endline than the boys. Despite this, the differences in gains between girls and boys for both intervention and control groups are not statistically significant..

Figure 24. Summary of Gains in Fluency Performance (WCPM) Across Baseline and Endline by Sex



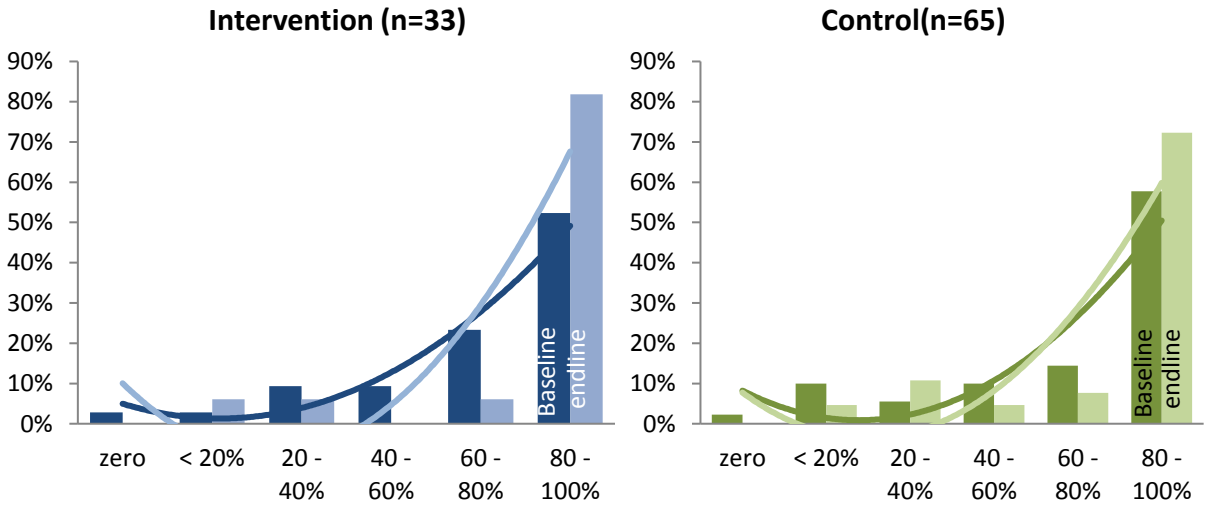
Oral reading fluency results by test and grade level. In addition to disaggregating by sex and grade level, it is also interesting to look at the distribution of students' abilities to read a grade-level text. This allows us to gauge how students fare when judging them by a grade benchmark. The results below show the distribution of accuracy in reading a grade-level for those students who read a text at their grade level at both baseline and endline. Because no CP-level students were present at the endline, results are only presented for CE and CM levels. The first two figures show that approximately 50% of CE students from both intervention and control groups couldn't read between 80 and 100% of the words in a grade-level text correctly with approximately one third of students unable to read more than 40% of a grade-level text correctly. At endline, however, over 80% of CE intervention students were able to read 80 to 100% of words correctly in a grade-level text. A higher percentage of control CE students at endline were able to read 80 to 100% of the grade-level text passage correctly but a third of students could only read up to 40% of the text passage correctly.

Figure 25. Percentage of WCPM for CE test



At baseline, CM students' accuracy in reading a grade-level text was largely around 60 to 100% of words read correctly for both groups. At endline, the trend line moved further to the right with over 80% of CM intervention students reading between 80 to 100% of the words correctly. For CM control students, this trend is similar though a smaller proportion of control students could read between 80 and 100% of a grade-level text at endline.

Figure 26. Percentage of WCPM for CM test



Overall, the above distributions reveal that the grade-level passages may have been too easy for the average CE or CM student. If the texts had been moderately more difficult, normal bell-shaped curves would have likely emerged. Still, these results provide a picture of how intervention students became better equipped to read a text at their grade-level as a result of the intervention.

Reading Comprehension Results

This section presents the distribution of reading comprehension at endline and the results disaggregated by grade level and sex across baseline and endline. During the assessment, comprehension questions were posed on two occasions: once after the timed first reading cut off at 60 seconds, and the second time after a timed but uncapped second reading of the same text. For questions after the first reading, the student did not have the text in front of them to refer to while they answered questions, which required drawing from memory to respond. In addition, the student only had to answer the questions associated with the portion of the text he or she read. For example, if the student was unable to read more than 3 lines in 60 seconds, only the questions associated with these three lines were asked. For questions after the second reading, the student retained the text in front of them and was free to refer to it to answer questions. For the second round of questions, an implicit (inferential) question was added to determine whether the student could use the explicit information in the text to respond to the question.

Distribution of reading comprehension results by group. For both groups at endline, students showed to be able to answer more comprehension questions correctly at the second reading when they were given the opportunity to read the full text in an unlimited amount of time and were able to refer back to it to answer questions. This is clearly shown in the figures below. These figures also demonstrate the differences in distribution of results in reading comprehension for both intervention and control groups at endline. The differences in mean gain scores between intervention and control groups in reading comprehension after the first reading are statistically significant, favoring the intervention group (p=.000, d=0.68, ES=0.32). This same trend was found in the differences in mean gain scores in reading comprehension after the second reading (p=.000, d=1.12, ES=0.49). For the intervention group, the distribution of reading comprehension questions answered correctly skews to the right with over 50% of students able to answer 5 out of 5 comprehension questions after the first reading and over 70% of students able to answer 6 out of 6 comprehension questions after the second reading. For control students, the distribution of questions answered correctly after both the first and second readings is more evenly distributed with around 30% of students able to answer 5 out of 5 comprehension questions after the first reading and just under 50% of students able to answer 6 out of 6 comprehension questions after the second reading. Overall, these distributions illustrate how intervention students had an easier time answering comprehension questions whether from memory or using a text support in front of them. This is not surprising given HYVALL reading sessions focused heavily on developing students' language skills, vocabulary, and comprehension strategies of stories read.

Figure 27. Number of Comprehension Questions Answered Correctly After the First and Second Reading for the Intervention Group

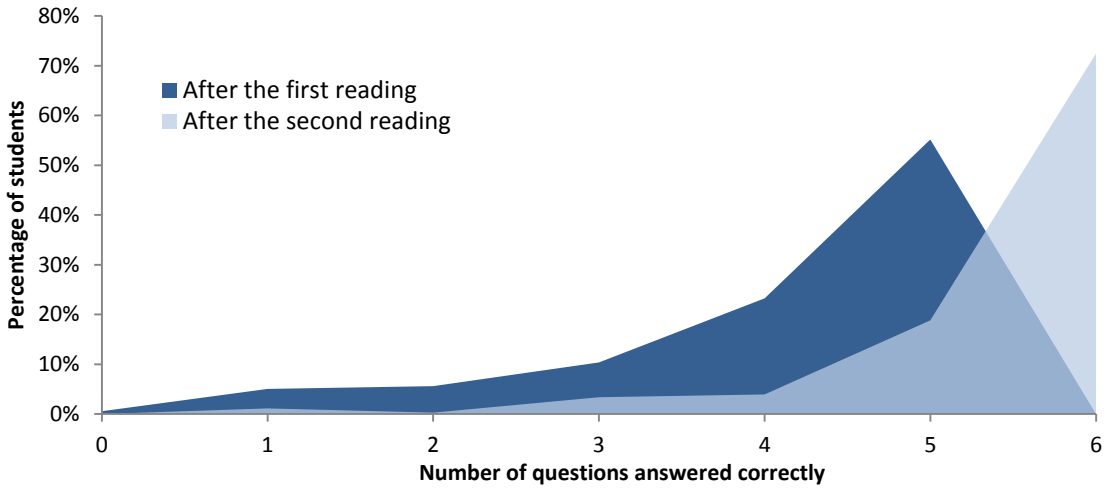
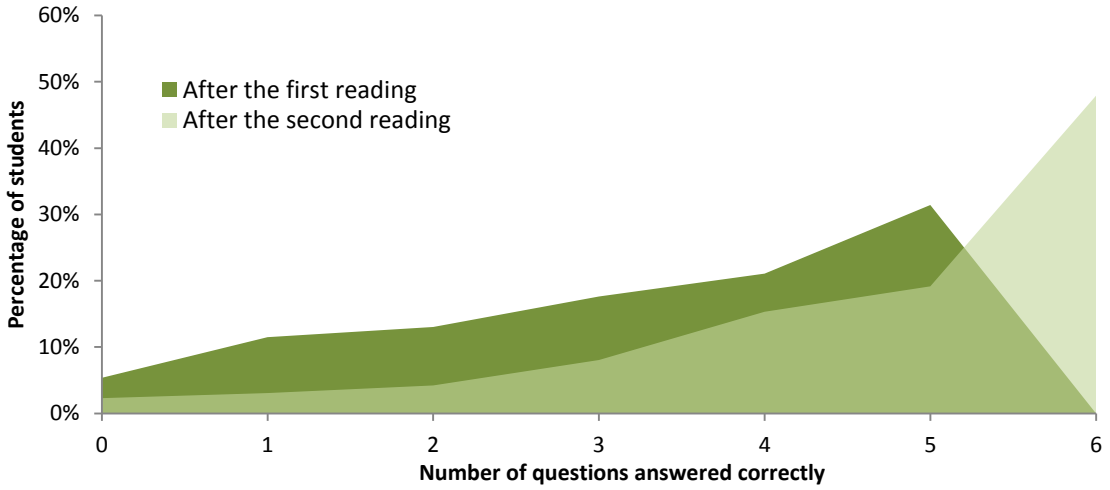


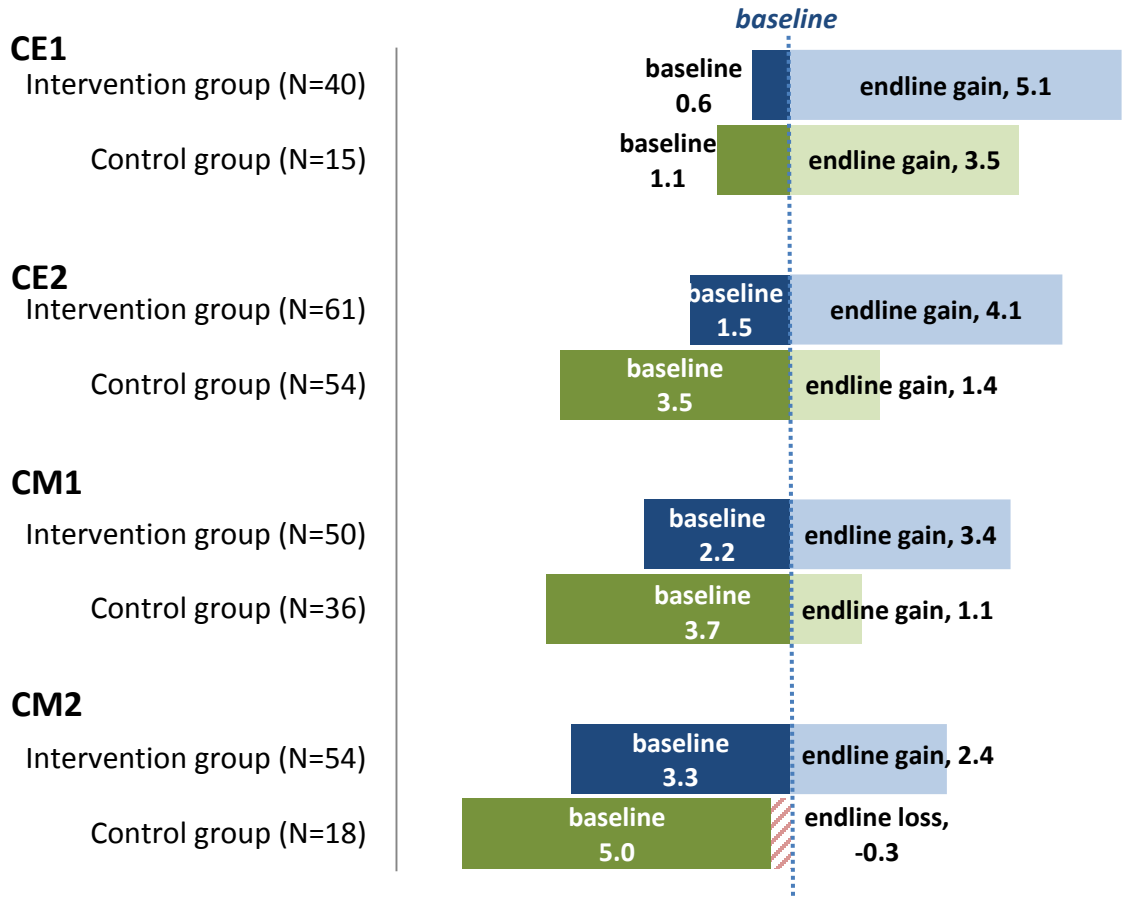
Figure 28. Number of Comprehension Questions Answered Correctly After the First and Second Reading for the Control group



Reading comprehension results by grade level. Looking across baseline and endline and disaggregating by grade-level, it is interesting to note that, like in reading fluency, the largest gains in reading comprehension (at the second reading) are also at the CE1 and CE2 grade levels for both groups. However, the differences in gains for these grade levels are statistically significant, favoring the intervention group (for CE1: p=.001, d=0.96, ES=0.43, for CE2: p=.000, d=1.4, ES=0.57). This suggests that HYVALL not only served to improve reading fluency for the early learners but also was simultaneously successful in building their reading comprehension. Gains were also significant for intervention and control students in the older grades except for CM2 control students who showed no

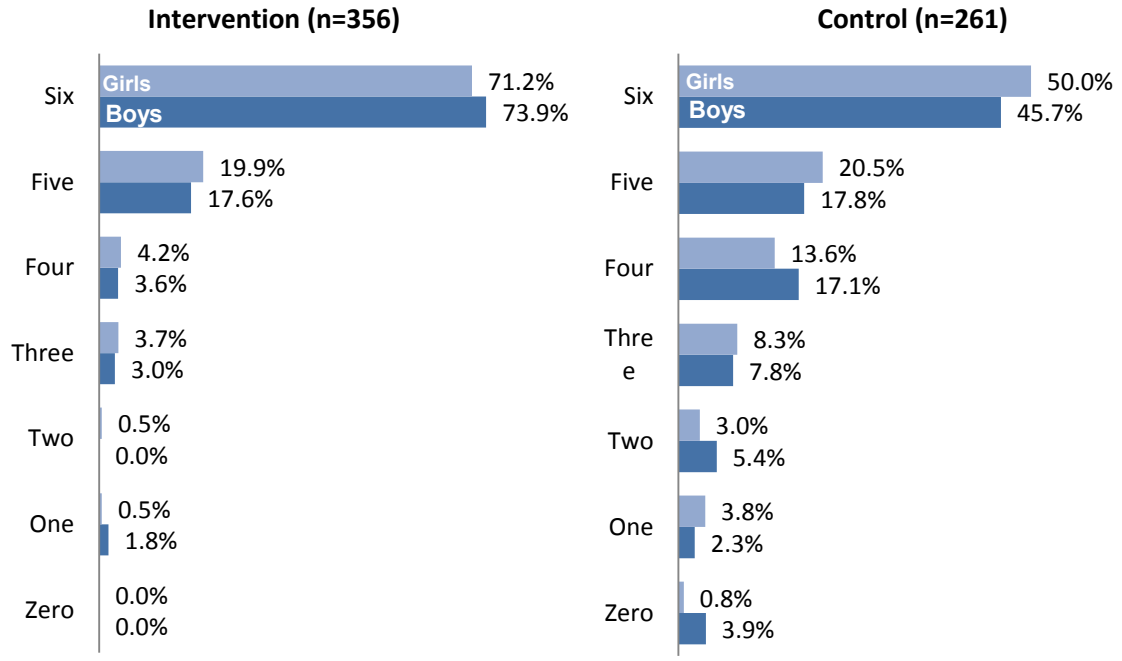
gains across baseline and endline. This can be explained by their high reading comprehension scores at baseline. Still, the CM intervention group surpassed their average reading comprehension scores at endline. This same trend also recurs across all grade levels suggesting that despite weaknesses in reading comprehension at baseline, the intervention group was able to catch up and surpass their control counterparts after 1.5 years of intervention.

Figure 29. Summary of Gains In Reading Comprehension (Out Of 6) Across Baseline and Endline by Grade Level (At Endline)



Reading comprehension results by sex. The figures below show the distributions of reading comprehension scores for boys and girls at endline. For the intervention group, the results show little difference between the performance of girls and boys in reading comprehension. For the control group, differences between boys and girls were more apparent: about 50% of girls were able to answer 6 comprehension questions when only 45% of boys were able to answer 6 questions. Overall in the control group, girls tend to outscore boys.

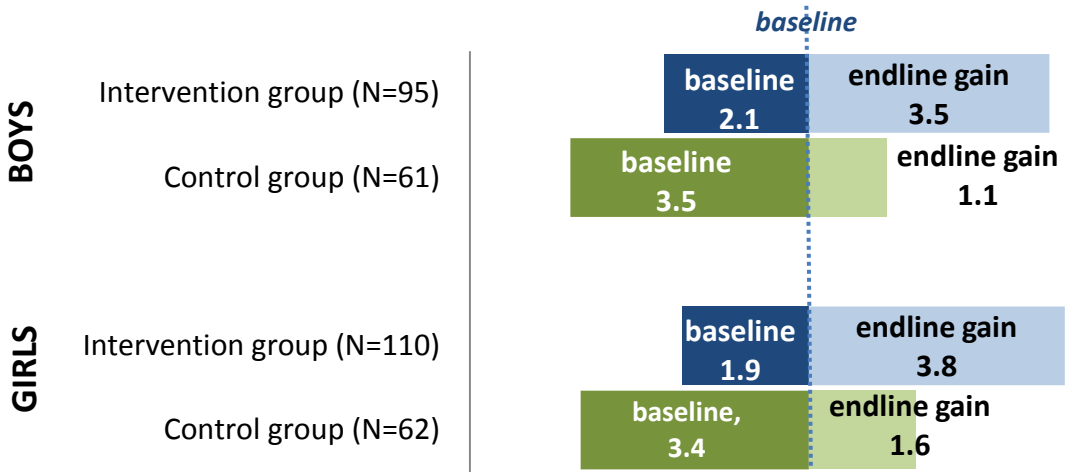
Figure 30. Number of Comprehension Questions Answered Correctly After the Second Reading, By Group and Sex



Looking across baseline and endline, girls gained more in reading comprehension than boys, in both intervention and control groups. Comparing the gains made by intervention and control girls, intervention girls were also found to gain significantly more (p=.000, d=1.08, ES=0.48). Intervention girls also outscored the control boys at endline.

Intervention boys also gained significantly more than their control boy counterparts, surpassing the control boys at endline (p=.000, d=1.07, ES=0.47). However, differences between boys' and girls' gains in the intervention group were not significant. These results suggest that though the HYVALL intervention likely contributed to strong gains in reading comprehension; it did not seem to contribute to differentiated gains for boys and girls. This makes sense when considering participation rates in HYVALL activities was equally high for both girls and boys alike.

Figure 31. Summary of Gains in Reading Comprehension (Out Of 6) Across Baseline and Endline by Sex



Conclusions and recommendations

The results of the analysis of reading gains between the baseline and endline showed that the HYVALL reading intervention was effective in improving oral reading fluency and comprehension for the students they served, regardless of gender and grade level. Though the effect was greater for younger students than older students, older students were still found to greatly benefit from the intervention. The program effect can be confirmed by the strong implementation and attendance data which indicates high fidelity of program implementation. The success of the intervention with respect to increasing reading gains is likely due to the increased exposure to literacy activities it offered, as well as the close and regular mentorship it provided its beneficiary students. In addition to the activities it provided outside of school, the program also promoted student practice at home allowing students to take home writing samples they had produced in the centers, to practice reading and writing further. This not only served to continue supporting students' literacy development beyond the centers but also linked family members to students' literacy development, contributing to the establishment of a culture of reading. In future design of this program in Senegal or other contexts, it is strongly recommended that a similar design be considered as it has potential for scalability without losing quality. However, the success of this model depends on the commitment of community volunteers. To foster this commitment, it is recommended to implement an incentives system that formally recognizes the contribution of the volunteers and grant them credits that could be allocated for university, employment or for the civil service. In ensuring volunteers engagement and in providing them with the necessary support and quality materials (guide, books) a program like HYVALL promises significant and positive results for student literacy gains and school readiness.

References

Entwisle, D., Alexander, K., & Olson, L. 2000. "Summer Learning and Home Environment." In R. Kahlenberg Ed., A Notion at Risk: Preserving Public Education as an Engine for Social Mobility pp. 9-30. New York: Century Foundation Press.

Save the Children (2013). Beyond School Walls: A Boost for Readers. Found at: http://www.savethechildren.org/atf/cf/%7B9def2ebe-10ae-432c-9bd0-df91d2eba74a%7D/BEYOND_SCHOOL_WALLS_LITERACY_BOOST_2013.PDF

Appendices

Appendix A: Methodology

The reading assessments were designed to provide a snapshot of student's reading capabilities in fluency and reading comprehension at their instructional or independent level. This assessment took an average of 6 minutes to administer and included the following sub-tests:

- *First reading of a passage.* Students 'reading accuracy and automaticity in reading a passage aloud was assessed. The subtest was timed and capped at 60 seconds, allowing the student to read until the 60 second mark. This yielded a score of correct words per minute and served to determine students' instructional or independent level. If students made 75% of more errors in reading the text or could not get through 75% of the text in a 1-minute period, they were asked to restart the first reading of a passage at a grade-level below. Below is the breakdown of number of words per passage for each assessment:

Assessment level	Number of words in passage
CP (grade 2)	34 words
CE (grade 4)	56 words
CM (grade 6)	75 words

- *Reading comprehension questions (after first reading).* Students' capacity for reading comprehension was assessed based on their first timed reading of the passage. A total of five literal questions were asked. Students could not refer back to the text to help them answer questions.
- *Second reading of a passage.* Students 'reading accuracy and automaticity in reading a passage aloud was assessed. The passage is the same "instructional" or "independent" level identified and read during the first reading. The subtest was timed but not capped at 60 seconds, allowing the student to read until the end. This yielded a score of correct words per minute. At the second reading, students' "instructional" or "independent" level is already determined so students who struggle with this reading are not dropped down to a lower level.
- *Reading comprehension questions (after second reading).* Students' capacity for reading comprehension was assessed based on their first timed reading of the passage. A total of five literal questions and a sixth inferential question were asked. Students were allowed to refer back to the text to help them answer questions.

A group of administrators drawn from the HYVALL team from the five regions were trained on test administration following a specific protocol. The test was piloted to assess and equate three different texts with a randomly selected group of CP, CE, and CM classes in Dakar. 40 students from each grade level (120 total) participated in the pilot. Following this pilot, texts that provided statistically close fluency and comprehension results were selected to use at baseline, midline, and endline. Because 120 students were insufficient to provide accurate psychometrics scores, the baseline and endline texts for each grade level were the same. This was deemed to have no impact on results given students who had started with a text at baseline had moved up to another grade level at endline.

Sampling: At baseline, in March 2013, EDC randomly selected control and intervention schools. From these schools, test administrators were instructed in how to randomly select 5 girls and 5 boys from each grade-level in these schools to retain for the assessment for a total of 570 students. For midline and endline assessments, those same students who were still present at their schools were assessed. For those who had left the school, test administrators were trained to randomly select replacement students at this same grade-level and sex.

Data entry: At baseline, assessments were conducted using a paper-based tool and thereafter, results were entered into a data shell in Excel. At midline and endline, data from assessments was collected live and electronically using Survey To Go. This was to in order to minimize data entry errors and to expedite data collection and entry process.

Missing data analysis

Missing data analysis of those students who dropped out of the sample across baseline and endline shows that the majority of students who left the sample were in grade 5 and 6 at the baseline which places them in secondary school at the endline. As these assessments were carried out in the primary school setting, these students would have been eliminated from the sample. Apart from these cases, twice as many control students left the sample in comparison to intervention students. T-test comparisons of mean scores of baseline show statistically significant differences between the fluency (WCPM) of intervention students who left and those control students who left (favoring intervention students) ($p=.012$). Furthermore, in comparing those students who left the sample to those who remained, statistically significant differences in baseline comprehension scores ($p=.022$) of those who were retained were found over those who dropped out. These are interesting to note because it suggests that those with weaker scores dropped out for certain reasons. Further study is needed to determine the reasons for these dropouts.

Data Analysis

All collected data were cleaned by EDC M&E staff and analyzed using standard statistical techniques, as needed for different analytical purposes. The results were disaggregated by sex and grade-level, as appropriate. Central tendency analysis (e.g. mean, frequencies, median) conducted for continuous demographic variables. Comparison of means statistical tests (paired and independent samples *t*-test) were conducted to estimate differences between groups such as grade level, type of participant (control

vs. intervention) and sex, where appropriate. Bivariate statistical analyses (e.g., correlations) were conducted to examine the relationship between different variables.

Study Limitations

The differences in mode of administration between the baseline and the endline assessment potentially complicated the collection and entry of data, especially in the baseline study. For the baseline study, data was collected using paper-based instruments and was then entered into an Excel shell. This may have led to a certain percentage of data entry error even despite 100% verification by two agents and again by an EDC agent prior to the endline assessment. At endline, instruments were administered using a tablet programmed with Survey-To-Go. This allowed for live data collection and included safeguards to minimized administration error (i.e. timing and calculations). Due to differences in administration, data quality may not be similar across baseline and endline.

At baseline, it was discovered that the control group had significantly better fluency and comprehension scores in comparison to the intervention group scores. Although the samples were selected at random (the control group was selected from comparable schools to the intervention schools in region). However, students selected from the control schools were randomly selected without considering their reading level. On the other hand, the students selected to participate in the HYVALL program and hence, students from the "intervention" were students who had been referred to the program due to their weaknesses in reading. This is because the goal of the intervention was to target students with reading difficulties. Hence, from the get-go students in the control group were already well ahead in reading compared to their "intervention" counterparts. This indicates that the samples are not really comparable.

Fidelity of implementation data was also incomplete for some of the variables. For example, no data was provided for Thies for the outside of center activities. The reason for this that volunteers in this region did not keep track of students' reading passports.

Instrument reliability and validity

It is not possible to provide reliability results for this instrument given the number of sub-tests on the test is limited to two per level (fluency and comprehension). However, it is possible to establish the concurrent validity of the fluency scores between the first and second reading scores and between fluency and comprehension scores. Overall, bivariate correlations between these scores were found statistically significant for all instrument levels. That is to say, a linear regression analysis between the fluency scores for the first and the second reading show that test scores are significantly and positively correlated. Fluency and comprehension scores provide similar results. This is shown in the tables below.

Concurrent validity: correlations for CP test instrument at endline

Correlations					
		CP_COMP_Total_correct_1st_reading_endline	WCPM_1st_reading_CP_endline	CP_COMP_Total_correct_2nd_reading_endline	WCPM_2nd_reading_CP_endline
CP_COMP_Total_correct_1st_reading_endline	Pearson Correlation	1	.773**	.815**	.764**
	Sig. (2-tailed)		.000	.000	.000
	N	61	61	60	60
WCPM_1st_reading_CP_endline	Pearson Correlation	.773**	1	.607**	.712**
	Sig. (2-tailed)	.000		.000	.000
	N	61	68	60	60
CP_COMP_Total_correct_2nd_reading_endline	Pearson Correlation	.815**	.607**	1	.613**
	Sig. (2-tailed)	.000	.000		.000
	N	60	60	60	60
WCPM_2nd_reading_CP_endline	Pearson Correlation	.764**	.712**	.613**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	60	60	60	60
**. Correlation is significant at the 0.01 level (2-tailed).					

Concurrent validity: correlations for CE test instrument at endline

Correlations					
		CE_COMP_Total_correct_1st_reading_endline	WCPM_1st_reading_CE_endline	CE_COMP_Total_correct_2nd_reading_endline	WCPM_2nd_reading_CE_endline
CE_COMP_Total_correct_1st_reading_endline	Pearson Correlation	1	.633**	.644**	.353**
	Sig. (2-tailed)		.000	.000	.000
	N	309	309	260	260
WCPM_1st_reading_CE_endline	Pearson Correlation	.633**	1	.258**	.728**
	Sig. (2-tailed)	.000		.000	.000
	N	309	328	260	260
CE_COMP_Total_correct_2nd_reading_endline	Pearson Correlation	.644**	.258**	1	.282**
	Sig. (2-tailed)	.000	.000		.000
	N	260	260	260	260
WCPM_2nd_reading_CE_endline	Pearson Correlation	.353**	.728**	.282**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	260	260	260	260
**. Correlation is significant at the 0.01 level (2-tailed).					

Concurrent validity: correlations for CM test instrument at endline

Correlations					
		WCPM_1st_reading_CM_endline	CM_COMP_Total_correct_1st_reading_endline	WCPM_2nd_reading_CM_endline	CM_COMP_Total_correct_2nd_reading_endline
WCPM_1st_reading_CM_endline	Pearson Correlation	1	.669**	.692**	.226**
	Sig. (2-tailed)		.000	.000	.000
	N	393	383	299	299
CM_COMP_Total_correct_1st_reading_endline	Pearson Correlation	.669**	1	.352**	.680**
	Sig. (2-tailed)	.000		.000	.000
	N	383	383	299	299
WCPM_2nd_reading_CM_endline	Pearson Correlation	.692**	.352**	1	.190**
	Sig. (2-tailed)	.000	.000		.001
	N	299	299	299	299
CM_COMP_Total_correct_2nd_reading_endline	Pearson Correlation	.226**	.680**	.190**	1
	Sig. (2-tailed)	.000	.000	.001	
	N	299	299	299	299
**. Correlation is significant at the 0.01 level (2-tailed).					

Correspondence between fluency (WCPM) and comprehension

The figure below provides an illustrative snapshot of the correspondence between fluency in comprehension. For the purpose of setting benchmarks, these numbers are useful because they can serve to guide stakeholders in the identification of thresholds for fluency based on real data.

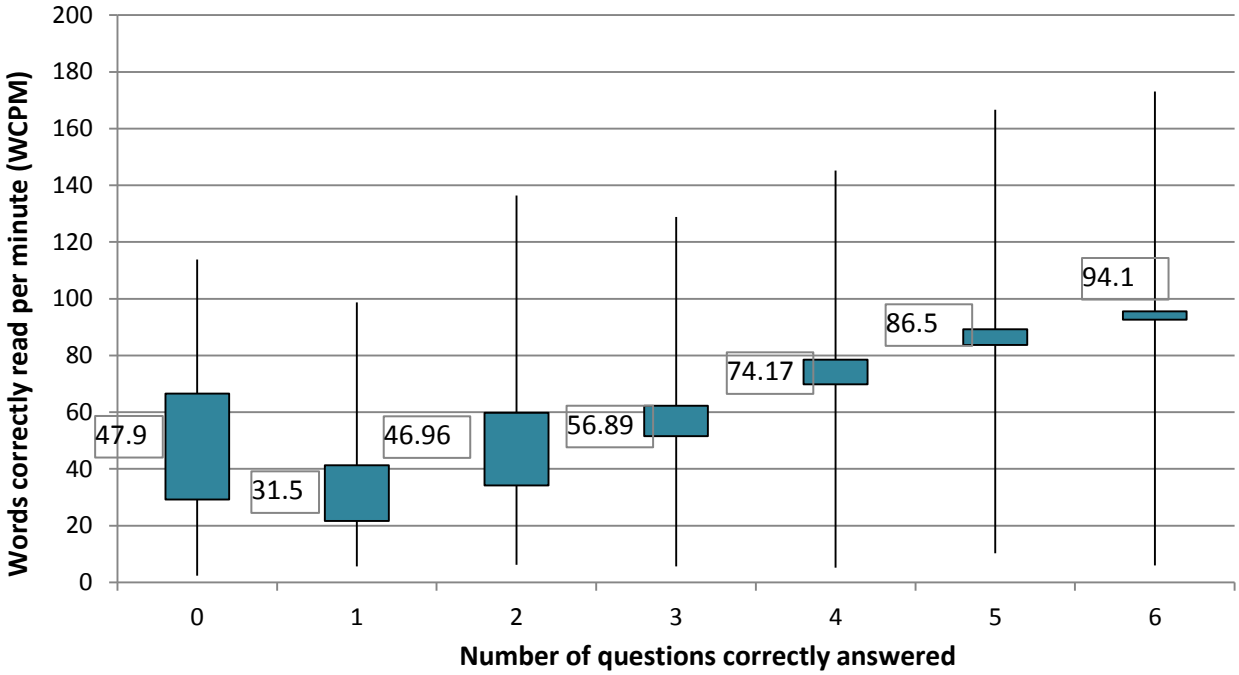


Table 3. Summary results from intervention groups (HYVALL) across baseline and endline

Description		N	Mean	St. Dev	P-value	Effect size
Percentage of words read correctly (accuracy)						
CE level (percentage)	*Baseline*	95	26.14%	29.8%		
	Endline	101	93.22%	13.09%	.000	0.9
CM level (percentage)	*Baseline*	90	60.55%	39.89%		
	Endline	104	95.57%	7.38	.000	0.66
Number of words read correctly (WCPM)						
CE level	*Baseline*	101	10.22	16.87		
	Endline	101	77.21	27.92	.000	0.9
CM level	*Baseline*	115	47.55	35.76		
	Endline	104	95.29	26.25	.000	0.74
Comprehension questions answered correctly after the first reading (percentage)						
CE level (out of 5)	*Baseline*	101	20.2%	29.4%		
	Endline	101	88.4%	19.6%	.000	0.89
CM level (out of 5)	*Baseline*	104	39.4%	38%		
	Endline	104	82.6%	20.1%	.000	0.71
Comprehension questions answered correctly after the second reading (percentage)						
CE level (out of 6)	*Baseline*	101	19.1%	30.8%		
	Endline	101	94.1%	12.1%	.000	0.92
CM level (out of 6)	*Baseline*	104	46.6%	40.8%		
	Endline	104	94.2%	11.3%	.000	0.75

Table 4. Summary results from control groups across baseline and endline

Description		N	Mean	St. Dev	P-value	Effect size
Percentage of words read correctly (accuracy)						
CE level (percentage)	*Baseline*	73	47.64%	35.09%		
	Endline	73	80.4%	28.48%	.000	0.7
CM level (percentage)	*Baseline*	54	71.23%	31.15%		
	Endline	54	87.92%	17.18%	.000	0.57

Number of words read correctly (WCPM)

CE level	*Baseline*	73	19.62	18.23		
	Endline	73	55.95	31.05	.000	0.79
CM level	*Baseline*	54	52.1	33.15		
	Endline	54	71.27	27.57	.000	0.65
Comprehension questions answered correctly after the first reading (percentage)						
CE level (out of 5)	*Baseline*	69	35.4%	30.2%		
	Endline	69	72%	30.9%	.000	0.71
CM level (out of 5)	*Baseline*	54	56.6%	31.2%		
	Endline	54	66.2%	30.04%	.035	0.28
Comprehension questions answered correctly after the second reading (percentage)						
CE level (out of 6)	*Baseline*	69	50.6%	34.1%		
	Endline	69	80.2%	28.2%	.000	0.65
CM level (out of 6)	*Baseline*	54	69.1%	32.3%		
	Endline	54	79.6%	23.5%	.041	0.27

NB. Grade levels represent student level at endline