



USAID
FROM THE AMERICAN PEOPLE

PAQUED: DRC

Projet d'Amélioration de la Qualité de l'Éducation (PAQUED)

2014 Endline Report of Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA)

October 2014

This publication was produced for review by the United States Agency for International Development and by Education Development Center. It was prepared by RTI International.

PAQUED: DRC

Projet d'Amélioration de la Qualité de l'Éducation (PAQUED)

**Endline Report of Early Grade Reading Assessment (EGRA) and
Early Grade Math Assessment (EGMA)**

Contract: AID-623-A-09-00010

EDC Project: 11224

Date: October 2014

Prepared for
USAID/DRC and EDC

Prepared by
Jennae Bulat, Aarnout Brombacher, Timothy Slade
Jennifer Iriondo-Perez, Maureen Kelly, Susan Edwards
RTI International 3040 Cornwallis Road
Post Office Box 12194 Research Triangle Park, NC 27709-2194

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.

RTI International is one of the world's leading research institutes, dedicated to improving the human condition by turning knowledge into practice. Our staff of more than 3,700 provides research and technical services to governments and businesses in more than 75 countries in the areas of health and pharmaceuticals, education and training, surveys and statistics, advanced technology, international development, economic and social policy, energy and the environment, and laboratory testing and chemical analysis. For more information, visit www.rti.org. RTI International is a trade name of Research Triangle Institute.

Table of Contents

	Page
List of Figures	vi
List of Tables	ix
Acronyms	xx
Executive Summary	1
Evolving Donor, Country, and DRC Education Policy Context: 2009-2014.....	1
Study Design and Limitations	2
Data Collection	5
Timing of Studies.....	5
Tools Used to Measure Impact	5
Findings	7
Impact on Student Reading Performance	7
Discussion of Mathematics Outcomes.....	20
Recommendations.....	26
Introduction.....	28
Brief Chronological History of Impact Evaluation Activities under PAQUED.....	28
Structure of the Report.....	30
Chapter A: PAQUED Project Context and Evolution; Study Design, Objectives, and Limitations	31
1. Project Context and Evolution	31
Project Implementation pre-2012	32
Realignment of PAQUED’s Focus for 2012–2014	33
Elements of the PAQUED Intervention pre-Realignment.....	33
PAQUED Inputs Targeting Students	33
PAQUED Inputs Targeting Teachers	36
Elements of the “Reading Program” Intervention Post-Realignment	38
Reading Program Inputs Targeting Students	38
Reading Program Inputs Targeting Teachers	38
2. Study Design and Objectives	40
Research Design	40
The EGRA Tool.....	40
The EGMA Tool.....	43
Student Context, Teacher, and Head Teacher Interviews.....	45
Assessor Training	45
3. Limitations	46
Misalignment of the Intervention and Assessment.....	46
Curricular Expectations of Grade 2 Students’ Reading Abilities.....	47

Changes in Sampling Design.....	47
Chapter B: Results and Analysis of Student Reading Performance in Reading Program Schools	50
1. Descriptive Statistics.....	50
2014 Sample	51
2. EGRA Subtask Outcomes	60
Vocabulary.....	61
Initial Sound Identification	65
Listening Comprehension.....	68
Grapheme Recognition	71
Familiar Word Reading	74
Invented Word Reading.....	76
Oral Reading Fluency	78
Reading Comprehension.....	80
Dictation	83
Student Characteristics Analyses.....	85
Teacher Characteristics Analyses	90
Head Teacher Characteristics Analyses.....	91
3. Summary and Conclusions.....	91
Chapter C: Results and Analysis of Student Reading Performance in Accessible PAQUED and Accessible Control Schools.....	94
1. Student Performance on EGRA Measures in Accessible PAQUED and Accessible Control Schools.....	94
Descriptives	94
2014 Sample	96
2. EGRA Subtest Outcomes	105
Vocabulary.....	107
Initial Sound Identification	113
Listening Comprehension.....	117
Grapheme Recognition	122
Familiar Word Reading	127
Invented Word Reading.....	131
Oral Reading Fluency	134
Reading Comprehension.....	137
Dictation	140
Student Characteristics Analyses.....	143
Teacher Characteristics Analyses	148
Head Teacher Characteristics Analyses.....	148
3. Summary and Conclusions.....	149
Chapter D: Results and Analysis of Student Reading Performance in PAQUED and Control Schools.....	151

1. Student Performance on EGRA Measures in PAQUED and Control Schools	151
Descriptives	151
2014 Sample	153
2. EGRA Subtest Outcomes	162
Vocabulary	164
Initial Sound Identification	171
Listening Comprehension	176
Grapheme Recognition	182
Familiar Word Reading	188
Invented Word Reading	193
Oral Reading Fluency	197
Reading Comprehension	201
Dictation	207
Student Characteristics Analyses	211
Teacher Characteristics Analyses	217
Head Teacher Characteristics Analyses	217
3. Summary and Conclusions	217
Chapter E: Results and Analysis of Student Mathematics Performance in Reading Program, Accessible PAQUED, and Accessible Control Schools	220
1. EGMA Results in Reading Program, Accessible PAQUED, and Accessible Control Schools	220
EGMA Bandundu	222
2. EGMA Results Disaggregated by Sex in Reading Program, Accessible PAQUED, and Accessible Control Schools	232
EGMA Bandundu Accessible PAQUED	232
EGMA Equateur Accessible PAQUED schools	234
EGMA Orientale Accessible PAQUED schools	235
EGMA Bandundu Reading Program schools	237
EGMA Equateur Reading Program schools	238
EGMA Orientale Reading Program schools	240
Chapter F: Results and Analysis of Student Mathematics Performance in PAQUED and Control Schools	242
1. EGMA Results in PAQUED and Control Schools	242
EGMA Bandundu	243
EGMA Equateur	246
EGMA Orientale	249
2. EGMA Results Disaggregated by Sex in PAQUED and Control Schools	252
EGMA Bandundu PAQUED Schools	252
EGMA Equateur PAQUED Schools	254
EGMA Orientale PAQUED Schools	255

Chapter G: Conclusions and Recommendations	257
Discussion of Reading Outcomes	258
Discussion of Mathematics Outcomes.....	259
Recommendations.....	260
Annex I: Details on Study Methodology	262
Annex II: The Early Grade Reading Assessment (EGRA)/Early Grade Mathematics Assessment (EGMA) Data Collection Process	280
Annex III: About the Assessment Instruments	284
Annex IV: Reading Program Schools: 2012-2014 Comparisons	296

List of Figures

Figure ES1. Subdivisions Containing Schools Assessed at Endline and Relative Size of the Study Populations	4
Figure ES2. Performance of Students in PAQUED Schools Relative to DRC Benchmarks.....	13
Figure ES3. Performance of Students in Accessible PAQUED Schools Relative to DRC Benchmarks	17
Figure ES4. Performance of Students in Reading Program Schools Relative to DRC Benchmarks.....	20
Figure A1. Subdivisions Containing Schools Assessed at Endline and Relative Size of the Study Populations	49
Figure B1. Grade 2 Scores on <i>Vocabulary</i> at Endline by Province.....	64
Figure B2. Grade 4 Scores on <i>Vocabulary</i> at Endline by Province.....	65
Figure B3. Grade 2 Scores on <i>Initial Sound Identification</i> at Endline by Province	67
Figure B4. Grade 4 Scores on <i>Initial Sound Identification</i> at Endline by Province	68
Figure B5. Grade 2 Scores on <i>Listening Comprehension</i> at Endline by Province	70
Figure B6. Grade 4 Scores on <i>Listening Comprehension</i> at Endline by Province	71
Figure B7. Grade 2 Scores on <i>Grapheme Recognition</i> at Endline by Province	73
Figure B8. Grade 4 Scores on <i>Grapheme Recognition</i> at Endline by Province	74
Figure B9. Grade 4 Scores on <i>Familiar Word Reading</i> at Endline by Province	76
Figure B10. Grade 4 Scores on <i>Invented Word Reading</i> at Endline by Province	78
Figure B11. Grade 4 Scores on <i>Oral Reading Fluency</i> at Endline by Province.....	80
Figure B12. Grade 4 Scores on <i>Reading Comprehension</i> at Endline by Province.....	82

Figure B13. Grade 4 Overall Correspondence between Oral Reading Fluency and Reading Comprehension.....	83
Figure B14. Grade 4 Scores on <i>Dictation</i> at Endline by Province	85
Figure B15. Performance of Students in Reading Program Schools Relative to DRC Benchmarks.....	93
Figure C1. Grade 2 Scores on <i>Vocabulary</i> at Endline by Province.....	112
Figure C2. Grade 4 Scores on <i>Vocabulary</i> at Endline by Province.....	112
Figure C3. Grade 2 Scores on <i>Initial Sound Identification</i> at Endline by Province	116
Figure C4. Grade 4 Scores on <i>Initial Sound Identification</i> at Endline by Province	117
Figure C5. Grade 2 Scores on <i>Listening Comprehension</i> at Endline by Province	121
Figure C6. Grade 4 Scores on <i>Listening Comprehension</i> at Endline by Province	122
Figure C7. Grade 2 Scores on <i>Grapheme Recognition</i> at Endline by Province	126
Figure C8. Grade 4 Scores on <i>Grapheme Recognition</i> at Endline by Province	127
Figure C9. Grade 4 Scores on <i>Familiar Word Reading</i> at Endline by Province	130
Figure C10. Grade 4 Scores on <i>Invented Word Reading</i> at Endline by Province	133
Figure C11. Grade 4 Scores on <i>Oral Reading Fluency</i> at Endline by Province.....	136
Figure C12. Grade 4 Scores on <i>Reading Comprehension</i> at Endline by Province.....	139
Figure C13. Grade 4 Overall Correspondence between Oral Reading Fluency and Reading Comprehension.....	140
Figure C14. Grade 4 Scores on <i>Dictation</i> at Endline by Province	143
Figure C15. Performance of Students in Accessible PAQUED Schools Relative to DRC Benchmarks	150
Figure D1. Grade 2 Scores on <i>Vocabulary</i> at Endline by Province.....	169
Figure D2. Grade 4 Scores on <i>Vocabulary</i> at Endline by Province.....	170
Figure D3. Grade 6 Scores on <i>Vocabulary</i> at Endline by Province.....	170
Figure D4. Grade 2 Scores on <i>Initial Sound Identification</i> at Endline by Province	175
Figure D5. Grade 4 Scores on <i>Initial Sound Identification</i> at Endline by Province.....	175
Figure D6. Grade 6 Scores on <i>Initial Sound Identification</i> at Endline by Province	176
Figure D7. Grade 2 Scores on <i>Listening Comprehension</i> at Endline by Province	181
Figure D8. Grade 4 Scores on <i>Listening Comprehension</i> at Endline by Province	181
Figure D9. Grade 6 Scores on <i>Listening Comprehension</i> at Endline by Province	182
Figure D10. Grade 2 Scores on <i>Grapheme Recognition</i> at Endline by Province	187

Figure D11. Grade 4 Scores on <i>Grapheme Recognition</i> at Endline by Province	187
Figure D12. Grade 6 Scores on <i>Grapheme Recognition</i> at Endline by Province	188
Figure D13. Grade 4 Scores on <i>Familiar Word Reading</i> at Endline by Province.....	192
Figure D14. Grade 6 Scores on <i>Familiar Word Reading</i> at Endline by Province.....	192
Figure D15. Grade 4 Scores on <i>Invented Word Reading</i> at Endline by Province	196
Figure D16. Grade 6 Scores on <i>Invented Word Reading</i> at Endline by Province	196
Figure D17. Grade 4 Scores on <i>Oral Reading Fluency</i> at Endline by Province	200
Figure D18. Grade 6 Scores on <i>Oral Reading Fluency</i> at Endline by Province	201
Figure D19. Grade 4 Scores on <i>Reading Comprehension</i> at Endline by Province.....	205
Figure D20. Grade 6 Scores on <i>Reading Comprehension</i> at Endline by Province.....	205
Figure D21. Grade 4 Overall Correspondence between Oral Reading Fluency and Reading Comprehension.....	206
Figure D22. Grade 6 Overall Correspondence between Oral Reading Fluency and Reading Comprehension.....	206
Figure D23. Grade 4 Scores on <i>Dictation</i> at Endline by Province	210
Figure D24. Grade 6 Scores on <i>Dictation</i> at Endline by Province	210
Figure D25. Performance of Students in PAQUED Schools Relative to DRC Benchmarks.....	219
Figure E1. EGMA Bandundu: 2014 Mean Percentage Scores by Subtask, Grade, and Treatment Group	224
Figure E2. EGMA Equateur: 2014 Mean Percentage Scores by Subtask, Grade, and Treatment Group.....	228
Figure E3. EGMA Orientale: 2014 Mean Percentage Scores by Subtask, Grade, and Treatment Group.....	231
Figure E4. EGMA Bandundu Accessible PAQUED Schools: 2014 Mean Percentage Scores by Subtask, Grade, and Sex.....	234
Figure E5. EGMA Equateur Accessible PAQUED Schools: 2014 Mean Percentage Scores by Subtask, Grade, and Sex.....	235
Figure E6. EGMA Orientale Accessible PAQUED schools: 2014 Mean Percentage Scores by Subtask, Grade, and Sex.....	237
Figure E7. EGMA Bandundu Reading Program Schools: 2014 Mean Percentage Scores by Subtask, Grade, and Sex.....	238
Figure E8. EGMA Equateur Reading Program Schools: 2014 Mean Percentage Scores by Subtask, Grade, and Sex.....	240

Figure E9. EGMA Orientale Reading Program Schools: 2014 Mean Percentage Scores by Subtask, Grade, and Sex.....	241
Figure F1. EGMA Bandundu: 2014 Mean Percentage Scores by Subtask, Grade, and Treatment Group	245
Figure F2. EGMA Equateur: 2014 Mean Percentage Scores by Subtask, Grade, and Treatment Group.....	248
Figure F3. EGMA Orientale: 2014 Mean Percentage Scores by Subtask, Grade, and Treatment Group.....	250
Figure F4. EGMA Bandundu Treatment: 2014 Mean Percentage Scores by Subtask, Grade, and Sex	253
Figure F5. EGMA Equateur Treatment: 2014 Mean Percentage Scores by Subtask, Grade, and Sex	255
Figure F6. EGMA Orientale Treatment: 2014 Mean Percentage Scores by Subtask, Grade, and Sex	256
Figure A1.1 Subdivisions Containing Schools Assessed at Endline and Relative Size of the Study Populations	268

List of Tables

Table ES1. Study Populations and Analyses Permitted.....	3
Table ES2. Subtask Composition of Assessment Instruments	6
Table ES3. Reading Performance in PAQUED Schools, in Mean Scores, by Province	9
Table ES4. Student Performance in PAQUED Schools Relative to National Benchmarks.....	11
Table ES5. Reading Performance in Accessible PAQUED Schools, in Mean Scores, by Province	13
Table ES6. Student Performance in Accessible PAQUED Schools Relative to National Benchmarks.....	16
Table ES7. Reading Performance in Subset of Reading Program Schools (n=20), in Mean Scores.....	18
Table ES8. Student Performance in Reading Program Schools Relative to National Benchmarks.....	19

Table ES9. Change in Mathematics Performance in PAQUED Schools over Time, in Percentage of Items Correct, by Province.....	21
Table ES10. Change in Mathematics Performance in Accessible PAQUED Schools over Time, in Percentage of Items Correct, by Province.....	23
Table ES11. Change in Mathematics Performance in Reading Program Schools ($n = 20$) over Time, in Percentage of Items Correct.....	25
Table A1. PAQUED Intervention—Inputs Targeting Students	35
Table A2. PAQUED Intervention—Training Inputs Targeting Teachers	36
Table A3. PAQUED Intervention—Materials Inputs Targeting Teachers.....	37
Table A4. Reading Program—Inputs Targeting Students	38
Table A5. Reading Program—Additional Inputs Targeting Teachers	39
Table A6. Study Populations and Analyses Permitted	48
Table B1. Intended Sample by Province and Grade	50
Table B2. Actual Sample by Province and Grade	51
Table B3. School Management Type, by Province ($n = 43$)	51
Table B4. General Characteristics of the Overall Student Sample ($n = 1,103$).....	52
Table B5. Student SES Indicators ($n = 1,103$).....	53
Table B6. Student Responses to Select Demographic Questions ($n = 1,097$)	53
Table B7. Student Indication of Language Spoken in the Home	54
Table B8. General Characteristics of the Teacher Sample ($n = 86$)	55
Table B9. Teacher Responses to Select Survey Questions ($n = 86$).....	56
Table B10. General Characteristics of the Head Teacher Sample ($n = 42$).....	59
Table B11. Head Teacher Responses to Select Survey Questions ($n = 42$)	60
Table B12. Overall Percent Zero Scores and Mean Scores by Grade and Subtask at Endline	60
Table B13. Vocabulary Zero Scores and Percentage of Items Attempted, by Grade and Province.....	62
Table B14. Vocabulary Mean Scores by Type of Vocabulary Word, by Grade and Province	63
Table B15. Comparison of Vocabulary Mean Scores by Grade, Province, and Sex.....	63
Table B16. Initial Sound Identification Zero Scores and Percentage of Items Attempted by Grade and Province	66

Table B17. Comparison of Initial Sound Identification Mean Scores by Grade, Province, and Sex.....	66
Table B18. Listening Comprehension Zero Scores and Percentage of Items Attempted by Grade and Province.....	69
Table B19. Comparison of Listening Comprehension Mean Scores by Grade, Province, and Sex.....	69
Table B20. Grapheme Recognition Zero Scores and Percentage of Items Attempted by Grade and Province.....	72
Table B21. Comparison of Grapheme Recognition Mean Scores by Grade, Province, and Sex.....	72
Table B22. Familiar Word Reading Zero Scores and Percent Attempted for Grade 4 by Province.....	75
Table B23. Comparison of Familiar Word Reading Mean Scores for Grade 4 by Province and Sex.....	75
Table B24. Invented Word Reading Zero Scores and Percent Attempted for Grade 4 by Province.....	77
Table B25. Comparison of Invented Word Reading Mean Scores for Grade 4 by Province and Sex.....	77
Table B26. Oral Reading Fluency Zero Scores and Percent Attempted for Grade 4 by Province.....	78
Table B27. Comparison of Oral Reading Fluency Mean Scores for Grade 4 by Province and Sex.....	79
Table B28. Reading Comprehension Zero Scores and Percent Attempted for Grade 4 by Province.....	81
Table B29. Comparison of Reading Comprehension Mean Scores for Grade 4 by Province and Sex.....	81
Table B30. Dictation Zero Scores and Percent Attempted, for Grade 4 by Province.....	83
Table B31. Comparison of Dictation Mean Scores for Grade 4 by Province and Sex.....	84
Table B32. Chi-Squared Analyses of Grade 2 Student Characteristics with Student High Performance on the Grapheme Recognition Subtask.....	86
Table B33. Grade 2 Student Characteristics Logistic Regressions.....	87
Table B34. Chi-Squared Analyses of Grade 4 Student Characteristics with Student High Performance on the Grapheme Recognition Subtask.....	88
Table B35. Grade 4 Student Characteristics Logistic Regressions.....	90

Table B36. Teacher Characteristics Logistic Regressions.....	91
Table B37. Student Performance in Reading Program Schools Relative to National Benchmarks.....	92
Table C1. Intended School Sample by Province	95
Table C2. Intended Student Sample by Province and Grade.....	95
Table C3. Actual Sample by Province and Grade	96
Table C4. School Management Type, by Province ($n = 82$)	96
Table C5. General Characteristics of the Overall Student Sample ($n = 1,965$).....	97
Table C6. Student SES Indicators ($n = 1,965$).....	97
Table C7. Select Student Responses to Demographic Questions ($n = 1,097$)	98
Table C8. Student Indication of Language Spoken in the Home	99
Table C9. General Characteristics of the Teacher Sample ($n = 154$)	100
Table C10. Select Teacher Responses to Survey Questions ($n = 86$).....	100
Table C11. General Characteristics of the Head Teacher Sample ($n = 81$).....	104
Table C12. Select Head Teacher Responses to Survey Questions ($n = 42$)	104
Table C13. Overall Percent Zero Scores and Mean Scores by Subtask, Grade and Group at Endline	105
Table C14. Vocabulary Zero Scores and Percent Attempted by Grade and Province	108
Table C15. Vocabulary Difference-in-Differences Analyses by Grade, Region, and Group	109
Table C16. Comparison of Vocabulary Mean Scores by Grade, Province, Group, and Sex.....	110
Table C17. Initial Sound Identification Zero Scores and Percent Attempted by Grade and Province.....	113
Table C18. Initial Sound Identification Difference-in-Differences Analyses by Grade, Region, and Group	114
Table C19. Comparison of Initial Sound Identification Mean Scores by Grade, Province, Group, and Sex	115
Table C20. Listening Comprehension Zero Scores and Percent Attempted by Grade and Province.....	118
Table C21. Listening Comprehension Difference-in-Differences Analyses by Grade, Region, and Group	119

Table C22. Comparison of Listening Comprehension Mean Scores by Grade, Province, Group, and Sex	120
Table C23. Grapheme Recognition Zero Scores, Percent Attempted, and Mean Scores by Grade and Province	123
Table C24. Grapheme Recognition Difference-in-Differences Analyses by Grade, Region, and Group	124
Table C25. Comparison of Grapheme Recognition Mean Scores by Grade, Province, and Sex	125
Table C26. Familiar Word Reading Zero Scores and Percent Attempted for Grade 4 by Province	128
Table C27. Familiar Word Reading Difference-in-Differences Analyses by Region and Group.....	129
Table C28. Comparison of Familiar Word Reading Mean Scores for Grade 4 by Province, Group, and Sex	129
Table C29. Invented Word Reading Zero Scores and Percent Attempted for Grade 4 by Province	131
Table C30. Invented Word Reading Difference-in-Differences Analyses by Region and Group.....	132
Table C31. Comparison of Invented Word Reading Mean Scores for Grade 4 by Province, Group, and Sex	132
Table C32. Oral Reading Fluency Zero Scores and Percent Attempted for Grade 4 by Province	134
Table C33. Oral Reading Fluency Difference-in-Differences Analyses by Region and Group.....	135
Table C34. Comparison of Oral Reading Fluency Mean Scores for Grade 4 by Province, Group, and Sex	135
Table C35. Reading Comprehension Zero Scores and Percent Attempted for Grade 4 by Province	137
Table C36. Reading Comprehension Difference-in-Differences Analyses by Region and Group.....	138
Table C37. Comparison of Reading Comprehension Mean Scores for Grade 4 by Province, Group, and Sex	138
Table C38. Dictation Zero Scores and Percent Attempted for Grade 4 by Province	141
Table C39. Dictation Difference-in-Differences Analyses by Region and Group	142
Table C40. Comparison of Dictation Mean Scores for Grade 4 by Province and Sex.....	142

Table C41. Chi-Squared Analyses of Grade 2 Student Characteristics with Student High Performance on the Grapheme Sound Knowledge Subtask, Accessible PAQUED and Accessible Control Schools	144
Table C42. Chi-Squared Analyses of Grade 4 Student Characteristics with Student High Performance on the Grapheme Sound Knowledge Subtask, Accessible PAQUED and Accessible Control Schools	146
Table C43. Grade 4 Student Characteristics Logistic Regressions.....	147
Table C44. Teacher Characteristics Logistic Regressions.....	148
Table C45. Student Performance in Accessible PAQUED Schools Relative to National Benchmarks.....	149
Table D1. Intended School Sample by Province and Grade.....	151
Table D2. Intended Student Sample by Province and Grade ($n = 5,616$).....	152
Table D3. Actual School Sample by Province and Grade	153
Table D4. Actual Student Sample by Province and Grade ($n = 5,255$).....	153
Table D5. School Type, by Group ($n = 145$)	153
Table D6. General Characteristics of the Overall Student Sample ($n = 5,255$).....	154
Table D7. Student SES Indicators ($n = 5,255$)	155
Table D8. Select Student Responses to Demographic Questions ($n = 5,204$).....	155
Table D9. Student Indication of Language(s) Spoken in the Home	156
Table D10. General Characteristics of the Teacher Sample ($n = 389$).....	157
Table D11. Select Teacher Responses to Survey Questions ($n = 389$).....	158
Table D12. General Characteristics of the Head Teacher Sample ($n = 142$).....	161
Table D13. Select Head Teacher Responses to Survey Questions ($n = 142$)	161
Table D14. Overall Percent Zero Scores and Mean Scores by Grade and Subtask at Endline	162
Table D15. Vocabulary Zero Scores and Percent Attempted, by Grade and Province	165
Table D16. Vocabulary Difference-in-Differences Analyses by Grade, Province, and Group	166
Table D17. Comparison of Vocabulary Mean Scores by Grade, Province, Group, and Sex.....	167
Table D18. Initial Sound Identification Zero Scores and Percent Attempted, by Grade and Province.....	171

Table D19. Initial Sound Identification Difference-in-Differences Analyses by Grade, Province, and Group.....	172
Table D20. Comparison of Initial Sound Identification Mean Scores by Grade, Province, Group, and Sex	173
Table D21. Listening Comprehension Zero Scores and Percent Attempted by Grade and Province.....	176
Table D22. Listening Comprehension Difference-in-Differences Analyses by Grade, Province, and Group	178
Table D23. Comparison of Listening Comprehension Mean Scores by Grade, Province, Group, and Sex	179
Table D24. Grapheme Recognition Zero Scores and Percent Attempted by Grade and Province.....	183
Table D25. Grapheme Recognition Difference-in-Differences Analyses by Grade, Province, and Group	184
Table D26. Comparison of Grapheme Recognition Mean Scores by Grade Province, Group, and Sex.....	185
Table D27. Familiar Word Reading Zero Scores and Percent Attempted by Grade and Province.....	189
Table D28. Familiar Word Reading Difference-in-Differences Analyses by Grade, Province, and Group	190
Table D29. Comparison of Familiar Word Reading Mean Scores by Grade, Province, Group, and Sex	190
Table D30. Invented Word Reading Zero Scores and Percent Attempted by Grade and Province.....	193
Table D31. Invented Word Reading Difference-in-Differences Analyses by Grade, Province, and Group	194
Table D32. Comparison of Invented Word Reading Mean Scores by Grade, Province, Group, and Sex	195
Table D33. Oral Reading Fluency Zero Scores and Percent Attempted by Grade and Province	197
Table D34. Oral Reading Fluency Difference-in-Differences Analyses by Grade, Province, and Group	198
Table D35. Comparison of Oral Reading Fluency Mean Scores by Grade, Province, Group, and Sex.....	199

Table D36. Reading Comprehension Zero Scores and Percent Attempted by Grade and Province.....	201
Table D37. Reading Comprehension Difference-in-Differences Analyses by Grade, Province, and Group	202
Table D38. Comparison of Reading Comprehension Mean Scores by Grade, Province, Group, and Sex	203
Table D39. Dictation Zero Scores and Percent Attempted by Grade and Province	207
Table D40. Dictation Difference-in-Differences Analyses by Grade, Province, and Group	208
Table D41. Comparison of Dictation Mean Scores by Grade, Province, Group, and Sex.....	208
Table D42. Chi-Squared Analyses of Grade 2 Student Characteristics with Student High Performance on the Grapheme Recognition Subtask	211
Table D43. Chi-Squared Analyses of Grade 4 Student Characteristics with Student High Performance on the Grapheme Recognition Subtask	213
Table D44. Grade 4 Student Characteristics Logistic Regressions	215
Table D45. Chi-Squared Analyses of Grade 6 Student Characteristics with Student High Performance on the Grapheme Sound Knowledge Subtask	215
Table D46. Student Performance in PAQUED Schools Relative to National Benchmarks.....	218
Table E1. EGMA Bandundu: 2014 Percent Zero Scores and Mean Percentage Scores by Grade.....	222
Table E2. EGMA Bandundu: 2012 vs. 2014 Difference-in-Differences of Mean Percentage Scores by Grade.....	225
Table E3. EGMA Equateur: 2014 Percent Zero Scores and Mean Percentage Scores by Grade.....	226
Table E4. EGMA Equateur: 2012 vs. 2014 Difference-in-Differences of Mean Percentage Scores by Grade.....	228
Table E5. EGMA Orientale: 2014 Percent Zero Scores and Mean Percentage Scores by Grade.....	229
Table E6. EGMA Orientale: 2012 vs. 2014 Difference-in-Differences of Mean Percentage Scores by Grade.....	231
Table E7. EGMA Bandundu Accessible PAQUED Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade	233

Table E8. EGMA Equateur Accessible PAQUED Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade	234
Table E9. EGMA Orientale Accessible PAQUED Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade	236
Table E10. EGMA Bandundu Reading Program Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade	237
Table E11. EGMA Equateur Reading Program Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade	239
Table E12. EGMA Orientale Reading Program Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade	240
Table F1. EGMA Bandundu: 2014 Percent Zero Scores and Mean Percentage Scores by Grade.....	243
Table F2. EGMA Bandundu: 2010 vs. 2014 Difference-in-Differences of Mean Percentage Scores by Grade.....	245
Table F3. EGMA Equateur: 2014 Percent Zero Scores and Mean Percentage Scores by Grade.....	246
Table F4. EGMA Equateur: 2010 vs. 2014 Difference-in-Differences of Mean Percentage Scores by Grade.....	248
Table F5. EGMA Orientale: 2014 Percent Zero Scores and Mean Percentage Scores by Grade.....	249
Table F6. EGMA Orientale: 2010 vs. 2014 Difference-in-Differences of Mean Percentage Scores by Grade.....	251
Table F7. EGMA Bandundu PAQUED Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade.....	252
Table F8. EGMA Equateur PAQUED Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade.....	254
Table F9. EGMA Orientale PAQUED Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade.....	255
Table G1. Study Populations and Analyses Permitted	257
Table A1.1: 2010 Sample: Geographic Distribution by Treatment Type.....	264
Table A1.2: 2012 Sample Geographic Distribution by Treatment Type	265
Table A1.3: 2014 Sample A Number of Schools Sampled by Geographic Distribution and Treatment Type.....	269
Table A1.4: 2014 Sample A Number of Students Sampled by Geographic Distribution, Treatment Type, and Grade	269

Table A1.5: 2014 Sample B Number of Schools Sampled by Geographic Distribution and Treatment Type.....	270
Table A1.6: 2014 Sample B Number of Students Sampled by Geographic Distribution, Treatment Type, and Grade	270
Table A1.7: 2014 Sample A Frame and Sample Counts: Subdivisions (Stage 1).....	272
Table A1.8: 2014 Sample A Frame and Sample Counts for PAQUED Schools (Stage 2).....	273
Table A1.9: 2014 Sample A Frame and Sample Counts for Control Schools (Stage 2).....	273
Table A1.10: 2014 Sample B Frame and Sample Counts for Subdivisions (Stage 1).....	274
Table A1.11: 2014 Sample B Frame and Sample Counts for Accessible PAQUED Schools (Stage 2)	274
Table A1.12: 2014 Sample B Frame and Sample Counts for Accessible Control Schools (Stage 2)	275
Table A1.13: Chapter B Bonferroni Adjustment Families for Reading Program Study, EGRA Results.....	277
Table A1.14: Chapter C Section 1 - Bonferroni Adjustment Families for Accessible PAQUED Study, EGRA Results	278
Table A1.15: Chapter C Section 2 - Bonferroni Adjustment Families for PAQUED Study, EGRA Results.....	278
Table A1.16: Chapter D Sections 1 and 2 - Bonferroni Adjustment Families for Accessible PAQUED Study, EGMA Results	279
Table A1.17: Chapter D Sections 3 and 4 - Bonferroni Adjustment Families for Accessible PAQUED Study, EGMA Results	279
Table A3.1: Correlation Matrix for EGRA Tasks	287
Table A3.2: Correlation Matrix for EGMA Tasks.....	288
Table A3.3: Reliability Analysis of EGRA Tool.....	290
Table A3.4: Reliability analysis of EGMA Tool	290
Table A3.5: Comparison of EGRA and EGMA Instruments	291
Table A3.6: Example of Application of the Means Equating Method	294
Table A4.1: Reading Program Schools Change Over Time, Grade 2	296
Table A4.2: Reading Program Schools Change Over Time, Grade 4	297
Table A4.3: Reading Program Schools (n=20) Change Over Time, Mathematics, Grade 2.....	297

Table A4.4: Reading Program Schools (n=20) Change Over Time, Mathematics, Grade 4.....	298
---	-----

Acronyms

CELTA	<i>Centre de Linguistique Théorique et Appliquée</i>
cgpm	correct graphemes per minute
cnonwpm	correct non-words per minute (invented words)
cwpm	correct words per minute (familiar words)
DEFF	design effect
DRC	Democratic Republic of Congo
EDC	Education Development Center
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
FPC	Finite Population Correction
IAI	interactive audio instruction
IRR	inter-rater reliability
MEPSP	Ministry of Primary, Secondary, and Professional Education
orf	oral reading fluency
PAQUED	<i>Projet d'Amélioration de la Qualité de l'Éducation (PIEQ)</i>
PIEQ	Package for Improving Education Quality (PAQUED)
PCA	principal components analysis
RTI	RTI International
SBD	subdivisions
SES	socioeconomic indicators
TLM	teaching and learning material
USAID	U.S. Agency for International Development
WT	weight

Executive Summary

Funded by the U.S. Agency for International Development (USAID) and led by the Education Development Center (EDC), the *Projet d'Amélioration de la Qualité de l'Éducation* (PAQUED) project was a five-year initiative that aimed to raise student learning through improved teaching and school environments in the Democratic Republic of the Congo (DRC). As a partner in PAQUED, RTI International (RTI) was responsible for adapting and applying the EGRA and EGMA instruments to assess the program impact on student learning outcomes.

This document presents a discussion of PAQUED's impact on three distinct populations of schools. Each discussion reflects a different stage in the significant evolution of PAQUED's focus and scope that took place between project design in 2009 and a formal realignment that occurred in January 2013. These three populations cannot be compared directly to each other, so each is dealt with in a separate chapter of the report.¹ They are presented in the introduction below in chronological order of their emergence as the point of focus for the PAQUED project; however, in the main body of the report they are presented in reverse chronological order, reflecting the degree to which each was expected to produce meaningful impact.

Evolving Donor, Country, and DRC Education Policy Context: 2009-2014

The PAQUED project design has evolved significantly since implementation began in 2009. The most significant characteristic of the project evolution was the measured shift away from an emphasis on primary education quality broadly construed toward a focus on improving early grade reading outcomes. As is often the case, this project evolution both influenced and was influenced by the evolving government of the DRC (GDRC) and donor education priorities.

An important contextual development during this period was USAID's release of its 2011-2015 Education Strategy, which strongly emphasized early grade reading. The GDRC also concurrently followed up on its participation in the 2012 Global Program for Education (GPE) *All Children Reading* Conference in Kigali by committing to put in place the policy framework and investments necessary to reach ambitious targets for increasing the number of readers in the DRC.

The original PAQUED teacher training and classroom interventions conformed to the Ministry of Education's (MEPSP) requirement that they align with the existing curriculum and textbooks. However, in response to the changing context, redesigned teacher training and classroom interventions were implemented in 45 Reading Program

¹ Please see *Annex I* for a detailed discussion of the sampling approaches and why they must be treated separately.

schools in the final academic year of the project. These new approaches included approaches characteristic of successful reading programs in other contexts.

Study Design and Limitations

The goals of the 2014 studies and this report were to describe (1) changes in student performance in reading and mathematics that took place in PAQUED schools between 2010 and 2014, (2) similar changes in the Accessible PAQUED schools between 2012 and 2014, and (3) the current state of student learning outcomes in PAQUED’s Reading Program schools.

The impact of the intervention in the project’s 3,000 target schools (1,000 each in the provinces of Bandundu, Equateur, and Orientale) was evaluated using a sample of 145 schools. Of this number, 36 were control schools and were not expected to have received any intervention and 109 were treatment schools (called “PAQUED” schools throughout this report) that were expected to have received the intervention. The results presented in this report for PAQUED schools are generalizable to the provincial level. The findings of this study are presented in *Chapter D* of this report.

The impact of the intervention in the 618 schools out of the original 3,000 that PAQUED deemed most easily accessible—and therefore most likely to have received the foreseen inputs—was evaluated using a sample of 75 schools. Of this number, 36 were “Accessible Control” that were not expected to have received any intervention and 39 were treatment schools (called “Accessible PAQUED” schools throughout this report) that were expected to have received the intervention. The results presented in this report for Accessible PAQUED schools are generalizable to schools which meet the same criteria for accessibility.² The findings of this study are presented in *Chapter C* of this report.

The impact of the intervention in the subset of 45 schools (out of the 618 Accessible PAQUED schools) on which PAQUED focused intensely from January 2013 through June 2014—and to which it provided a modified, far more robust set of inputs, including student texts, regular coaching, and peer-led continuous professional development—was evaluated by attempting to assess all 44 schools (only 43 were reached). These are called “Reading Program” schools throughout this report, and their results are not generalizable to any other population. There was no counterfactual or control group for these schools. The majority of the analysis for this population focuses on performance at endline in 2014. However, a subset of Reading Program schools (20 out of 44) were assessed in both 2012 and 2014; with caveats in place about small sample sizes, a brief discussion of change in performance over time for those 20 schools is included as well. The findings of this study are presented in *Chapter B* of this report.

² The accessibility criterion specified that the schools had to be located within approximately 20 km of an urban center. For a subdivision to be eligible for inclusion in the sample population for the 2012 study which serves as the baseline for the Accessible PAQUED evaluation, there had to be a minimum of six intervention schools within the subdivision that were within 20 km of an urban center.

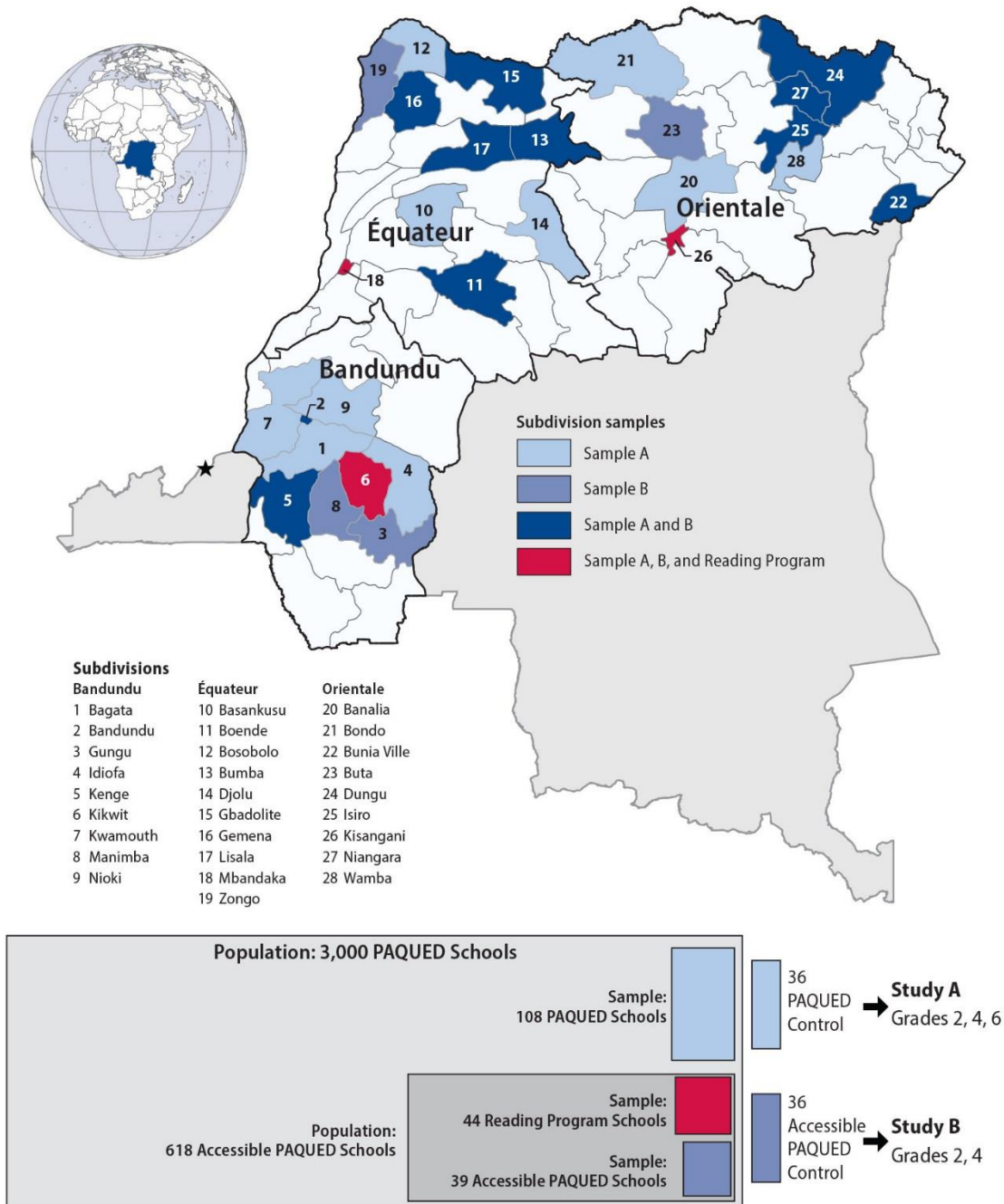
Table ESI below provides a visual representation of the studies whose results are presented in this report.

Table ES1. Study Populations and Analyses Permitted

Study	Treatment Groups	2010	2012	2014	Evaluation
PAQUED (Grades 2, 4, 6)	PAQUED Control	Baseline		Endline	Relative growth in performance in PAQUED vs. Control over the period from 2010-2014
Accessible PAQUED (Grades 2, 4)	Accessible PAQUED Accessible Control		Baseline	Endline	Relative growth in performance in Accessible PAQUED vs. Accessible Control over the period from 2010-2014
Reading Program (Grades 2, 4)	Reading Program			Snapshot	Performance in 2014

Figure ESI, on the next page, provides two visual representations of the school populations assessed during the 2014 endline Early Grade Reading Assessment (EGRA)/Early Grade Mathematics Assessment (EGMA). The top portion of the image includes a map of the educational subdivisions in which schools were assessed, which is color-coded according to the population(s) whose schools were located in that subdivision. The bottom of the image provides scale representation of the relative sizes of the populations being studied and the samples drawn (i.e., the ratio of the area of the light grey *Population: 3,000 PAQUED Schools* rectangle to the area of the dark grey *Population: 618 Accessible PAQUED Schools* is 3,000:618, or ~4.85:1.) In the map and visualization, Study A refers to the PAQUED and Control schools; Study B refers to the Accessible PAQUED and Accessible Control schools; and the Reading Program schools are nested within the Study B shape because they were a subset of the Accessible PAQUED schools.

Figure ES1. Subdivisions Containing Schools Assessed at Endline and Relative Size of the Study Populations



Limitations

The current report presents data and analyses that are subject to a few limitations. Some are due to the differences between the original (2009) vision for the project and the post-realignment (2013) vision; some are a function of how the DRC educational system’s standards have evolved since 2010, when adaptation of the EGRA instrument was

conducted; and others are due to the change in sampling design between baseline (2010) and midterm (2012).

Data Collection

Timing of Studies

Data for the 2014 Endline EGRA/EGMA was collected from May 1, 2014, through June 9, 2014, by teams of assessors using Tangerine® electronic data collection software. Teams spent an average of two days per school sampling 13 students per grade in Grades 2, 4, and 6 (the PAQUED schools study) or Grades 2 and 4 (the Accessible PAQUED and Reading Program studies). Teams also conducted interviews with the sampled students' classroom teachers and the school's headteacher.

Tools Used to Measure Impact

PAQUED's impact on student learning outcomes was measured using the EGRA and the EGMA.

The EGRA instrument is a 15- to 20-minute orally administered set of subtasks designed to assess foundational literacy skills crucial to becoming a fluent reader. The EGRA assessment is designed to be a method-independent approach to assessment, in that the instrument does not reflect a particular method of reading instruction (i.e., “whole language” or “phonics-based” approach). Rather, EGRA measures basic skills that a child must possess to eventually be able to read fluently and with comprehension—the ultimate goal of reading. The EGRA subtasks are based on research regarding a comprehensive approach to reading acquisition across languages, including five essential components: phonemic awareness, phonics, reading fluency, vocabulary, and comprehension.³

Similarly, the EGMA is an orally administered instrument that individually tests children and is designed to assess student performance on foundational mathematics competencies which they are expected to master in the early grades. The instrument includes subtasks which are designed to do the following:

- include key skills that developing country and developed country curricula have determined should be acquired in early grades;
- reflect those skills that are most predictive of future performance, according to available research and scientific advice;
- represent a progression of skills that lead toward proficiency in mathematics; and
- target both conceptual understanding and procedural fluency.

³ The definitions are adapted from U.S. Department of Health and Human Services, National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel—Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4769). Available at <http://www.nichd.nih.gov/publications/nrp/smallbook.cfm>

The particular subtasks included in the EGRA and EGMA used in a given context are determined through a consultative adaptation process. Subtasks are selected according to the skills that are of interest to the stakeholders involved in the study, the curricular expectations of the country, and the particular characteristics of the language(s) being assessed.⁴ *Table ES2* describes the subtasks included in the 2014 EGRA/EGMA tools.

Table ES2. Subtask Composition of Assessment Instruments

Instrument	Subtasks	Grade		
		2	4	6
EGRA	Vocabulary	X	X	X
	Initial Sound Identification	X	X	X
	Grapheme Recognition	X	X	X
	Familiar Word Reading		X	X
	Invented Word Reading		X	X
	Oral Reading Fluency		X	X
	Reading Comprehension		X	X
	Listening Comprehension	X	X	X
Dictation		X	X	
EGMA	Rote Counting	X	X	X
	Rational Counting	X	X	
	Number Identification	X	X	X
	Quantity Comparison	X	X	X
	Missing Number	X	X	X
	Word Problems	X	X	X
	Calculations			
	Addition	X	X	X
	Subtraction	X	X	X
	Multiplication		X	X

⁴ At the time of the adaptation of the tools in the DRC, the policy regarding the use of national languages was in question. In consultation with the donor and government counterparts, it was decided to use French for the project interventions and as the language of assessment for the EGRA instruments.

Instrument	Subtasks	Grade		
		2	4	6
	Division		X	X

Findings

Despite the complexity of the situation being examined, the findings can be summed up fairly simply: these studies do not permit us to make a positive statement regarding the characteristics of an intervention—whether in terms of its components or its intensity—required in order to realize meaningful improvements in student performance in reading or mathematics.⁵ However, these studies do allow us to state with confidence that the support provided to Accessible PAQUED and PAQUED schools was not sufficient to generate learning gains as measured by the DRC EGRA and EGMA instruments that are either significant or substantive.

Impact on Student Reading Performance

In the study of 3000 PAQUED schools, only two comparisons revealed statistically significant differences: (1) while Grade 4 male students in Control schools in Equateur initially outperformed their female classmates on the *Vocabulary* subtask in 2010, by 2014 the difference had disappeared, and (2) while Grade 6 male students in PAQUED schools had not outperformed their female classmates on the *Familiar Word Reading* subtask in 2010, by 2014 they were doing so. (While performance by both boys and girls appears to have decreased over time, girls' performance decreased much more, thus accounting for the difference.) There were no statistically significant differences in performance between students in PAQUED intervention schools and Control schools, whether cross-sectionally or over time. The intervention did not measurably improve student reading performance between the baseline assessment in 2010 and the endline assessment in 2014. Given that anecdotal sources and qualitative studies, such as the 2012 Midterm Report conducted by School-to-School International⁶, highlighted the complexities and challenges of operating in the DRC and raised serious questions about whether the intervention was actually implemented with any fidelity to speak of in these schools, this result was not surprising.

⁵ In 2013-2014 EDC carried out an evaluation of the growth over time in the content knowledge and pedagogical practices of teachers in PAQUED schools. Those teacher data were matched with student performance data collected using different instruments than those used for the RTI EGRA/EGMA studies. The report on that study suggests that the Reading Program intervention had a significant and positive effect on teacher pedagogical practice as well as some student learning outcomes. Please see Louge, N. (2014). *2014 final evaluation report teachers' literacy knowledge, instructional practices, and their students' reading performance in PAQUED supported schools in the Democratic Republic of Congo*. Education Development Center.

⁶ Lynd, M. (2012, September). *Final report: Mid-term review: Package for improving educational quality (PIEQ) project*. Washington, DC: USAID

The Accessible PAQUED study began in 2012 was based on a revised sampling frame drawn from the population of schools the project estimated would be most likely to have received the intervention's inputs. As was the case with the PAQUED schools, the study of Accessible PAQUED schools did not reveal any discernible impact on student reading performance. While an absence of robust data on fidelity of implementation cautions against making definitive statements, this suggests that the package of inputs that constituted the original PAQUED intervention was inadequate to the task of improving student reading, and that the lack of results in either of these first two study populations was not simply a matter of poor implementation.

The intensified, focused intervention delivered to the Reading Program schools in response to changing USAID and GDRRC priorities certainly bore the hallmarks of the sort of well-designed reading-improvement strategies that have demonstrated success elsewhere, i.e., books in the hands of students, including decodable and leveled texts; read-aloud books and teacher guides for the instructors; comprehensive training for teachers on effective reading strategies and the use of the materials provided; at least semi-regular coaching and mentoring by someone trained in reading pedagogy; and—crucially—adequate, daily instructional time focused on reading.

While scores on most reading subtasks appeared to improve slightly from 2012 to 2014, relatively few gains were shown to be statistically significant.⁷ Such statistical confirmation would be particularly valuable because the *Vocabulary*, *Initial Sound Identification*, *Grapheme Recognition*, *Familiar Word Reading*, *Invented Word Reading*, *Oral Reading Fluency*, *Reading Comprehension*, and *Dictation* subtasks—that is, every subtask save *Listening Comprehension*—included at least one instance of student performance in Reading Programs schools that appeared to decline, rather than improve, from 2012 to 2014.

Some evidence emerged from these studies that support for students appears to improve student performance. For instance, for students in Reading Program schools, having a teacher who engaged actively with the intervention (as measured by attendance at school- and cluster-level forums, participating in training, and receiving visits by PAQUED personnel trained in reading) correlated positively and significantly with being a high performer (top quintile of all students in the same grade in the same study) on the *Grapheme Recognition* subtask. It is encouraging that something over which the school can exert influence could have such an impact. (This is in contrast to other factors that also correlated with being a high performer, such as speaking predominantly French at home or having a family member at home who also knows how to read.)

The following section provides summaries of student performance in reading for each of the populations of interest. The first table in each case summarizes the change in

⁷ More insight on the outcomes of the intervention in the Reading Program schools can be gleaned from Louge, N. (2014). *2014 Final Evaluation Report: Teachers' Literacy Knowledge, Instructional Practices, and Their Students' Reading Performance in PAQUED Supported Schools in the Democratic Republic of Congo*. Education Development Center.

performance over time from the population’s baseline (2010 or 2012) to endline (2014). The second table compares student performance at endline to national benchmarks⁸ for the various reading subtasks, and the figure provides a visualization of the same.

Table ES3 provides a summary of student performance in reading for the PAQUED population, disaggregated by Province.

Table ES3. Reading Performance in PAQUED Schools, in Mean Scores, by Province

Province	Subtask	Group	Grade 2		Grade 4		Grade 6	
			2010	2014	2010	2014	2010	2014
Bandundu	Vocabulary	Control	6.45	5.69	8.59	8.58	11.73	12.30
		Treatment	6.15	6.12	8.47	8.53	11.22	11.78
	Initial Sound Identification	Control	0.82	0.90	3.12	1.96	4.08	4.18
		Treatment	0.64	0.64	2.23	1.70	3.66	2.74
	Grapheme Recognition	Control	4.34	5.98	15.53	14.44	32.03	32.87
		Treatment	3.57	3.49	14.18	14.01	29.54	30.16
	Familiar Word Reading	Control	-	-	8.33	2.05	16.38	14.15
		Treatment	-	-	4.85	2.55	14.42	14.68
	Invented Word Reading	Control	-	-	3.25	2.20	10.99	11.91
		Treatment	-	-	2.11	2.02	10.11	10.68
	Oral reading fluency	Control	-	-	5.40	4.32	16.55	24.43
		Treatment	-	-	4.41	5.52	15.50	24.72
	Reading Comprehension	Control	-	-	0.23	0.13	0.48	0.85
		Treatment	-	-	0.18	0.08	0.48	0.77
	Listening Comprehension	Control	1.04	0.96	0.91	0.66	1.39	1.65
		Treatment	0.79	0.35	0.86	0.74	1.29	1.57
Dictation	Control	-	-	2.03	0.59	2.97	1.39	
	Treatment	-	-	1.98	0.71	2.85	1.46	

⁸ Benchmarks for student performance in French do not exist for every skill in every grade. In some cases, a benchmark may exist for that skill, but in a national language that was not assessed by these studies.

Province	Subtask	Group	Grade 2		Grade 4		Grade 6	
			2010	2014	2010	2014	2010	2014
Equateur	Vocabulary	Control	5.53	7.22	8.43	9.67	10.92	12.53
		Treatment	6.21	8.40	8.78	9.94	11.12	11.97
	Initial Sound Identification	Control	0.33	1.55	2.94	2.84	4.17	5.03
		Treatment	1.09	1.30	3.55	2.68	4.63	3.81
	Grapheme Recognition	Control	3.77	4.04	16.21	19.98	33.08	42.42
		Treatment	5.39	5.49	20.05	17.36	36.11	31.66
	Familiar Word Reading	Control	-	-	6.86	4.76	18.48	23.73
		Treatment	-	-	9.06	4.55	20.04	13.77
	Invented Word Reading	Control	-	-	2.92	3.81	12.64	18.70
		Treatment	-	-	5.23	3.94	13.74	10.46
	Oral reading fluency	Control	-	-	5.59	10.46	24.64	40.75
		Treatment	-	-	8.77	9.51	25.40	24.42
	Reading Comprehension	Control	-	-	0.40	0.15	1.04	0.87
		Treatment	-	-	0.63	0.22	1.08	0.53
	Listening Comprehension	Control	0.25	0.44	0.45	0.68	0.90	1.55
		Treatment	0.24	0.72	0.52	0.90	0.90	1.50
Dictation	Control	-	-	1.91	0.97	2.76	1.93	
	Treatment	-	-	2.03	0.97	2.83	1.32	
Orientale	Vocabulary	Control	7.25	7.72	10.91	11.20	13.89	13.39
		Treatment	7.08	7.72	10.34	10.59	14.75	13.78
	Initial Sound Identification	Control	0.37	0.34	1.16	0.43	1.87	1.12
		Treatment	0.40	0.24	1.14	0.81	2.39	1.66
	Grapheme Recognition	Control	3.53	5.34	13.56	19.52	37.10	36.58
		Treatment	4.15	3.49	17.67	16.62	39.23	37.78
	Familiar Word Reading	Control	-	-	6.75	6.59	23.27	19.67

Province	Subtask	Group	Grade 2		Grade 4		Grade 6	
			2010	2014	2010	2014	2010	2014
		Treatment	-	-	12.61	3.94	27.26	20.14
	Invented Word Reading	Control	-	-	3.79	5.42	16.74	15.24
		Treatment	-	-	6.72	3.54	21.29	15.66
	Oral reading fluency	Control	-	-	6.72	12.09	30.78	33.16
		Treatment	-	-	11.20	8.41	39.42	34.63
	Reading Comprehension	Control	-	-	0.22	0.48	0.51	0.93
		Treatment	-	-	0.44	0.09	0.69	1.00
	Listening Comprehension	Control	0.61	0.37	0.84	1.27	1.74	2.02
		Treatment	0.83	0.48	1.17	1.02	1.98	2.00
	Dictation	Control	-	-	1.51	0.81	2.90	1.42
		Treatment	-	-	2.07	0.68	3.23	1.60

Table ES4 compares student performance at endline to the appropriate national benchmarks, and *Figure ES2* illustrates the same content graphically.

Table ES4. Student Performance in PAQUED Schools Relative to National Benchmarks⁹

Grade	Reading or Pre-reading Skill	Below benchmark		Within benchmark range		Above benchmark range	
		<i>n</i>	(%)	<i>N</i>	(%)	<i>n</i>	(%)
	Listening Comprehension	-	-	-	-	-	-
2	Phonemic Awareness	1647	(92.1%)	50	(2.6%)	98	(5.4%)
	Graphemes	1769	(98.3%)	9	(0.5%)	17	(1.2%)
	Familiar Words	-	-	-	-	-	-
	Oral Reading Fluency	-	-	-	-	-	-

⁹ The *n* for each grade presented in this table is as follows: Grade 2 = 1795; Grade 4 = 1745; Grade 6 = 1715. Each *n* presented is unweighted, and each percentage presented is weighted.

Grade	Reading or Pre-reading Skill	Below benchmark		Within benchmark range		Above benchmark range	
		<i>n</i>	(%)	<i>N</i>	(%)	<i>n</i>	(%)
4	Comprehension	-	-	-	-	-	-
	Listening Comprehension	1541	(87.5%)	96	(5.6%)	108	(6.9%)
	Phonemic Awareness	-	-	-	-	-	-
	Graphemes	1602	(89.6%)	88	(6.6%)	55	(3.8%)
	Familiar Words	1715	(98.5%)	5	(0.1%)	19	(1%)
	Oral Reading Fluency	1597	(88.3%)	73	(6.8%)	75	(4.9%)
	Comprehension	1688	(95.9%)	23	(2%)	34	(2.1%)
6	Listening Comprehension	1277	(69.9%)	201	(12.3%)	237	(17.8%)
	Phonemic Awareness	-	-	-	-	-	-
	Graphemes	-	-	-	-	-	-
	Familiar Words	-	-	-	-	-	-
	Oral Reading Fluency	1507	(80.4%)	111	(8.8%)	97	(10.8%)
	Comprehension	1551	(86.7%)	92	(7.2%)	72	(6%)

Figure ES2. Performance of Students in PAQUED Schools Relative to DRC Benchmarks

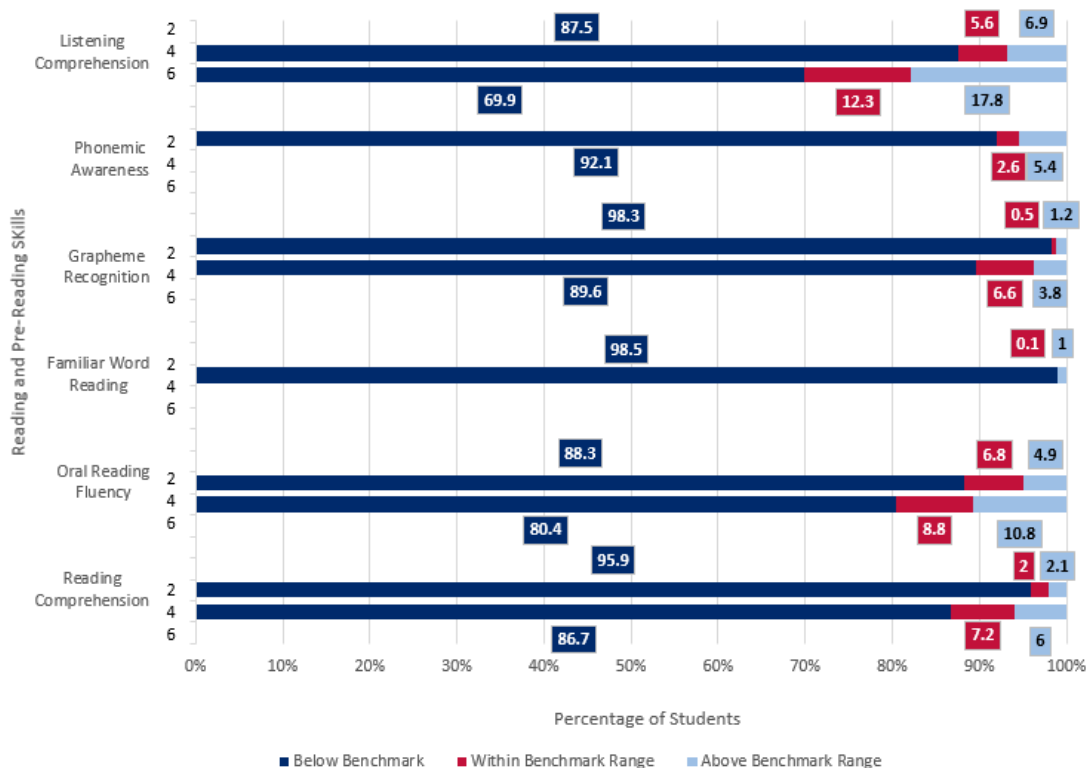


Table ES5 provides a summary of student performance in reading for the Accessible PAQUED population, disaggregated by Province.

Table ES5. Reading Performance in Accessible PAQUED Schools, in Mean Scores, by Province

Province	Subtask	Group	Grade 2		Grade 4	
			2010	2014	2010	2014
Bandundu	Vocabulary	Control	6.90	7.34	9.36	10.03
		Treatment	7.75	9.23	10.70	12.23
	Initial Sound Identification	Control	1.03	1.20	2.00	2.22
		Treatment	1.09	1.32	2.84	3.20
	Grapheme Recognition	Control	6.45	5.42	17.60	13.80
		Treatment	8.73	9.26	19.98	26.09
	Familiar Word Reading	Control	-	-	4.38	2.56

Province	Subtask	Group	Grade 2		Grade 4	
			2010	2014	2010	2014
Equateur		Treatment	-	-	6.84	7.25
		Control	-	-	3.39	1.99
	Invented Word Reading	Treatment	-	-	5.20	5.55
		Control	-	-	3.86	4.28
	Oral reading fluency	Treatment	-	-	8.20	12.88
		Control	-	-	0.13	0.06
	Reading Comprehension	Treatment	-	-	0.17	0.34
		Control	0.37	0.64	0.61	0.95
	Listening Comprehension	Treatment	0.41	1.37	0.85	1.96
		Control	-	-	0.57	0.79
	Dictation	Treatment	-	-	0.80	0.94
		Control	8.76	9.00	11.24	11.43
	Vocabulary	Treatment	8.72	8.74	10.35	11.38
		Control	1.66	2.73	3.43	3.40
Initial Sound Identification	Treatment	1.66	2.79	3.54	4.31	
	Control	6.74	11.34	24.79	23.17	
Grapheme Recognition	Treatment	6.10	8.65	20.29	24.42	
	Control	-	-	8.67	7.90	
Familiar Word Reading	Treatment	-	-	8.50	7.23	
	Control	-	-	7.53	10.65	
Invented Word Reading	Treatment	-	-	6.33	5.90	
	Control	-	-	15.72	12.73	
Oral reading fluency	Treatment	-	-	10.09	12.68	
	Control	-	-	0.63	0.41	
Reading Comprehension	Treatment	-	-	0.35	0.30	

Province	Subtask	Group	Grade 2		Grade 4	
			2010	2014	2010	2014
Orientale	Listening Comprehension	Control	0.49	1.03	1.49	1.18
		Treatment	0.72	0.61	1.09	1.05
	Dictation	Control	-	-	1.09	0.92
		Treatment	-	-	1.23	0.99
	Vocabulary	Control	6.73	8.32	10.00	12.31
		Treatment	9.21	10.09	12.44	12.37
	Initial Sound Identification	Control	0.70	0.11	1.50	0.95
		Treatment	1.03	0.50	1.81	0.83
	Grapheme Recognition	Control	6.64	6.07	22.65	26.77
		Treatment	7.41	9.16	25.18	27.39
	Familiar Word Reading	Control	-	-	10.93	11.00
		Treatment	-	-	12.73	13.36
	Invented Word Reading	Control	-	-	8.38	9.21
		Treatment	-	-	9.48	10.47
Oral reading fluency	Control	-	-	13.25	18.42	
	Treatment	-	-	16.89	21.42	
Reading Comprehension	Control	-	-	0.34	0.39	
	Treatment	-	-	0.59	0.68	
Listening Comprehension	Control	0.21	0.92	0.78	1.80	
	Treatment	0.71	0.87	1.75	2.17	
Dictation	Control	-	-	1.30	1.30	
	Treatment	-	-	1.23	1.14	

Table ES6 compares student reading performance to the appropriate national benchmarks, and **Figure ES3** which illustrates the same content graphically.

Table ES6. Student Performance in Accessible PAQUED Schools Relative to National Benchmarks¹⁰

Grade	Reading or Pre-reading Skill	Below benchmark		Within benchmark range		Above benchmark range	
		<i>n</i>	(%)	<i>N</i>	(%)	<i>n</i>	(%)
2	Listening Comprehension	-	-	-	-	-	-
	Phonemic Awareness	897	(89.2%)	36	(3.1%)	61	(7.7%)
	Graphemes	978	(98.3%)	6	(1%)	10	(0.7%)
	Familiar Words	-	-	-	-	-	-
	Oral Reading Fluency	-	-	-	-	-	-
	Comprehension	-	-	-	-	-	-
4	Listening Comprehension	769	(68.8%)	70	(10.1%)	132	(21%)
	Phonemic Awareness	-	-	-	-	-	-
	Graphemes	812	(74.3%)	113	(19.4%)	46	(6.3%)
	Familiar Words	940	(95.5%)	9	(1.1%)	12	(2.3%)
	Oral Reading Fluency	813	(79.1%)	81	(10.4%)	77	(10.5%)
	Comprehension	935	(92.4%)	16	(3.5%)	20	(4%)

¹⁰ The *n* for each grade presented in this table is as follows: Grade 2, 994; Grade 4, 971. Each *n* presented is unweighted, and each percentage presented is weighted.

Figure ES3. Performance of Students in Accessible PAQUED Schools Relative to DRC Benchmarks

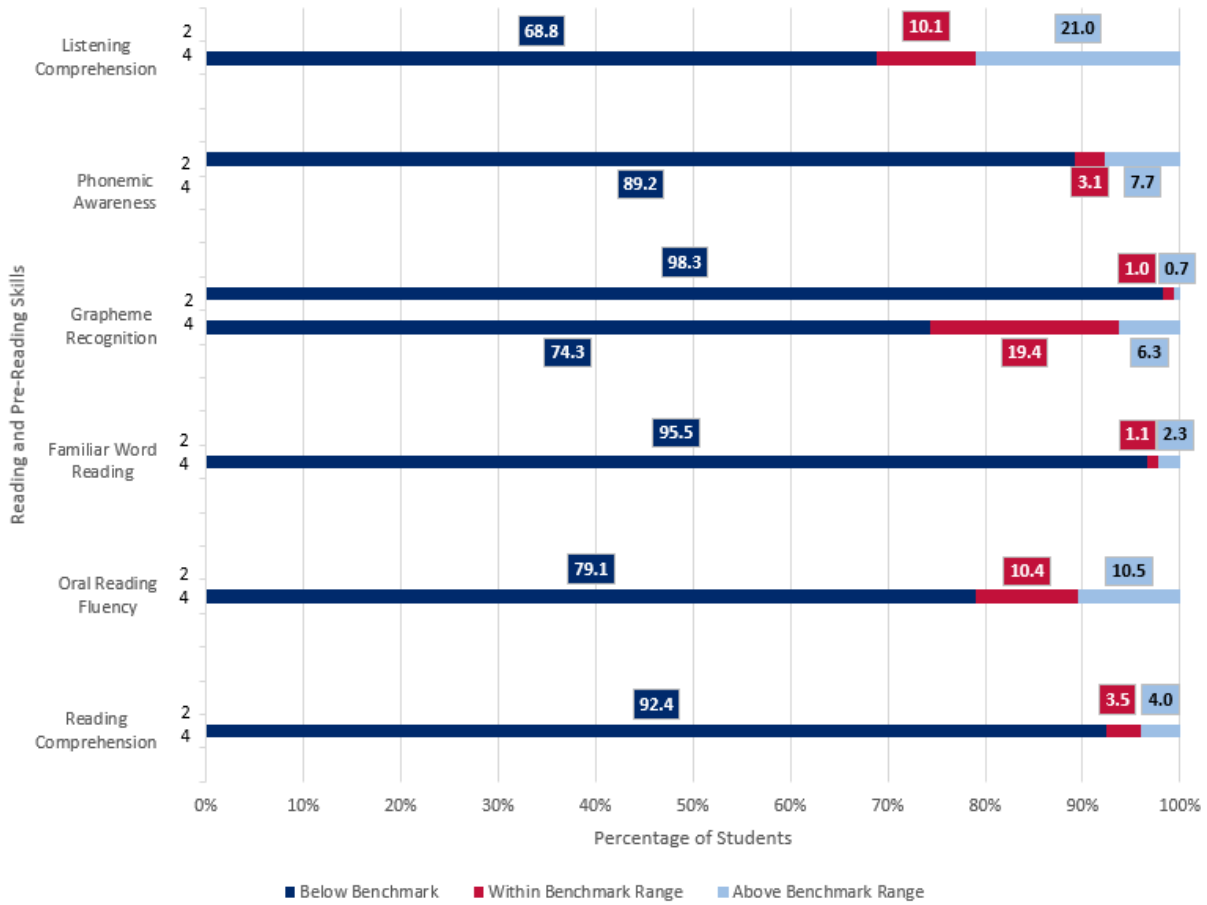


Table ES7 provides a summary of student performance in reading for the Reading Program population, disaggregated by Province.

Table ES7. Reading Performance in Subset of Reading Program Schools (n=20), in Mean Scores¹¹

Subtask	Group	Grade 2		Grade 4	
		2012	2014	2012	2014
Vocabulary	Control	-	-	-	-
	Treatment	8.21	9.34 *	11.08	12.07
Initial Sound Identification	Control	-	-	-	-
	Treatment	1.27	1.47	2.73	1.91
Grapheme Recognition	Control	-	-	-	-
	Treatment	8.16	16.02 *	21.26	21.91
Familiar Word Reading	Control	-	-	-	-
	Treatment	-	-	8.67	6.74
Invented Word Reading	Control	-	-	-	-
	Treatment	-	-	6.57	4.81
Oral reading fluency	Control	-	-	-	-
	Treatment	-	-	11.09	13.74
Reading Comprehension	Control	-	-	-	-
	Treatment	-	-	0.32	0.37
Listening Comprehension	Control	-	-	-	-
	Treatment	0.66	1.12 *	1.2	1.83
Dictation	Control	-	-	-	-
	Treatment	-	-	1.01	0.96

* significant at $p < 0.017$

Table ES8 compares student reading performance at endline to the appropriate national benchmarks, and **Figure ES4** illustrates the same content graphically.

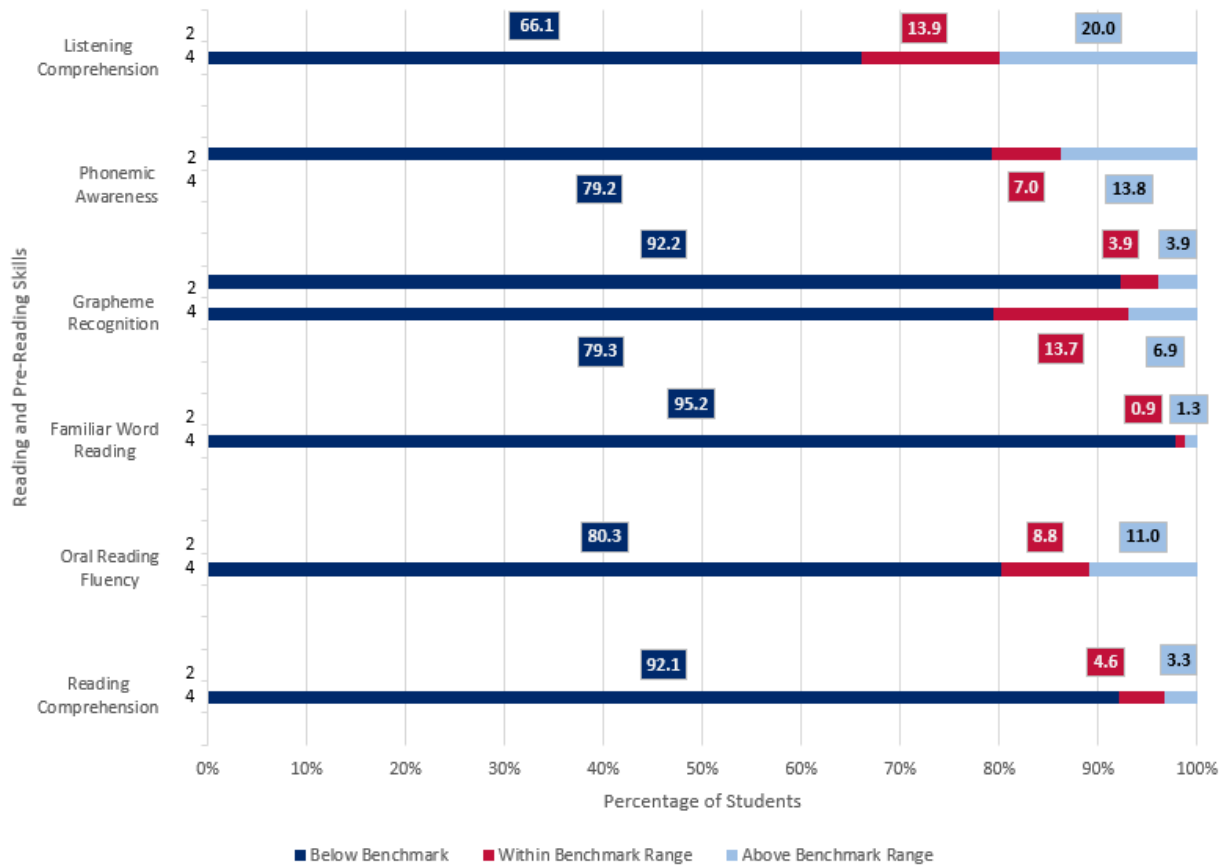
¹¹ There were no control schools against which performance in Reading Program schools could be compared. Because the number of schools and student records is so small, this table does not disaggregate results by province.

Table ES8. Student Performance in Reading Program Schools Relative to National Benchmarks¹²

Grade	Reading or Pre-reading Skill	Below benchmark		Within benchmark range		Above benchmark range	
		<i>n</i>	(%)	<i>n</i>	(%)	<i>n</i>	(%)
2	Listening Comprehension	-	-	-	-	-	-
	Phonemic Awareness	437	(79.2%)	39	(7%)	78	(13.8%)
	Graphemes	510	(92.2%)	21	(3.9%)	23	(3.9%)
	Familiar Words	-	-	-	-	-	-
	Oral Reading Fluency	-	-	-	-	-	-
	Comprehension	-	-	-	-	-	-
4	Listening Comprehension	368	(66.1%)	72	(13.9%)	109	(20%)
	Phonemic Awareness	-	-	-	-	-	-
	Graphemes	433	(79.3%)	72	(13.7%)	44	(6.9%)
	Familiar Words	521	(95.2%)	6	(0.9%)	8	(1.3%)
	Oral Reading Fluency	440	(80.3%)	46	(8.8%)	63	(11%)
	Comprehension	504	(92.1%)	25	(4.6%)	20	(3.3%)

¹² The *n* for each grade presented in this table is as follows: Grade 2 = 554; Grade 4 = 549. Each *n* presented is unweighted, and each percentage presented is weighted.

Figure ES4. Performance of Students in Reading Program Schools Relative to DRC Benchmarks



Ultimately, as *Tables ES3-ES8* and *Figures ES2-ES4* above illustrate, overall student performance in 2014 was below what was needed to meet national benchmarks across all subtasks. Even for oral skills such as vocabulary, phonemic awareness, and listening comprehension, students failed to demonstrate the French oral skills required to effectively read in French. Student mean scores on grapheme and word recognition, as well as connected text reading, were also lower than necessary for reading with fluency and comprehension.

Discussion of Mathematics Outcomes

Few significant differences were seen, and the apparent non-significant trend suggested that Control students often outperformed PAQUED students and the Accessible PAQUED students sometimes outperformed the Accessible Control students. The strong conclusion is that the interventions delivered to PAQUED, Accessible PAQUED, and Reading Program schools had no meaningful impact on student performance in mathematics. *Table ES9* provides a summary of the change in student performance over time from 2010-2014 in PAQUED schools.

Table ES9. Change in Mathematics Performance in PAQUED Schools over Time, in Percentage of Items Correct, by Province¹³

Province	Subtask	Group	Grade 2		Grade 4		Grade 6	
			2010	2014	2010	2014	2010	2014
Bandundu	Number Identification	Control	33.6%	43.4%	52.9%	49.3%	59.6%	64.9%
		Treatment	30.4%	40.6%	55.3%	55.5%	53.3%	66.0%
	Quantity Comparison	Control	61.2%	45.7%	62.7%	54.4%	35.5%	39.7%
		Treatment	57.1%	45.7%	60.3%	57.0%	34.7%	38.5%
	Missing Number	Control	17.9%	16.8%	23.7%	35.5%	22.3%	42.6%
		Treatment	17.8%	16.2%	30.0%	33.1%	22.9%	38.2%
	Word Problems	Control	44.8%	30.9%	50.3%	41.5%	62.6%	55.7%
		Treatment	37.6%	31.2%	51.0%	35.6%	60.8%	51.8%
	Calculations: Addition	Control	40.0%	37.8%	28.0%	34.1%	47.2%	60.0%
		Treatment	33.9%	32.1%	33.8%	40.8%	37.8%	56.9%
	Calculations: Subtraction	Control	38.5%	26.1%	37.4%	25.2%	33.4%	45.5%
		Treatment	32.3%	25.4%	42.3%	36.1%	37.3%	42.1%
	Calculations: Multiplication	Control	-	-	21.1%	13.5%	16.3%	27.6%
		Treatment	-	-	24.8%	17.1%	16.6%	19.2%
Calculations: Division	Control	-	-	21.5%	7.5%	21.2%	28.4%	
	Treatment	-	-	28.7%	14.2%	19.9%	21.7%	
Equateur	Number Identification	Control	29.3%	36.2%	56.6%	62.2%	59.5%	67.7%
		Treatment	38.0%	45.8%	61.3%	57.2%	59.8%	64.9%
	Quantity Comparison	Control	55.0%	49.5%	61.2%	63.0%	34.1%	46.1%
		Treatment	60.0%	62.7%	64.6%	61.9%	33.6%	44.0%
	Missing Number	Control	8.8%	18.2%	15.0%	37.1%	13.8%	43.1%
		Treatment	11.9%	24.3%	18.9%	38.5%	12.1%	40.3%

¹³ The Rote Counting and Rational Counting tasks do not lend themselves to meaningful reporting of means, and are therefore omitted from this report. Results can be obtained by requesting the datasets.

Province	Subtask	Group	Grade 2		Grade 4		Grade 6		
			2010	2014	2010	2014	2010	2014	
	Word Problems	Control	29.4%	30.5%	38.4%	49.8%	56.4%	59.6%	
		Treatment	34.7%	43.0%	44.6%	50.0%	56.5%	59.1%	
	Calculations: Addition	Control	32.5%	35.4%	31.2%	53.3%	45.5%	62.8%	
		Treatment	37.9%	49.0%	32.9%	44.4%	40.1%	57.2%	
	Calculations: Subtraction	Control	23.8%	24.2%	39.1%	48.6%	43.1%	53.3%	
		Treatment	28.8%	41.5%	45.8%	41.5%	46.2%	48.8%	
	Calculations: Multiplication	Control	-	-	22.0%	30.3%	22.1%	32.6%	
		Treatment	-	-	20.9%	34.0%	23.6%	31.9%	
	Calculations: Division	Control	-	-	14.9%	32.9%	18.4%	46.9%	
		Treatment	-	-	20.4%	35.0%	19.4%	36.9%	
	Orientale	Number Identification	Control	37.0%	47.6%	62.1%	66.3%	69.2%	70.1%
			Treatment	48.7%	41.2%	67.8%	63.4%	71.7%	72.2%
		Quantity Comparison	Control	52.9%	64.6%	66.7%	70.9%	43.2%	51.1%
			Treatment	64.0%	67.7%	71.4%	72.0%	47.4%	48.8%
		Missing Number	Control	15.9%	20.5%	26.3%	44.3%	25.8%	46.2%
			Treatment	19.1%	19.8%	38.4%	47.6%	38.6%	47.5%
		Word Problems	Control	29.8%	24.9%	39.3%	33.8%	68.4%	46.4%
			Treatment	40.4%	27.2%	47.5%	36.1%	71.1%	49.3%
Calculations: Addition		Control	33.7%	46.8%	34.9%	56.1%	57.6%	73.6%	
		Treatment	50.5%	46.7%	47.7%	50.2%	61.9%	76.5%	
Calculations: Subtraction		Control	26.8%	32.3%	41.7%	53.5%	45.9%	64.3%	
		Treatment	44.2%	34.4%	54.7%	49.1%	63.6%	69.5%	
Calculations: Multiplication		Control	-	-	17.3%	41.8%	24.7%	41.0%	
		Treatment	-	-	29.3%	30.8%	42.7%	46.4%	
			Control	-	-	17.2%	39.1%	25.0%	56.1%

Province	Subtask	Group	Grade 2		Grade 4		Grade 6	
			2010	2014	2010	2014	2010	2014
	Calculations: Division	Treatment	-	-	31.4%	28.0%	35.1%	53.4%

Table ES10 provides a summary of the change in student performance over time from 2012-2014 in Accessible PAQUED schools.

Table ES10. Change in Mathematics Performance in Accessible PAQUED Schools over Time, in Percentage of Items Correct, by Province¹⁴

Province	Subtask	Group	Grade 2		Grade 4	
			2012	2014	2012	2014
Bandundu	Number Identification	Control	45.20%	48.30%	60.10%	56.80%
		Treatment	49.40%	56.30%	71.20%	70.30%
	Quantity Comparison	Control	47.70%	62.90%	54.30%	58.50%
		Treatment	50.20%	66.10%	60.50%	67.80%
	Missing Number	Control	17.10%	23.80%	26.20%	39.20%
		Treatment	15.70%	24.90%	31.30%	42.60%
	Word Problems	Control	28.30%	22.70%	56.10%	42.30%
		Treatment	29.00%	29.80%	59.30%	41.00%
	Calculations: Addition	Control	49.50%	45.30%	30.90%	40.30%
		Treatment	49.00%	52.70%	44.60%	58.90%
	Calculations: Subtraction	Control	43.60%	31.40%	29.20%	42.10%
		Treatment	45.00%	42.00%	36.50%	51.80%
	Calculations: Multiplication	Control	-	-	14.30%	23.70%
		Treatment	-	-	18.00%	33.90%
	Calculations: Division	Control	-	-	18.50%	31.50%
		Treatment	-	-	22.00%	21.50%

¹⁴ The Rote Counting and Rational Counting tasks do not lend themselves to meaningful reporting of means, and are therefore omitted from this report. Results can be obtained by requesting the datasets.

Province	Subtask	Group	Grade 2		Grade 4	
			2012	2014	2012	2014
Equateur	Number Identification	Control	43.10%	49.50%	67.00%	60.30%
		Treatment	44.60%	53.00%	65.70%	62.50%
	Quantity Comparison	Control	56.50%	64.90%	60.30%	63.00%
		Treatment	54.50%	62.50%	64.40%	66.80%
	Missing Number	Control	15.70%	34.90%	31.60%	44.10%
		Treatment	18.20%	28.30%	29.40%	45.70%
	Word Problems	Control	30.30%	29.50%	55.90%	50.20%
		Treatment	26.10%	33.80%	44.50%	51.80%
	Calculations: Addition	Control	47.40%	57.90%	44.10%	49.10%
		Treatment	42.40%	47.30%	39.50%	54.20%
	Calculations: Subtraction	Control	35.10%	47.90%	51.70%	50.10%
		Treatment	34.80%	40.90%	43.10%	45.90%
	Calculations: Multiplication	Control	-	-	26.20%	44.40%
		Treatment	-	-	22.50%	35.70%
Calculations: Division	Control	-	-	36.20%	38.40%	
	Treatment	-	-	26.30%	30.10%	
Orientale	Number Identification	Control	51.30%	56.50%	77.50%	73.30%
		Treatment	57.20%	57.60%	77.50%	70.50%
	Quantity Comparison	Control	56.30%	69.10%	71.30%	72.80%
		Treatment	73.60%	71.70%	75.50%	77.00%
	Missing Number	Control	16.10%	22.60%	33.20%	47.10%
		Treatment	20.60%	25.40%	37.70%	49.10%
	Word Problems	Control	23.40%	17.00%	49.60%	36.80%
		Treatment	22.10%	15.60%	48.90%	28.50%
		Control	47.70%	60.50%	53.30%	68.40%

Province	Subtask	Group	Grade 2		Grade 4	
			2012	2014	2012	2014
	Calculations: Addition	Treatment	58.60%	62.50%	53.70%	59.20%
	Calculations: Subtraction	Control	22.90%	41.10%	58.40%	72.50%
		Treatment	35.60%	53.80%	50.10%	58.50%
	Calculations: Multiplication	Control	-	-	35.40%	45.00%
		Treatment	-	-	36.60%	43.60%
	Calculations: Division	Control	-	-	42.50%	55.00%
		Treatment	-	-	43.70%	43.40%

Table ES11 provides a summary of the change in student performance over time from 2012-2014 in the subset ($n = 20$) of Reading Program schools tested at both points.

Table ES11. Change in Mathematics Performance in Reading Program Schools ($n = 20$) over Time, in Percentage of Items Correct¹⁵

Subtask	Group	Grade 2		Grade 4	
		2012	2014	2012	2014
Number Identification	Treatment	48.05%	52.35%	70.42%	63.65% *
Quantity Comparison	Treatment	58.82%	60.12%	62.24%	65.09%
Missing Number	Treatment	16.22%	23.12% *	30.36%	42.78% *
Word Problems	Treatment	42.18%	29.48% *	58.40%	38.67% *
Calculations: Addition	Treatment	46.91%	49.80%	43.84%	52.99% *
Calculations: Subtraction	Treatment	31.90%	36.61%	45.28%	48.84%
Calculations: Multiplication	Treatment			23.80%	35.52% *

¹⁵ There were no control schools against which performance in Reading Program schools could be compared. Because the number of schools and student records is so small, this table does not disaggregate results by province.

Subtask	Group	Grade 2		Grade 4	
		2012	2014	2012	2014
Calculations: Division	Treatment			31.44%	36.55%

* significant at $p < 0.017$

Recommendations

The absence of a clear narrative of student reading or math performance driven by the PAQUED interventions—whether the original package of inputs or the modified package provided to Reading Program schools—precludes issuing definitive recommendations for the improvement of student learning outcomes in the DRC.

However, it is important to bear in mind that the research design that concluded with the 2014 Endline EGRA and EGMA never foresaw the realignment of the project to focus on Reading Program schools. Because this realignment was concluded following the 2012 EGRA/EGMA data collection (which because of PAQUED focus on Accessible schools actually served as a second baseline rather than a true midterm), it was not possible to establish a true baseline or true counterfactuals for the Reading Program. As a result, the current Reading Program study was severely limited in the conclusions it could present.¹⁶

Therefore, if (as is suggested by the national Reading Roadmap, the *feuille de route*) the DRC education community wants discussions about the future path of reading instruction in the Congo to be concretely informed by *quantitative* data on the viability of the Reading Program’s approach, designing a research study explicitly for that purpose would be an important next step.¹⁷

The above notwithstanding, there are elements of the Reading Program intervention that appear promising and are in line with both the MEPSP’s vision and strategic documents and also with approaches that some USAID implementers have successfully applied elsewhere (including in Kenya, Liberia, Malawi, Egypt, and Jordan). Provision of books to students—including decodable and leveled readers—that are aligned with teacher guides that focus explicitly on the teaching of reading is one such element. Another, is

¹⁶ EDC designed and conducted a study in the Reading Program schools that examined changes in teacher knowledge, attitudes, and practice as a result of the modified intervention, and the impact of those changes on student performance. Please see Louge, N. (2014). *2014 final evaluation report teachers’ literacy knowledge, instructional practices, and their students’ reading performance in PAQUED supported schools in the Democratic Republic of Congo*. Education Development Center.

¹⁷ A team led by Professor Pierre Mukendi of the *Université de Kinshasa* (UNIKIN), a member of the National Reading Commission, completed a *qualitative* endline study of the PAQUED project. The report his team produced, and the Lessons Learned Workshop that was convened in August 2014 to discuss both its findings and the preliminary analyses of the 2014 endline EGRA/EGMA studies, provide valuable insights that go beyond what the data that are the subject of this report can state. To request a copy of Prof. Mukendi’s report please contact Susan Ross at sross@edc.org.

the incorporation of content on reading pedagogy into in-service teacher training modules (such as the school- and cluster-level *forums d'échange*) is another. Finally, providing teachers with regular, ongoing support by coaches who have been trained in reading pedagogy is a third. That the composite regression model of teacher participation demonstrated a positive and significant correlation with student performance on the *Grapheme Recognition* subtask is an encouraging early sign of the Reading Program's potential.

Introduction

Funded by the U.S. Agency for International Development (USAID) and led by the Education Development Center (EDC), the *Projet d'Amélioration de la Qualité de l'Éducation* (PAQUED) project was a five-year initiative that aimed to raise student learning through improved teaching and school environments in the Democratic Republic of the Congo (DRC). As a partner in PAQUED, RTI International (RTI) was responsible for assessing the impact of the program using EGRA and EGMA instruments adapted for the DRC context.

This document presents a discussion of PAQUED's impact on three distinct populations of schools; each reflects a different stage in the significant evolution of PAQUED's focus and scope that took place between project design in 2009 and a formal realignment that occurred in January 2013. These different populations cannot be compared directly to each other, so each is dealt with in a separate section of the report.¹⁸ This introduction provides a chronological overview of the impact evaluation activities that took place from 2010 to 2014; however, the body of the report itself presents the results in reverse chronological order, reflecting the relative level of importance accorded to each population as the project was gradually brought into alignment with USAID's reading strategy.

Brief Chronological History of Impact Evaluation Activities under PAQUED

In September and October of 2010, as part of the PAQUED project, RTI collaborated with the DRC's Ministry of Primary, Secondary, and Professional Education (*Ministère de l'Enseignement Primaire, Secondaire, et Professionnel* [MEPSP]) to conduct a baseline assessment of the reading and mathematics skills of students enrolled in Grades 2, 4, and 6 in three provinces participating in this project: Bandundu, Equateur, and Orientale. The instruments used in the 2010 assessment were RTI's Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA). The instruments are designed to collect information on the level of competency in foundational reading and mathematical skills areas. All of the competencies that each instrument measures have been shown through research to be highly predictive of later reading and mathematical ability and susceptible to improvement through teaching. (For more information on the structure and contents of the EGRA and EGMA instruments, please see the *Research Design* section of this report and *Annex III: About the Assessment Instruments*.)

The project's initial target population included 3,000 primary schools, divided evenly across the three provinces; as a result the sample for the 2010 study was designed to be representative at the provincial level. When changes in performance from 2010 to 2014

¹⁸ Please see *Annex I* for a detailed discussion of the sampling approaches and why they must be treated separately.

for this largest of target populations are discussed in this report, the schools will be referred to as the “PAQUED schools.”¹⁹

In May 2012, near the midpoint of the PAQUED program, a second EGRA/EGMA study was conducted with the goal of understanding the impact of the PAQUED intervention after two years of implementation. Numerous technological and logistical problems delayed and impeded implementation of planned interventions, however, with the result that very few schools received the planned package of services and support. As a result, EDC preferred that the assessment’s sampling frame be modified to focus on those schools that, by virtue of being “accessible” (defined loosely as being within approximately 20 km of an urban center), might be more likely to have received the intervention as planned and might, therefore, present a more accurate portrait of the intervention’s potential impact.

A critical practical result of changing the sampling frame was that data were no longer comparable between 2010 and 2012. The Midterm EGRA and EGMA Reports produced in 2012, therefore, focused exclusively on describing differences between treatment schools and control schools, without attempting to attribute those changes conclusively to the intervention.²⁰

The new population frame defined by incorporating the new accessibility criterion contained 618 schools out of PAQUED’s original 3,000 schools. When changes in performance from 2012 to 2014 for this intermediate target population are discussed in this report, the schools will be referred to as the “Accessible PAQUED schools.”

As a result of the PAQUED mid-term review and in response to the recommendation to realign the project with the new USAID Education strategy, in January 2013 PAQUED underwent a formal project realignment that significantly diminished the size of the priority target population and shifted the scope to focus more narrowly on reading (see *Chapter A, Realignment of PAQUED’s Focus for 2012–2014* for a fuller discussion of this process and the reasons behind it). The realigned project intervention targeted 45 schools located in the *grands centres* of Kikwit (Bandundu Province), Mbandaka (Equateur Province), and Kisangani (Orientale Province). While the Reading Program schools all meet the 2012 criteria for accessibility, and they therefore represent a subset of the Accessible PAQUED population, they were selected purposively. As a result, they

¹⁹ A full discussion of the 2010 baseline assessment can be found in the EGRA and EGMA baseline reports, accessible at <https://www.eddataglobal.org/countries/index.cfm?fuseaction=showdir&pubcountry=CD&statusID=3&shotypes=0>

²⁰ A full discussion of the 2012 midterm assessment can be found in the EGRA and EGMA midterm reports, accessible at <https://www.eddataglobal.org/countries/index.cfm?fuseaction=showdir&pubcountry=CD&statusID=3&shotypes=0>

For an analysis of the comparability of the 2010 and 2012 samples, please see Davies, C. (2013). *Projet d’Amélioration de la Qualité de l’Éducation (PAQUED) 2010 comparison: Reduced sample versus full sample*. Washington, DC: USAID. Available from: <https://www.eddataglobal.org/documents/index.cfm?fuseaction=pubDetail&id=667>.

are a non-random sample and results from those schools cannot be generalized to the Accessible PAQUED population. When performance results for this target population are discussed in this report, the schools will be referred to as “Reading Program schools.”

In May 2014, RTI collected a third set of EGRA/EGMA data. Two separate samples were drawn, the first of which would permit comparisons with the 2010 baseline data (PAQUED schools) and the second of which would permit comparisons to the 2012 data (Accessible PAQUED and Reading Program schools).²¹

Structure of the Report

This report is presented in seven chapters. *Chapter A* focuses on briefly describing the project’s context and the evolution of its scope and alignment over time, as well as more fully describing the study’s research design, objectives, and limitations. *Chapter B* presents an analysis of student results from the Reading Program schools. *Chapters C* and *D* present analyses of student reading performance for the Accessible PAQUED and PAQUED schools. *Chapters E* and *F* presents analyses of student mathematics performance for the Accessible PAQUED and PAQUED schools. *Chapter G* presents conclusions and recommendations informed by both the foregoing analyses and a four-day workshop conducted in August 2014 with the National Reading Commission to review and explore the data.

Analytical *Chapters B-D* will contain descriptive statistics for each of the samples; student, teacher, and head teacher characteristics drawn from context interviews; student performance at endline (2014); change in student performance over time (from 2012 to 2014 or 2010 to 2014, as appropriate); and a selection of significant regressions and correlations.

The report’s annexes more detail regarding the sampling and analytical methods applied, the instruments, and the data collection process itself.

²¹ In February 2013, EDC conducted a baseline assessment of teacher pedagogical knowledge and classroom practices in reading and writing in the Reading Program schools. The endline assessment for this study, conducted in May and June 2014, included qualitative focus group elements, classroom observation tools, and teacher interview protocols and was informed by data on teachers’ fidelity of implementation of the reading program. In response to a request from USAID to generate performance data for Grade 2 children on the reading of connected text, EDC added a student assessment element to their study. The student assessment tool included untimed identification of 26 letter sounds; untimed reading of 8 high-frequency French words; and timed reading of a 26-word passage of connected text. For more details please consult Louge, N. (2014). *2014 final evaluation report teachers’ literacy knowledge, instructional practices, and their students’ reading performance in PAQUED supported schools in the Democratic Republic of Congo*. Education Development Center.

Chapter A: PAQUED Project Context and Evolution; Study Design, Objectives, and Limitations

1. Project Context and Evolution

The intended outcomes of PAQUED as originally proposed focused on providing safe, high-quality learning environments for children in the DRC that would enable them to develop the skills in literacy and numeracy that could help the country recover post-conflict. However, significant technological, logistical, and budgetary complications arose over the first two years of the project leading the project to commission a study by School to School International (STS) to evaluate whether the original design was still adequate to meet the project's goals. To understand the environment in which PAQUED operated, an excerpt from STS's Midterm Review (MTR) report is illustrated:

...working in the DRC is difficult. Beyond the ever-present problems imposed by logistical barriers, dilapidated infrastructure, and the sheer size of the country, there lies a deeper, more problematic issue. As one interviewee said, "It's hard to work in a system where people have given up, they no longer believe in good things. The system is very, very broken. No matter how much you put in there, it's difficult to feel the results." DRC is a country in transition – a country that is at once post-conflict country [*sic*] and struggling to emerge from conflict – and many people, not least those living in the DRC, are seeking ways to turn the tide of cynicism, corruption, and systemic decay. But it will take time. It is in this context that PIEQ and other assistance programs are struggling to deliver results within reasonable time and budget constraints, and to achieve genuine improvements in the quality of education – in short, an extremely challenging task (Lynd, 2012, p. 28).²²

²²Lynd, M. (2012, September). *Final report: Mid-term review: Package for improving educational quality (PIEQ) project*. Washington, DC: USAID

Project Implementation pre-2012

PAQUED was designed as a post-conflict recovery program with a heavy emphasis on teacher professional development. Significant inputs were devoted to the reconstruction of damaged or destroyed school buildings, the training of large numbers of teachers, the delivery of basic school materials for teachers and students, the development of a large number of instructional video modules aimed at delivering teacher professional development content, and production of interactive audio instruction (IAI) lessons for teachers to work through along with their students.²³

The MTR noted a number of successes in the first two years of implementation, most notably in terms of the strength of collaboration between PAQUED and Ministry of Education (MEPSP – *Ministère d’Enseignement Primaire, Secondaire, et Professionnel*) officials; increased involvement of parents and community members in their students’ schools; the collaborative development of classroom based IAI programs which complemented the approved primary school curriculum; the provision of very large numbers of kits of school materials; and the revitalization of the MEPSP’s lapsed cluster- and school-based structures for delivery of continuous professional development (CPD).

However, actual delivery of the core instructional elements of PAQUED’s IAI intervention in the period from 2010 to 2012 was beset by serious technical, logistical, and other difficulties. IAI lessons targeting students in Grades 5–6 had yet to be delivered to schools by the time of the midterm assessment, which is the reason Grade 6 students were excluded from the sample. Although lessons had been developed for students in Grades 1–4, manufacturing defects affecting the vast majority of the radio and video players delivered to schools likely prevented most students from being exposed to them. Even where radios were functioning as desired, the ratio of teachers to radios was too high for most teachers to use them consistently. In fact, PAQUED project staff interviewed for the MTR speculated that no more than 20–30% of intervention teachers were using the IAI materials.

Finally, while the project design called for Ministry of Education Inspectors to provide regular follow-up visits to schools, the combination of long distances, poor infrastructure, and a lack of financial incentives contributed to school visits occurring rarely if at all.

Even before the results of the 2012 EGRA/EGMA study suggested that the intervention had resulted in minimal impact, the MTR recommended several significant changes to the PAQUED approach that should be implemented if the project were to expect to meet its goals.

²³ In prior reports this was referred to as Interactive Radio Instruction (IRI); PAQUED updated the name to reflect different technological modes of delivery.

Realignment of PAQUED's Focus for 2012–2014

In early 2012, as preparations for the midterm EGRA and EGMA study were beginning in earnest, PAQUED commissioned School to School International to conduct an external Midterm Review to obtain a sense of the effectiveness of implementation to date and what changes might be effected in order to maximize impact.²⁴ The review found that the PAQUED project was spread too thinly. While PAQUED had earned plaudits for its successes in training nearly 30,000 teachers during two summer institutes; rapidly developing nearly 500 IAI lessons; developing and deploying its community mobilization training modules; and making significant inroads in the policy arena by partnering closely with the MEPSP, it had struggled to providing significant support for and oversight of the ongoing delivery of its intervention in all 3,000 PAQUED schools. (This was corroborated by the Midterm EGRA and EGMA reports, which found essentially no statistically significant improvements in learning outcomes two years after project launch.) Furthermore, the Midterm Review noted that the project as originally designed and implemented was not very well aligned with USAID's new education strategy, especially Goal 1, which focuses on reading.

EDC proposed, and USAID accepted, a realignment of PAQUED's goals and some substantive changes to the implementation strategy. The most substantive change came in the form of a new, intense, reading-focused intervention to be implemented in a drastically reduced number of schools during the final academic year of the project. The schools which received this intensified intervention are referred to throughout this report as the Reading Program schools.

Elements of the PAQUED Intervention pre-Realignment

The section immediately below describes the project intervention as implemented in the PAQUED and Accessible PAQUED schools; it addresses inputs targeted at students first before turning to inputs targeted at teachers. The subsequent section highlights the changes effected following the project realignment and introduction of the Reading Program schools; it also presents student inputs followed by teacher inputs.

PAQUED Inputs Targeting Students

Since the program was originally designed to support the existing curriculum, the IAI lessons for Grades 1 and 2 did not focus strongly on reading; students in classrooms where the teacher was implementing the program with fidelity would be exposed to a single 30-minute lesson on reading. They would also be exposed to a single 30-minute lesson on mathematics and five 30-minute lessons on French and life skills. Students were not provided with reading books.

²⁴ Lynd, M. (2012). *Final Report: Mid-Term Review: Package for Improving Educational Quality (PIEQ) Project*. Reported by School to School International.

The IAI lessons for Grades 3 and 4 were more substantively focused on reading and mathematics than were those for the earlier grades. The curriculum provided for 50 30-minute lessons on reading and 50 30-minute lessons on mathematics distributed across 100 units. Units were completed at a rate of one every two weeks, meaning that a student whose teacher was implementing the program with fidelity would be exposed to an average of 2.5 lessons, or 1 hour and 15 minutes, of reading-related content per week and 2.5 lessons of mathematics-related content per week. However, students in Grades 3 and 4 were not provided with reading books either. Further complicating matters, the reading lessons were not aligned with the gaps in students' skills that were identified by the 2010 EGRA because the development of the materials had to begin before the 2010 data collection, analysis, and reporting were complete.

The development of the Grade 5 and 6 IAI lessons began after the reporting of the 2010 EGRA/EGMA results was complete. As a result, the reading lessons were informed by concrete knowledge of Grade 6 students' strengths and weaknesses. The content that students in Grades 5 and 6 received was similar to that provided to students in Grades 3 and 4, resulting in a similar 2.5-lessons-per-week (1 hour and 15 minutes) level of exposure to reading and mathematics material. Students in Grades 5 and 6 also did not receive reading books. The IAI lessons for Grades 5 and 6 were not delivered to the field until October 2012; i.e., after the 2012 Midterm EGRA/EGMA had already been conducted. The absence of meaningful inputs to Grade 6 students by the time preparations for the 2012 EGRA/EGMA were underway was the main reason PAQUED leadership elected not to evaluate Grade 6 student performance in 2012.

Table illustrates the inputs that were provided to students in the PAQUED schools and the Accessible PAQUED schools over the lifetime of the project.

Table A1. PAQUED Intervention—Inputs Targeting Students

Inputs	2010–2012, PAQUED Schools			2012–2014, Accessible PAQUED Schools		
	Grades 1–2	Grades 3–4	Grades 5–6	Grades 1–2	Grades 3–4	Grades 5–6
General French & life skills	5 lessons/week	--	--	--	--	--
IAI Lessons French reading (30 min/lesson)	1 lesson/week	2.5 lessons/week	--	1 lesson/week	2.5 lessons/week	2.5 lessons/week
Key math skills (30 min/lesson)	1 lesson/week	2.5 lessons/week	--	1 lesson/week	2.5 lessons/week	2.5 lessons/week
Student Readers	None	None	None	None	None	None
Student Learning Kits	Slates, pencils, chalk	Notebooks, pencils, geometry kits	--	Slates, pencils, chalk	Notebooks, pencils, geometry kits	Notebooks, pencils, geometry kits

PAQUED Inputs Targeting Teachers

Teachers were the central focus of the PAQUED intervention. They received training through multiple avenues, ranging from large-scale 10-day workshops to instructional videos screened on Ran-10 video players at cluster-level peer-learning events called *forums d'échange*. Although teachers were intended to receive regular follow-up visits from Ministry inspectors, this happened so infrequently during the first two years of the project that the responsibility was shifted to school principals and other personnel for the final two years. **Table A2** illustrates the training inputs provided to teachers for Grades 1–6 in PAQUED and Accessible PAQUED schools.

Table A2. PAQUED Intervention—Training Inputs Targeting Teachers

Inputs		2010–2012, PAQUED Schools			2012–2014, Accessible PAQUED Schools		
		Grades 1–2	Grades 3–4	Grades 5–6	Grades 1–2	Grades 3–4	Grades 5–6
Interactive Audio	Face-to-face training on IAI usage	2–3 days/year, with coaching by inspectors		--	2–3 days/year, with coaching by inspectors and school principals		
	Follow-up visits	Maximum of 1 visit per quarter by Ministry inspectors		--	Maximum of 1 visit per quarter by PAQUED personnel		
Content Knowledge	Face-to-face training, French language	10 days, 2011 Summer Institute			--		
	Face-to-face training, math skills		--	10 days, 2012 Summer Institute			
	Cluster-level meetings	Minimum of 1 per month					
Video Instruction	Ran-10	1 Ran-10 video player per cluster					
	Video training modules	Active learning in IAI Making and using instructional materials (for French and mathematics)			No new modules		

Teachers were given kits and video-based guides for making their own instructional materials from locally available materials, access to MP3 players pre-loaded with IAI lesson content, teacher guides to accompany the IAI lesson plans, and a few reading-oriented inputs. **Table A3** details the materials provided to teachers for Grades 1–6 in PAQUED and Accessible PAQUED schools.

Table A3. PAQUED Intervention—Materials Inputs Targeting Teachers

Inputs	2010–2012, PAQUED Schools			2012–2014, Accessible PAQUED Schools			
	Grades 1–2	Grades 3–4	Grades 5–6	Grades 1–2	Grades 3–4	Grades 5–6	
Teacher kits for making instructional materials		1 per school			No kits – only chalk		
IAI							
	MP3 player with SD card preloaded with lesson content	1 MP3 player per 12 teachers		1 MP3 player per 8 teachers			
	Teacher guide for IAI	1 guide per 3 teachers	1 guide per 2 teachers	--	1 guide per 3 teachers	1 guide per 2 teachers	1 guide per 1 teacher
Reading	Story cards (8–10 cards per set)	1 set per teacher	--		1 set per teacher		

The tables presented above describe the elements of the PAQUED intervention that teachers and students in PAQUED schools were intended to receive. In practice, delivery of IAI lesson content was impeded by faulty radio units; the solar panels would not adequately charge the batteries, and the hand cranks frequently snapped off from overly vigorous use. Similarly, problems with the batteries on the Ran10 video players prevented video modules from being used as frequently as intended.

Elements of the “Reading Program” Intervention Post-Realignment

Students and teachers in the Reading Program schools benefitted from a more robust set of inputs than schools in the broader PAQUED populations. These inputs, which began to be provided following project realignment in January 2013, are described below.

Reading Program Inputs Targeting Students

The most substantive change in the inputs provided to students was the introduction of student texts. For the first time under PAQUED, children were exposed to significant quantities of reading materials. The use of IAI lessons was also streamlined, with the lessons on French language and on life skills being discontinued. **Table A4** details the student inputs provided by the Reading Program intervention.

Table A4. Reading Program—Inputs Targeting Students

Input		2013–2014, Reading Program Schools		
		Grades 1–2	Grades 3–4	Grades 5–6
IAI lessons	Reading (30 min/lesson)	1 lesson/week	2.5 lessons/week	2.5 lessons/week
	Key math skills (30 min/lesson)	1 lesson/week	2.5 lessons/week	2.5 lessons/week
Student readers		1 per 2 students	N/A	N/A

Reading Program Inputs Targeting Teachers

The level of support provided to teachers for Grades 1–2 was drastically increased under the Reading Program. (Teachers for Grades 3–6 received very few additional inputs, but continued to receive the same support as was provided to their Accessible PAQUED colleagues.) These teachers for Grades 1–2 received teacher guides containing lessons for 30 minutes of daily reading instruction, “big books” for use in read-aloud exercises, and alphabet charts. They received 18 days of face-to-face training on reading pedagogy (spread out across three training workshops of 10, 3, and 5 days respectively), reinforced by monthly visits from PAQUED staff trained in reading pedagogy. They were also encouraged to more regularly attend the cluster- and school-level *forums d’échange*, the

content focus of which was altered to focus on reading. *Table A5* illustrates these changes.

Table A5. Reading Program—Additional Inputs Targeting Teachers

Inputs	2012–2014, Reading Program Schools		
	Grades 1–2	Grades 3–4	Grades 5–6
Teacher kits for making instructional materials	Alphabet charts	--	--
IAI	MP3 player with SD card preloaded with lesson content	1 MP3 player per grade level	
	Teacher guide for IAI	1 guide per 1 teacher	
Reading	Story cards (8–10 cards per set)	1 big book per teacher per week (24 total)	Story cards
	Teacher guide for Reading Program	1 per teacher	3 generic reading strategies focusing on vocabulary and comprehension
	Cluster-level peer learning meetings (with video modules)	1x per quarter; facilitator trained in reading pedagogy	--
	School-level peer learning meetings	2x per month	--
	Face-to-face training	18 days	--
	Follow-up visits	1 per month by trained PAQUED staff	--

2. Study Design and Objectives

Research Design

Each of the studies (2010, 2012, and 2014) included sufficiently large samples for results to be representative of the three PAQUED provinces²⁵ and to allow for comparisons across these provinces. The ability to allow interprovincial comparisons is important because of systematic differences between the provinces PAQUED serves, including recent history with conflict, geography, economy, and languages spoken.

The goal of the 2014 study and this report is to describe changes in student performance in reading and mathematics that took place in PAQUED schools between 2010 and 2014 and in the Accessible PAQUED schools between 2012 and 2014. Comparing the treatment schools to control schools with similar characteristics using a difference-in-differences approach will clarify the extent to which these changes may be attributable to PAQUED’s IAI, teacher-training, and Reading Program interventions.

The EGRA Tool

The EGRA instrument is a 15- to 20-minute orally administered set of subtasks designed to assess foundational literacy skills crucial to becoming a fluent reader. The EGRA assessment is designed to be a method-independent approach to assessment, in that the instrument does not reflect a particular method of reading instruction (i.e., “whole language” or “phonics-based” approach). Rather, EGRA measures basic skills that a child must have to eventually be able to read fluently and with comprehension—the ultimate goal of reading. The EGRA subtasks are based on research regarding a comprehensive approach to reading acquisition across languages, including five essential components: phonemic awareness, phonics, reading fluency, vocabulary, and comprehension.²⁶

EGRA assessments are adapted to the language(s) and locality where they are administered. The French EGRA contains various subtasks with items selected on the basis of the properties of French. Specific measures assessed in the DRC included vocabulary, phonemic awareness through identification of initial sounds, grapheme recognition, familiar and invented word reading, passage reading and comprehension, oral comprehension, and dictation. The DRC EGRA was designed to pinpoint which basic skills students were lacking and which were causing them difficulty in reading, and to provide this information to MEPSP and PAQUED. A powerful diagnostic tool, EGRA

²⁵ Note that this is different from stating that the provinces sampled are representative of the DRC as a whole.

²⁶ The definitions are adapted from *Report of the National Reading Panel—Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction: Reports Of The Subgroups*, Eunice Kennedy Shriver National Institute of Child Health and Human Development, 2000, National Institutes of Health Publication No. 00-4754, available at <http://www.nichd.nih.gov/publications/nrp/smallbook.cfm>

is able to assess competencies that cannot be measured using a traditional pen-and-pencil examination.

Adaptation of the French EGRA instrument for the DRC took place in December 2009. DRC ministry officials and *Centre de Linguistique Théorique et Appliquée* (CELTA) language experts participated in the instrument adaptation process, which was led by Dr. Souhila Messaoud-Galusi of RTI International.²⁷ The resulting EGRA instrument included the following subtasks:²⁸

The *Vocabulary* subtask assessed students' basic receptive vocabulary and their understanding of basic commands. In this untimed subtask, students were asked to point to body parts (such as “nose”) and objects in the environment (such as “eraser”) as well as to follow simple commands differentiated by prepositions (such as “put the pencil under the paper” followed by “put the pencil above the paper”). After two practice items, students were asked to identify eight body parts, identify six objects in the environment, and put a pencil in six different locations. The final score was the number of words and simple commands that students successfully identified, with the maximum possible overall score being 20 (eight possible points for parts of the body, six possible points for words of environment, and six possible points for spatial terms). This subtask was administered to children in Grades 2 and 4.

The *Initial Sound Identification* subtask assessed students' phonemic awareness (the ability to explicitly identify and manipulate the sounds of language). Phonemic awareness has been found to be one of the most robust predictors of reading acquisition and is often used to identify students at risk for reading difficulties in the primary grades in developed countries. In this subtask, students were asked to listen to a word (such as “*sac*”) and identify the first sound in that word (in this case, /s/). After two practice items, students were given 10 test items. The final score was the number of words of which students successfully identified the initial sound, with the maximum possible score being 10. This subtask was administered to children in Grades 2 and 4.

The *Grapheme Recognition* subtask assessed students' knowledge of the grapheme-sound relationships critical for sounding out new words. In this timed subtask, students were shown a chart containing 10 rows, each containing 10 graphemes (individual letters or combinations of letters) arranged in random order. Students were asked to tell the examiner either the sounds or the names of as many graphemes as possible within one minute, yielding a fluency score measured in correct graphemes per minute (cgpm).²⁹ This subtask was administered to children in Grades 2 and 4.

²⁷ Details of the adaptation process are outlined in the pilot report, *EGRA Pilot for PAQUED Baseline Assessment, DRC*, prepared by RTI International for EDC with funding from USAID, 2010.

²⁸ Five of the subtasks that tested more advanced reading skills—*Familiar Word Reading, Invented Word Reading, Oral Reading Fluency, Reading Comprehension, and Dictation*—were administered to students in Grade 4 only.

²⁹ Although the subtask presents 100 graphemes, it is possible for students to obtain scores greater than 100 correct graphemes per minute if they complete the subtask in less than one minute. The score is calculated by dividing the total letter sounds (graphemes) correctly identified by the time taken to complete the task

The *Familiar Word Reading* subtask assessed students' skill at reading high-frequency words (those words most frequently used in reading and writing by the target age in the target language). Recognizing familiar words is critical for developing reading fluency. In this timed subtask, students were presented a chart of 50 familiar words arranged in rows of five and asked to read as many words as they could within one minute, yielding a fluency score measured in correct words per minute (cwpm).³⁰ This subtask was administered only to children in Grade 4.

The *Invented Word Reading* subtask assessed students' skill at applying letter-sound correspondence rules to decode unfamiliar words. To ensure that students were sounding out the words rather than recognizing them by sight, administrators showed the students a chart of 50 pronounceable, but invented, words that followed legal spelling patterns of French. Students were asked to sound out as many invented words as they could within one minute, yielding a fluency score measured in correct non-words per minute (cnonwpm). This subtask was administered only to children in Grade 4.

The *Oral Reading Fluency* and associated *Reading Comprehension* subtasks assessed students' fluency in reading aloud a passage of grade-level text and their ability to understand what they had read. These subtasks were administered only to children in Grade 4.

- *Oral Reading Fluency*: The ability to read passages fluently is considered a necessary component for reading comprehension. In this subtask, students were given a 54-word story and asked to read it aloud in one minute. The Oral Reading Fluency score was reported in cwpm.³¹
- *Reading Comprehension*: After the students finished the passage, or after the minute ended, the passage was removed. Students were orally asked five questions that required them to recall basic facts from the passage. The Reading Comprehension score was the number of correct answers, with a maximum possible score of 5.

Listening comprehension is considered to be a critical skill for reading comprehension because it is the ability to make sense of oral language. In the *Listening Comprehension* subtask, the examiner read a 50-word passage to students. Students were then orally asked five questions about that passage. The Listening comprehension score was the total correct answers, with a maximum possible score of 5. This subtask was administered to children in Grades 2 and 4.

and multiplying the quotient by 60. For example, a child identifying 100 graphemes correctly in 50 seconds would receive a score of $(100 \text{ correct letters} \div 50\text{s}) * 60\text{s} = 120 \text{ cgpm}$.

³⁰ Although the subtask presents 50 familiar words, it is possible for students to obtain scores greater than 50 cwpm if they complete the task in less than 1 minute. The score is calculated by dividing the total words read correctly by the time taken to complete the task and multiplying the quotient by 60. For example, a child who read all 50 words correctly in 50 seconds would receive a score of $(50 \text{ correct words} \div 50\text{s}) * 60\text{s} = 60 \text{ cwpm}$.

³¹ This fluency measure is calculated in the same fashion as the cgpm, cwpm, and cnonwpm scores, by dividing the number of items read correctly by the amount of time elapsed, and multiplying the quotient by 60 seconds.

The *Dictation* subtask assessed students' skill at spelling. Students were allowed to spell from memory or by sounding out words and applying sound-spelling correspondences. In this subtask, the examiner read a short sentence to the students (“*Mon ami s'appelle Bola.*”), and students attempted to write the sentence. The three target words that were scored were “*mon,*” “*ami,*” and “*appelle.*” The score was the number of words spelled correctly. This subtask was administered only to children in Grade 4.

Administration of the EGRA includes an “early stop” rule, which requires assessors to discontinue the administration of a subtask if a student is unable to respond correctly to any of the items in the first line (i.e., the first 10 graphemes, the first five words, or the first line of the Oral Reading Fluency story). This rule was established to avoid frustrating children who do not understand the subtask or lack the skills to respond. Before administrators conducted the EGRA, they were required to read to each student explicit information about the test and how it would be used. Students were asked to provide verbal assent to participate in the assessment.

To prevent contamination of 2014 results due to leakage of the 2012 testing protocols, minor modifications were made to the EGRA subtasks between 2012 and 2014. With the exception of the oral reading passage and its associated reading comprehension questions, all items were retained across years. However, the items within a given subtask were redistributed in a random sequence. For example, the grapheme “e” appears three times within the *Grapheme Recognition* subtask. In 2012, it was in the 50th, 66th, and 86th positions; following random redistribution for the 2014 instrument, it appeared in the 1st, 7th, and 85th positions.

The 2012 oral reading passage and the associated comprehension questions were replaced in their entirety with a new passage and questions. To retain the validity of comparing scores across years, an equating study was conducted to determine the relative difficulty of the two passages and derive an equating factor that could be used to convert scores on both passages to a common scale. (For a more technical discussion of this process see the *Equating* section in *Annex III*.)

The EGMA Tool

EGMA is an orally administered instrument that tests children individually and is designed to assess student performance on foundational mathematics competencies that students are expected to master in the early grades. The instrument includes subtasks which are designed to do the following:

- include key skills that developing country and developed country curricula have determined should be acquired in early grades;
- reflect those skills that are most predictive of future performance, according to available research and scientific advice;
- represent a progression of skills that lead toward proficiency in mathematics; and
- target both conceptual understanding and procedural fluency.

Although RTI researchers have prepared a basic EGMA instrument that serves as a model, it requires local adaptation in each country in which it is applied, often including translation of test instructions. Adaptation of the EGMA instrument for the DRC took place in December 2009. Ministry officials participated in the instrument adaptation process, which was led by mathematics education expert Aarnout Brombacher and involved ensuring that the EGMA subtests corresponded to the DRC curricula for the targeted grades.³² The resulting EGMA instrument included the following subtasks:

Oral Counting Fluency. The assessment of oral counting fluency targets children's ability to produce numbers fluently. In this subtask, children are asked to count as far as they can. The score is based on the last correct number the child says previous to making an error or at the end of a minute. In the DRC in Grade 2, children were asked to count by ones. In Grade 4, children were asked to count by ones and then by tens. In grade 6, children were asked to count by tens and then by twenty-fives.

One-to-one Correspondence. One-to-one correspondence refers to counting objects. This subtask targets children's ability to recognize the items they need to count and to mentally tag those items that they have already counted. This is a timed subtask because the purpose is to elicit a fluency measure. In Grade 2, children were presented with a stimulus sheet with 30 circles and were asked to count them. In Grade 4, children were presented with one stimulus sheet with 80 circles and a second stimulus sheet with 300 circles arranged in groups of 10. This task was not administered to Grade 6 students.

Number Identification. Number identification assesses the student's knowledge and ability to identify written symbols. Here, the stimulus sheet consisted of number symbols presented in a grid. Students were asked to identify the numbers aloud. The Grade 2 instrument included single-, double-, and triple-digit numbers. The Grade 4 instrument included numbers with up to five digits, and the Grade 6 instrument included numbers with up to six digits, common fractions and percentages.

Quantity (Number) Discrimination. Quantity (number) discrimination assesses the student's ability to make judgments about differences by comparing quantities, represented by numbers. The difficulty of numbers included depended on grade level, as in number identification above.

Missing Number (Missing Number). Missing number (Missing Number) assesses the student's ability to discern and complete number patterns. Each item in this subtask consisted of four numbers in a number sequence and a placeholder (a blank line) for a next or missing number. The child was asked to name the missing number. The numbers used and pattern complexity were almost exactly the same for Grades 2, 4 and 6, allowing for clear examination of grade progression.

Word Problems. Word problems assess the student's conceptual understanding of basic operations. Children were presented with oral word problems, which were read to them, and they were asked to solve them. The children were provided with manipulatives to assist in solving the problems.

³² Details of the adaptation process are outlined in the pilot report *EGMA Pilot for PAQUED Baseline Assessment, DRC*, prepared by RTI International for EDC with funding from USAID, 2010.

Basic Operations—addition, subtraction, multiplication, and division problems. This subtask assesses the student’s procedural competency in these basic operations. In this subtask, children are presented with addition, subtraction, multiplication, and division items and asked to solve them. For DRC, the complexity of problems differed by grade level, such that Grade 2 involved only addition and subtraction problems, whereas Grade 4 and 6 also included multiplication and division. Grade 4 and Grade 6 learners were asked to solve problems in writing, whereas Grade 2 students were asked to state the answer aloud.

Student Context, Teacher, and Head Teacher Interviews

In addition to administering the EGRA and EGMA instruments, assessors asked all students a separate series of questions about their family situation, household, and simple indicators of socioeconomic status. Assessors also conducted structured interviews with the students’ classroom teachers and the school’s head teacher to learn more about the classroom practices and environmental factors which may have intervened in the learning process. All interviews were cleared of identifying information and the resultant data were incorporated into the regression analyses presented in this report.

Assessor Training

The assessors who administered the EGRA and EGMA instruments completed an 11-day training course, including three days of practice in schools with children. More candidates were invited to the training sessions than were selected to participate in the data collection process. Assessors were competitively selected based on several factors, including their facility with the data collection modality (Tangerine™ software on a Nexus 7 tablet), their ability to administer the instruments with fidelity to the protocol, and their rapport with the students. The most fundamental requirement, however, was that all selected assessors have achieved inter-rater reliability (IRR) scores in excess of 90% over the course of several assessments. (For more details, see *Annex II*.)

3. Limitations

The current report presents data and analyses that are subject to several limitations. Some are due to the differences between the original (2009) vision for the project and the post-realignment (2013) vision; some are a function of how the DRC educational system's standards have evolved since 2010, when adaptation of the EGRA instrument was conducted; and others are due to the change in sampling design between baseline (2010) and midterm (2012).

Misalignment of the Intervention and Assessment

In alignment with the DRC national curriculum, the inputs addressing education quality focused on French as a broad content area, and in keeping with the pedagogical and linguistic approaches commonly favored in France, reading was treated as merely one thread within the subject of French as a whole. Reading-focused inputs provided by the project's initial design were relatively few. As indicated in *Section 1: Elements of the PAQUED Intervention pre-Realignment*, the IAI lessons were designed as a complement to the existing national curriculum and there was only room to include a single 30-minute IAI lesson per week explicitly addressed the acquisition of reading skills. Children were not being taught the foundational skills needed to be fluent French readers. Instead, the focus was on development of their oral language skills.

EGRA was a relatively new methodology at the time, and the level of inputs that would be required to elicit observable change on reading scores in the Congolese context via a program focused predominantly on oral language was not well known. As a result, the EGRA as it was adapted in 2009 did not contain as many oral-language subtasks as it might include if it were re-adapted today to evaluate a circa-2009 PAQUED-style intervention.

Essentially, while the reading tasks were appropriate for students in each of the three Grades assessed, it is now clear that the level of reading-focused project inputs was significantly out of step with what would be required for significant improvements to be realized.

Curricular Expectations of Grade 2 Students' Reading Abilities

At the time of project launch, the DRC's national educational policy required that the first two years of primary instruction (*premier degré*) be conducted in one of the four national languages. (Depending on the location of the school, the appropriate language might be Ciluba, Kikongo, Lingala, or Kiswahili.³³) French was taught as a subject during these first two years, becoming the medium of instruction in Grade 3. Although a skill-based reading program was absent from the national French curriculum across the entire primary cycle, there were nonetheless expectations that students in Grade 3 and beyond be able to read with understanding in French.

The practical result of this national policy (which has since been revised) was that Grade 2 students were not expected to be able to read in French. During adaptation of EGRA for the Congolese context, it was therefore seen as unnecessary to include robust measures of Grade 2 students' abilities in the areas of decoding (the *Invented Words* subtask), global recognition (the *Familiar Words* subtask), and reading of connected text (the *Oral Reading Fluency* subtask). The subtasks that would have permitted this study to provide greater insight into changes in Grade 2 students' reading performance over time were thus excluded from the Grade 2 instrument. Because the reading intervention that was implemented post-realignment in 2013 focused predominantly on teachers and students in Grades 1 and 2, having data on student performance across a broader range of reading skills would have provided greater insight into the program's impact.

Changes in Sampling Design

The intervention PAQUED implemented in the Reading Program schools following the project realignment in 2013 included a tighter focus on reading and a far greater intensity of inputs on a per-school basis than the original intervention. Broad conclusions about the relevance of the Reading Program experience for the wider educational system in the DRC cannot be supported by the data collected and presented in this study. Non-probability samples—such as the schools in the Reading Program—are subject to unknown biases, and therefore statistical inference to the target population cannot be drawn. Further and more rigorous study would be needed and is warranted given the results of this current exploration.

Chapter C focuses on Accessible PAQUED schools, examining changes in Grade 2 and Grade 4 students' reading and mathematics performance over a two-year span in a subset of PAQUED intervention schools that were more easily accessible by project staff and resources. The results presented in that section of the report are therefore only generalizable to other schools in the three PAQUED provinces that meet similar criteria

³³ In practice, the great majority of government-supported schools lack access to appropriate teaching and learning materials (TLMs) in the national language of their area. (Schools which are *conventionnée catholique*—public schools that the government has permitted the Catholic Church to manage—are often the exception because they are able to bring additional resources to bear to equip their teachers and students with national-language TLMs.)

of safety and accessibility, and not to either the broader population of PAQUED schools or the broader population of DRC schools. The subpopulation in question—the roughly 618 “accessible” schools in the sampling frame—represents roughly 21% of the initial target population of 3,000 PAQUED intervention schools.

Chapter D focuses on PAQUED schools, examining changes in Grade 2, Grade 4, and Grade 6 students’ reading and mathematics performance over a four-year span. The results presented in that section are generalizable to the provincial level.

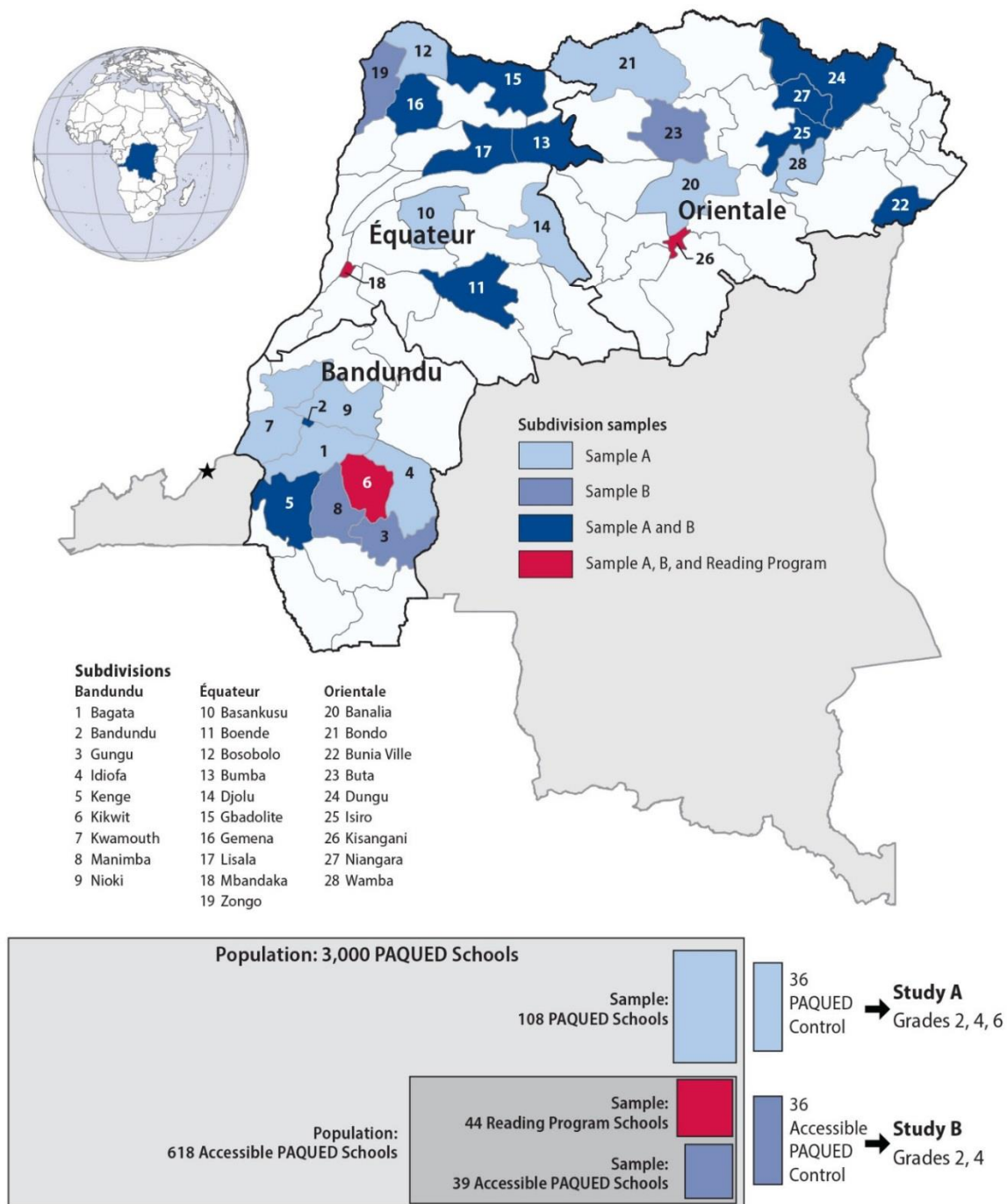
Table A6 below provides a visual representation of the studies whose results are presented in this report.

Table A6. Study Populations and Analyses Permitted

Study	Treatment Groups	2010	2012	2014	Evaluation
PAQUED (Grades 2, 4, 6)	PAQUED Control	Baseline		Endline	Relative growth in performance in PAQUED vs. Control over the period from 2010-2014
Accessible PAQUED (Grades 2, 4)	Accessible PAQUED Accessible Control		Baseline	Endline	Relative growth in performance in Accessible PAQUED vs. Accessible Control over the period from 2010-2014
Reading Program (Grades 2, 4)	Reading Program			Snapshot	Performance in 2014

Figure A1 on the next page provides two visual representations of the school populations assessed during the 2014 endline EGRA/EGMA. The top portion of the image includes a map of the educational subdivisions in which schools were assessed, color-coded according to the population(s) whose schools were located in that subdivision. The bottom of the image provides scale representation of the relative sizes of the populations being studied and the samples drawn. (e.g., the ratio of the area of the light grey *Population: 3,000 PAQUED Schools* rectangle to the area of the dark grey *Population: 618 Accessible PAQUED Schools* is 3,000:618, or ~4.85:1.) In the map and visualization, Study A refers to the PAQUED and Control schools; Study B refers to the Accessible PAQUED and Accessible Control schools; and the Reading Program schools are nested within the Study B shape because they were a subset of the Accessible PAQUED schools.

Figure A1. Subdivisions Containing Schools Assessed at Endline and Relative Size of the Study Populations



Chapter B: Results and Analysis of Student Reading Performance in Reading Program Schools

Chapter B presents the results and analyses of reading performance in Reading Program Schools. The absence of a control group and counterfactual that would permit the analysis to account for the secular trend—that is, the expected growth in student performance that should naturally occur over time—limits the in-depth discussion of performance in the 44 schools to considering the 2014 results. However, at the request of EDC, *Annex IV: Reading Program Schools: 2012-2014 Comparisons* has been added to the report. It examines the change over time in the subset (n = 20) of Reading Program schools that were assessed both in 2012 and 2014.

1. Descriptive Statistics

After the 2012 PAQUED EGRA assessment was completed, EDC selected 44 “accessible” schools from across the three provinces to receive an additional, intensive reading program. These Reading Program schools received a modified intervention that is discussed in more detail in Chapter A, *Elements of the “Reading Program” Intervention Post-Realignment*. Of the 44 Reading Program schools, 20 were both sampled and assessed during the 2012 EGRA/EGMA; they serve as the comparison group for analyses of the Reading Program schools’ change in performance over time. All 44 Reading Program schools were included in the sample drawn in 2014. Data were ultimately collected in 43 Reading Program schools. *Table B1* displays the intended sample size by province and grade.

Table B1. Intended Sample by Province and Grade³⁴

Province	Number of Schools	Number of Students		
		Grade 2	Grade 4	Total Students
Bandundu	17	221	221	442
Equateur	16	208	208	416
Orientale	11 ³⁵	143	143	286
Total	44	572	572	1,144

Due to data collection challenges (i.e., student absenteeism on the day of assessment that prevented a full sample from being collected within a given classroom, and incomplete

³⁴ Throughout this report, sums reported in tables may not equal 100% due to rounding.

³⁵ There were only 11 Reading Program schools in Orientale.

student assessments), 41 student cases were either not collected or removed from the datasets, leaving a total of 1,103 students. This represents a total data loss of only 4%, which is within acceptable limits.

Table B2 displays the actual student sample, by province and grade, used in the subsequent analyses in this report.

Table B2. Actual Sample by Province and Grade³⁶

Province	Number of Schools	Number of Students		Total Students
		Grade 2	Grade 4	
Bandundu	17	218	218	436
Equateur	15	191	189	380
Orientale	11	145	142	287
Total	43	554	549	1,103

Table B3 displays the proportion of schools in each group by school management type. As this table indicates, Catholic-managed and Protestant-managed schools were the most prominent type within this sample (overall $n = 13$ and 19 , respectively). Approximately 19% of the schools were government-managed.

Table B3. School Management Type, by Province ($n = 43$)

Province	Government-Managed	Catholic	Protestant	Kimbanguist	Islamic	Total Schools
Bandundu	4	6	7	-	-	17
Equateur	2	4	6	2	1	15
Orientale	2	4	5	-	-	11
Total	8	14	18	2	1	43

2014 Sample

As indicated above, the final 2014 sample included 1,103 students sampled from 43 schools in three provinces. While most students (40%) came from the Bandundu province and fewest (26%) from the Orientale province, the student sample was approximately even across grade level and gender. **Table B4** describes the general characteristics of the student sample.

³⁶Throughout this report, sums reported in tables may not equal 100% due to rounding.

Table B4. General Characteristics of the Overall Student Sample (n = 1,103)

Variable	Number of Students	Percent
Province		
Bandundu	436	40%
Equateur	380	35%
Orientale	287	26%
Grade level		
2	554	50%
4	549	50%
Sex		
Female	672	57%
Male	476	43%

Tables B5 and B6 show how students responded to a series of questions targeting socioeconomic indicators (SES) regarding possessions in the home. Although students were asked more questions than those listed in the tables below, the questions presented in this report were determined to be of greatest theoretical interest and the most likely to impact student performance. Because student SES has frequently been shown to impact student performance, regression analyses, which are reported later in this report, include an SES composite in their models.³⁷

³⁷ Having several highly correlated independent variables (such as the SES-related questions shown in *Table B5*) can produce unstable estimates; reducing those to a smaller number of principal components will yield a better test and/or more stable estimates of regression coefficients. Therefore, for the purposes of regression analyses incorporating SES, the students' answers to the questions regarding possessions in the home were collapsed into a single variable using principal components analysis (PCA).

Table B5. Student SES Indicators (n = 1,103)

SES Item	Number of Students	Percentage
Radio	883	80%
Telephone	995	90%
Electricity in the home	539	49%
Television	642	58%
Refrigerator	238	22%
Indoor toilets	195	18%
Bicycle	539	49%
Motorcycle	434	40%
Canoe	109	10%
Motor vehicle	120	11%

Table B6. Student Responses to Select Demographic Questions (n = 1,097)

Province	Number of Students	Total in Province	Percent of Province
Student has reading book in class			
Bandundu	93	435	21%
Equateur	52	376	14%
Orientale	87	286	30%
Someone in the student's home is able to read			
Bandundu	361	435	83%
Equateur	326	376	87%
Orientale	219	286	77%
Student has at least one book at home			
Bandundu	141	435	32%
Equateur	56	376	15%

Province	Number of Students	Total in Province	Percent of Province
Orientale	57	286	20%
Student attended kindergarten			
Bandundu	120	435	28%
Equateur	163	376	43%
Orientale	108	286	38%
If teacher assigns homework, student has someone at home to help with it			
Bandundu	130	435	30%
Equateur	68	376	18%
Orientale	90	286	32%

As indicated in **Table B6**, percentages of students across the three provinces were roughly comparable on whether they had a reading book in class and if someone in their home was able to read. However, students in Bandundu appeared to be more likely than students in other provinces to report having a book at home. Conversely, students in Equateur were more likely to have attended kindergarten.

Table B7 shows student self-reports in regard to what language is used in the home. Because students had the option of indicating more than one language, student responses exceed the total number of students in the sample.

Table B7. Student Indication of Language Spoken in the Home

Province	French Number / %	Kikongo Number / %	Lingala Number / %	Kiswahili Number / %	Other Number / %	Total Student Reports
Bandundu	59 / 10%	372 / 64%	87 / 15%	2 / 0%	58 / 10%	578
Equateur	68 / 16%	0 / 0%	332 / 79%	2 / 0%	17 / 4%	419
Orientale	73 / 21%	0 / 0%	187 / 55%	77 / 23%	5 / 2%	342
Total	200 / 15%	372 / 28%	606 / 45%	81 / 6%	80 / 6%	1,339

Perhaps not surprisingly, given differences in predominant languages across provinces, more students in Bandundu reported speaking Kikongo at home than any other language, while more students in Equateur and Orientale reported speaking Lingala at home. Use of

French in the home ranged from a total of 59 (Bandundu) to 73 (Orientale) student reports.

Teachers were also asked a series of questions at the time of the student testing. **Tables B8 and B9** describe the general characteristics and relevant survey responses (responses that will be included in subsequent regression analyses) of the teacher sample.

Table B8. General Characteristics of the Teacher Sample (n = 86)

Variable	Number of Teachers	Percent
Province		
Bandundu	34	40%
Equateur	30	35%
Orientale	22	26%
Grade		
2	43	50%
4	43	50%
Sex		
Female	55	64%
Male	31	36%

As seen in **Table B8**, the teacher sample includes more women (64%) than men (35%). Although 32 schools in Equateur were sampled, data were only available for 30 teachers. Note that in Bandundu and Orientale, two teachers were sampled from each school, as designed.

Table B9. Teacher Responses to Select Survey Questions (n = 86)

Province	Number / %	Number / %	Number / %	Number / %	Number / %	Total in Province
What language do you speak and write in the best?						
	Other	French	Kikongo	Lingala	Kiswahili	
Bandundu	0 / 0%	23 / 68%	11 / 32%	0 / 0%	0 / 0%	34
Equateur	0 / 0%	18 / 60%	0 / 0%	12 / 40%	0 / 0%	30
Orientale	2 / 9%	13 / 59%	0 / 0%	6 / 27%	1 / 5%	22
What is your highest level of education? ³⁸						
	D4	D6	G3	L2		
Bandundu	8 / 24%	26 / 76%	0 / 0%	0 / 0%		34
Equateur	5 / 17%	22 / 73%	3 / 10%	0 / 0%		30
Orientale	3 / 14%	16 / 73%	2 / 9%	1 / 5%		22
Other than PAQUED training, over the past two years how often did you receive in-service training in how to teach the French language?						
	Never	1 Time	2+ Times	No Response		
Bandundu	7 / 21%	4 / 12%	12 / 35%	11 / 32%		34
Equateur	0 / 0%	5 / 17%	16 / 53%	9 / 30%		30
Orientale	0 / 0%	2 / 9%	5 / 23%	15 / 68%		22
How do you characterize your students' competence in French?						
	Weak	Average	Strong	No Response		
Bandundu	4 / 12%	12 / 35%	18 / 53%	0 / 0%		34
Equateur	6 / 20%	18 / 60%	6 / 20%	0 / 0%		30
Orientale	5 / 23%	12 / 55%	5 / 23%	0 / 0%		22
How do you characterize your students' competence in mathematics?						
	Weak	Average	Strong	No Response		
Bandundu	4 / 12%	12 / 35%	18 / 53%	0 / 0%		34
Equateur	6 / 20%	16 / 53%	7 / 23%	1 / 3%		30
Orientale	3 / 14%	12 / 55%	7 / 32%	0 / 0%		22

³⁸ D4 = 4 years of post-primary education; incomplete secondary education (not a high school diploma)
D6 = 6 years of post-primary education; completed secondary education (a high school diploma)
G3 = 3 years of post-secondary education; completion of the first half of a course of study in an *institut supérieure* (a Bachelor's degree under the French educational system)
L2 = 5 years of post-secondary education; completion of a full course of study at either an *institut supérieure* or a *programme universitaire* (a Master's degree under the French educational system)

Province	Number / %	Number / %	Number / %	Number / %	Number / %	Total in Province
How often did you receive a visit by PAQUED personnel this past school year?						
	1	2-3	4-5	6+	No Response	
Bandundu	9 / 27%	11 / 32%	3 / 9%	9 / 27%	2 / 6%	34
Equateur	6 / 20%	8 / 27%	7 / 23%	6 / 20%	3 / 10%	30
Orientale	1 / 5%	5 / 25%	1 / 5%	7 / 32%	8 / 36%	22
How often did you participate in a teacher exchange forum at the cluster level?						
	At Least 1 Time per Trimester	At Least 1 Time per Month	Other	No Response		
Bandundu	7 / 21%	13 / 38%	1 / 3%	13 / 38%		34
Equateur	4 / 13%	7 / 23%	2 / 7%	17 / 57%		30
Orientale	9 / 41%	2 / 9%	1 / 5%	10 / 46%		22
How often did you participate in a teacher exchange forum at the school level?						
	At Least 1 Time per Trimester	At Least 1 Time per Month	Other	No Response		
Bandundu	5 / 15%	17 / 50%	1 / 3%	11 / 32%		34
Equateur	5 / 17%	18 / 60%	1 / 3%	6 / 20%		30
Orientale	3 / 14%	14 / 64%	0 / 0%	5 / 23%		22
If you participated in teacher exchange forums, what video modules were used? *						
	IAI Lessons	Teaching Materials	No Modules Used			
Bandundu	14 / 41%	12 / 35%	12 / 35%			34
Equateur	16 / 53%	17 / 57%	4 / 13%			30
Orientale	12 / 55%	15 / 68%	2 / 9%			22
What resources did you receive from the PAQUED project? *						
	IAI Guide	Reading Activities Guide	Read Aloud Books	Student Texts	Chalk	
Bandundu	28 / 82%	33 / 97%	30 / 88%	30 / 88%	31 / 91%	34
Equateur	22 / 73%	19 / 63%	19 / 63%	16 / 53%	17 / 57%	30
Orientale	2 / 9%	1 / 5%	14 / 64%	14 / 64%	5 / 23%	22

Province	Number / %	Number / %	Number / %	Number / %	Number / %	Total in Province
How many radios did your school receive from the PAQUED project?						
	1	2	3	Other	No Response	
Bandundu	5 / 15%	13 / 38%	13 / 38%	1 / 3%	2 / 6%	34
Equateur	3 / 10%	19 / 63%	3 / 10%	1 / 3%	4 / 13%	30
Orientale	0 / 0%	2 / 9%	3 / 14%	8 / 36%	9 / 41%	22
If you used a PAQUED kit, which one did you use?						
	Materials Fabrication	School Kit	Class Kit	No Response		
Bandundu	24 / 71%	32 / 94%	30 / 88%	0 / 0%		34
Equateur	8 / 27%	23 / 77%	16 / 53%	1 / 3%		30
Orientale	11 / 50%	20 / 91%	12 / 55%	0 / 0%		22
Did you follow the interactive IAI lessons at your school?						
	Yes	No				
Bandundu	30 / 88%	4 / 12%				34
Equateur	26 / 87%	4 / 13%				30
Orientale	19 / 86%	3 / 14%				22

*Because multiple responses were allowed per teacher, percentages do not necessarily sum to 100.

As demonstrated in *Table B9*, most teachers in all provinces reported speaking and writing best in French; relatively fewer teachers stated that Kikongo, Lingala, or Kiswahili were their strongest languages, although patterns in language preference emerged by province. Most teachers reported D6 as their highest level of education attained, with fewer teachers reporting completion only through D4. Very few teachers indicated having completed either G3 or L2 levels of education. Of the teachers who responded, most reported having attended at least two in-service trainings (other than PAQUED trainings) and having received two to three visits from PAQUED personnel. Interestingly, teachers were consistent in categorizing their students' proficiencies in French and in mathematics, with more teachers in Equateur and Orientale rating students as average in both subjects and teachers in Bandundu rating students as either average or good. Promisingly, among teachers who responded to this question, the majority reported having participated in cluster- and school-level exchange forums at least monthly.

The use of video modules in exchange forums was somewhat inconsistent. Teachers in Bandundu were more likely than their peers in other provinces to report not using any videos. However, teachers in Equateur and Orientale used both the IAI Lesson videos and the Materials Development videos nearly equally. Most teachers in Bandundu and Equateur reported receiving a range of PAQUED materials, although Orientale teachers reported receiving far fewer materials. Overall, the school and class kits appear to be

somewhat preferred over the materials fabrication kit. Finally, most teachers reported using the IAI interactive lessons.

Tables B10 and B11 describe the general characteristics and relevant survey responses (responses that will be included in subsequent regression analyses) of the head teacher sample.

Table B10. General Characteristics of the Head Teacher Sample (n = 42)

Variable	Number of Head Teachers	Percent
Province		
Bandundu	17	41%
Equateur	15	36%
Orientale	10	24%
Sex		
Female	10	24%
Male	32	76%

As shown in *Table B10*, head teachers were predominantly male (76% compared to 24% female). This is consistent across provinces, as shown in *Table 17* below.

Table B11. Head Teacher Responses to Select Survey Questions (n = 42)

Province	Number / %	Number / %	Number / %	Number / %	Number / %	Total in Province
What is the sex of the head teacher?	Female	Male				
Bandundu	4 / 24%	13 / 77%				17
Equateur	3 / 20%	12 / 80%				15
Orientale	3 / 30%	7 / 70%				10
How often did you receive a visit by PAQUED personnel this past school year?	1	2-3	4-5	6+	No Response	
Bandundu	0 / 0%	4 / 24%	2 / 12%	11 / 65%	0 / 0%	17
Equateur	0 / 0%	3 / 20%	4 / 27%	7 / 47%	1 / 7%	15
Orientale	1 / 10%	2 / 20%	1 / 10%	6 / 60%	0 / 0%	10

Table B11 also shows that the majority of teachers in Bandundu and Orientale reported receiving six or more visits by PAQUED personnel in the year prior to the survey, compared to Equateur where 27% of teachers reported four to five visits and 20% reported two to three visits.

2. EGRA Subtask Outcomes

Table 18 shows overall zero scores and means scores—both including and excluding students with zero scores—at endline in 2014 for each subtask. A student receives a score of zero on a subtask if that student (1) does not attempt even one item on the task despite being encouraged to try, or (2) attempts items but does not get any correct. On subtasks where relatively few students score zero, the difference between means that include these zero scores and those that exclude them is not large. However, on subtasks where a large proportion of students have zero scores, the difference can be substantial, and it is often useful to consider both means when attempting to understand student performance.

Table B12. Overall Percent Zero Scores and Mean Scores by Grade and Subtask at Endline

Grade	Subtask	% Zero Scores	Mean Including Zero Scores ³⁹	Mean Excluding Zero Scores
2	Vocabulary	2%	9.71	9.88

³⁹ Maximum possible scores for untimed subtasks are as follows: *Vocabulary* (20), *Listening Comprehension* (5), *Initial Sound Identification* (10), *Reading Comprehension* (5), and *Dictation* (3). For

4	Vocabulary	0%	11.97	12.02
2	Listening Comprehension	49%	1.28	2.48
4	Listening Comprehension	33%	1.79	2.67
2	Initial Sound Identification	66%	2.09	6.08
4	Initial Sound Identification	53%	2.84	6.09
2	Grapheme Recognition	22%	15.02	19.38
4	Grapheme Recognition	11%	25.64	28.72
4	Familiar Word Reading	53%	8.75	18.42
4	Invented Word Reading	58%	6.62	15.70
4	Oral Reading Fluency	40%	15.79	26.47
4	Reading Comprehension	77%	0.49	2.17
4	Dictation	38%	1.07	1.72

Subtasks with small percentages of zero scores—such as Vocabulary and, to a lesser degree, Grapheme Sound Knowledge—differences between means are not large. Differences between means that include and exclude students with zero scores are notable on subtasks such as *Initial Sound Identification*, the two word reading tasks, and *Reading Comprehension*. Overall, as anticipated, student mean scores are higher in Grade 4 than in Grade 2 for subtasks that were administered in both grades. The relative amount of increase, however, suggests an ongoing deficiency in skills even at the higher grade. The presence of substantial proportions of zero scores even in Grade 4 further indicates student performance that is lower than required to achieve Grade 4 benchmarks.

The following subsections of this report provide additional detail about each of the EGRA subtasks represented in this report.

Vocabulary

The *Vocabulary* subtask presented children with 20 vocabulary words. Data collectors asked children to identify several body parts and objects as well as to move objects in a variety of directions. As such, this subtest is an assessment of basic French receptive vocabulary. The type of vocabulary assessed is in the DRC curriculum for French in

timed subtasks there is no theoretical upper limit to the fluency score, but the number of items for each subtask is the following: *Grapheme Recognition* (100), *Familiar Word Reading* (50), *Invented Word Reading* (50), and *Oral Reading Fluency* (54).

Grade 1. The curriculum specifies that teaching should include common words and that students should be able to perform a gesture or action based on instructions given by the teacher in French.

Table B13 shows the percent of zero scores and percent of items attempted for the *Vocabulary* subtask.

Table B13. Vocabulary Zero Scores and Percentage of Items Attempted, by Grade and Province

Grade	Province	% Zero Scores		% Attempted		Mean Score	
		2012	2014	2012	2014	2012	2014
2	Bandundu	2%	0%	39%	52%	7.72	10.33
	Equateur	0%	2%	44%	49%	8.73	9.74
	Orientale	0%	3%	44%	46%	8.88	9.20
4	Bandundu	0%	0%	53%	62%	10.64	12.42
	Equateur	0%	0%	52%	56%	10.39	11.10
	Orientale	0%	1%	61%	61%	12.17	12.15

As shown in **Table B13**, even in 2012, very few students (the largest proportion being 2% of Grade 2 students for Bandundu) scored zero on this subtask, suggesting that students had at least a minimal level of oral competence in French. Similarly, across grades and provinces, on average, students were able to attempt close to or more than 50% of the vocabulary at endline. Zero scores notwithstanding, mean scores, even at endline, for Grade 4 students remain low (even in the highest group, correctly responding to only 12.42, or 62% of items). This suggests continuing deficiencies in French oral language ability. Certainly, by Grade 4, students would be expected to have acquired the basic level of oral language competence in French that would allow them to respond to simple vocabulary.

As described in **Chapter A** of the report, the *Vocabulary* subtask had three sections, each comprised of different types of vocabulary words: parts of the human body (e.g., head, foot), common classroom objects (e.g., pencil, eraser), and prepositions (e.g., under, over). As seen in **Table B14**, across overall grades and provinces, students got the highest percentage of correct responses when identifying common classroom objects and the lowest percentage of correct responses when identifying prepositions. Difficulty with prepositions is to be expected when students have not yet mastered a language, although a greater familiarity with vocabulary of classroom objects suggests that students primarily use the French language in the classroom at least some of the time.

Table B14. Vocabulary Mean Scores by Type of Vocabulary Word, by Grade and Province

Grade	Province	Parts of the Human Body (8 items)		Classroom Objects (6 items)		Prepositions (6 items)	
		n	Mean (%)	n	Mean (%)	n	Mean (%)
2	Bandundu	218	4.94 (62%)	218	4.27 (71%)	218	2.49 (42%)
	Equateur	190	5.09 (64%)	190	4.24 (71%)	190	1.94 (32%)
	Orientale	145	4.41 (55%)	145	4.25 (71%)	145	2.19 (37%)
4	Bandundu	218	6.01 (75%)	218	4.88 (81%)	218	3.55 (59%)
	Equateur	189	5.65 (71%)	189	4.68 (78%)	189	2.23 (37%)
	Orientale	142	5.44 (68%)	142	4.67 (78%)	142	3.32 (55%)

Given that students have not mastered this level of French vocabulary overall, it is interesting to explore levels of competence between girls and boys. **Table B15** provides a comparison of mean scores on this subtask by grade, province, and sex.

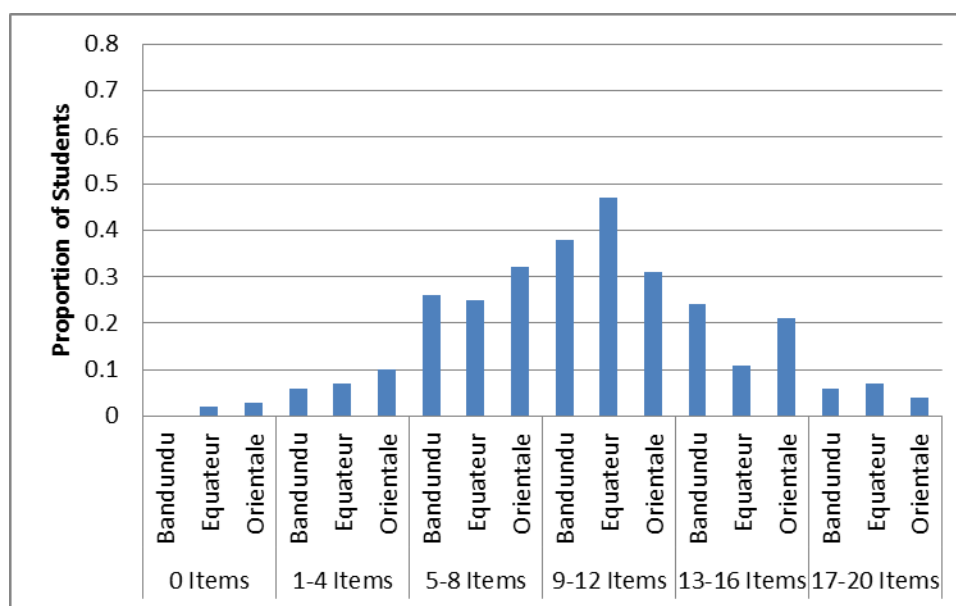
Table B15. Comparison of Vocabulary Mean Scores by Grade, Province, and Sex

Grade	Subtask	Province	Sex	Baseline 2012		Endline 2014	
				Mean (SE)	p_value	Mean (SE)	p_value
2	Vocabulary	Bandundu	Male	8.25 (0.84)		9.92 (0.48)	
			Female	7.93 (0.89)	0.65	10.66 (0.35)	0.25
		Equateur	Male	8.09 (0.23)		8.71 (0.42)	
			Female	8.43 (0.36)	0.49	10.42 (0.35)	0.01
		Orientale	Male	8.45 (0.36)		9.51 (0.59)	
			Female	8.63 (0.68)	0.74	8.92 (0.35)	0.45
4	Vocabulary	Bandundu	Male	11.38 (0.79)		12.29 (0.36)	
			Female	11.30 (0.83)	0.95	12.53 (0.25)	0.62
		Equateur	Male	9.35 (0.86)		10.09 (0.31)	
			Female	9.92 (0.82)	0.59	11.68 (0.27)	0.00
		Orientale	Male	12.26 (1.08)		12.21 (0.39)	
			Female	12.64 (0.91)	0.48	12.10 (0.37)	0.85

While two differences between sexes are significant at the $p < 0.05$ level (Grade 2 and Grade 4 in Equateur, both favoring girls), the number of hypothesis tests presented in this report require a more conservative threshold to avoid type 1 errors.⁴⁰ At this more conservative level ($p < 0.0006$), neither of these differences is statistically significant.

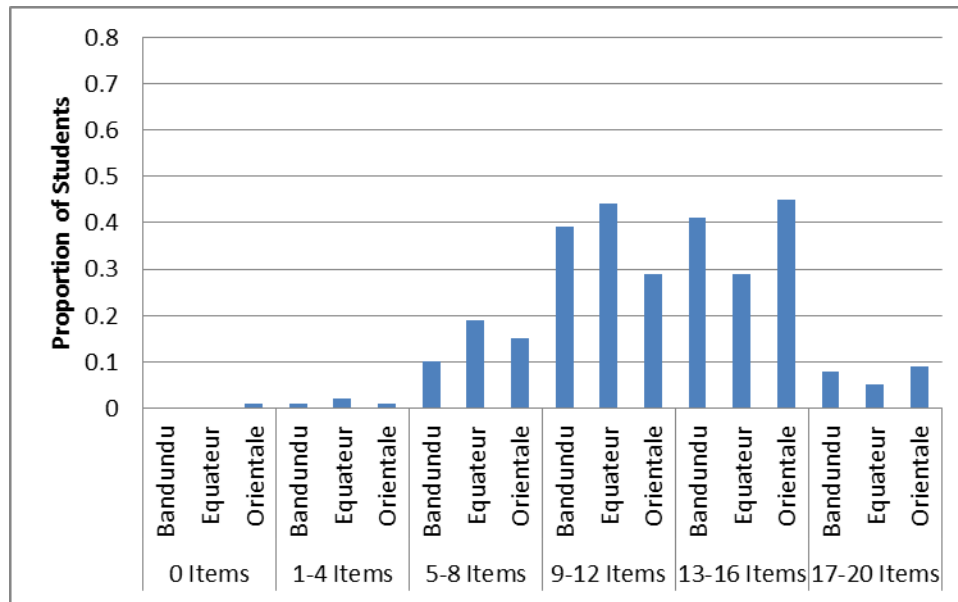
Exploring distributions of scores across the range of possible scores (0–20 items) again shows that relatively few students scored zero, but also relatively few achieved perfect scores on this subtask. By avoiding ceiling effects, this instrument still had room to show additional growth. **Figures B1** and **B2** illustrate these distributions by grade.

Figure B1. Grade 2 Scores on Vocabulary at Endline by Province



⁴⁰ Type 1 errors in statistics occur when a difference is thought to exist where one does not. (Put another way, a type 1 error is the rejection of the null hypothesis when it is actually true.) Due to the large number of means comparisons conducted for this section (Chapter B) of the report, the Bonferroni correction was used to determine the threshold of significance for these analyses. This chapter contains 78 t-tests (comparing performance between gender within treatment groups, by province), and as a result the threshold for significance was determined to be $p < 0.05 / 78 = 0.0006$ for the t-tests.

Figure B2. Grade 4 Scores on Vocabulary at Endline by Province



What is encouraging is that while the distribution for grade 2 shows most students scored in the 5 to 16 item range, this distribution is shifted up for Grade 4, where most students scored between 9 and 16 items.

Initial Sound Identification

In the *Initial Sound Identification* subtask, students listened to individual words, such as “*sac*,” and were asked to identify the first sound (i.e., phoneme) of that word (in this case, /s/). This subtask comprised 10 items, for a maximum possible score of 10. The first 5 items (e.g. *sac*, *jour*, *date*, *lac*, and *car*) were presented to all students. Students who identified the first sound of at least one of these words were presented the remaining 5 items (e.g., *balle*, *tour*, *par*, *vol*, and *fil*).

Table B16 shows the percent of zero scores and percent of items attempted for the *Initial Sound Identification* subtask.

Table B16. Initial Sound Identification Zero Scores and Percentage of Items Attempted by Grade and Province

Grade	Province	% Zero Scores		% Attempted		Mean Score	
		2012	2014	2012	2014	2012	2014
2	Bandundu	69%	48%	11%	30%	1.09	2.98
	Equateur	56%	56%	17%	31%	1.66	3.14
	Orientale	75%	86%	10%	7%	0.99	0.72
4	Bandundu	54%	34%	28%	39%	2.80	3.87
	Equateur	33%	36%	36%	45%	3.54	4.50
	Orientale	56%	81%	18%	9%	1.78	0.90

As shown in **Table B16**, the *Initial Sound Identification* subtask was difficult for students in both 2012 and 2014; even in Grade 4, percentages of zero scores at endline ranged from 34% in Bandundu to 81% in Orientale. Similarly, students were generally able to attempt relatively few items, particularly in Orientale (percent attempted in Grade 2 = 7% and in Grade 4 = 9%). Mean scores were also low, with the highest level of competence on this task was among Grade 4 students at endline in Equateur, who were able to identify on average 4.5 items.

Table B17 provides a comparison of mean scores on this subtask by grade, province, and sex.

Table B17. Comparison of Initial Sound Identification Mean Scores by Grade, Province, and Sex

Grade	Subtask	Province	Sex	Baseline 2012		Endline 2014	
				Mean (SE)	p_value	Mean (SE)	p_value
2	Initial Sound Identification	Bandundu	Male	1.67 (0.31)		2.75 (0.38)	
			Female	0.99 (0.22)	0.17	3.17 (0.32)	0.45
		Equateur	Male	1.70 (0.60)		2.55 (0.41)	
			Female	1.77 (0.64)	0.93	3.54 (0.35)	0.10
		Orientale	Male	0.55 (0.16)		0.86 (0.33)	
			Female	0.47 (0.21)	0.71	0.60 (0.20)	0.52
4	Initial Sound Identification	Bandundu	Male	2.27 (0.81)		3.93 (0.39)	
			Female	2.85 (0.55)	0.46	3.82 (0.35)	0.84

Grade	Subtask	Province	Sex	Baseline 2012		Endline 2014	
				Mean (SE)	p_value	Mean (SE)	p_value
		Equateur	Male	4.06 (0.62)	0.90	2.94 (0.56)	
			Female	3.95 (0.76)		5.39 (0.31)	0.00
		Orientale	Male	0.79 (0.19)	0.21	0.93 (0.31)	
			Female	1.65 (0.50)		0.87 (0.24)	0.89

At the more conservative ($p < 0.0006$) threshold, no statistically significant differences between sexes emerged.

Exploring distributions of scores across the range of possible scores (0–10 items) again shows the difficulty that students had with identifying the initial sound. **Figures B3** and **B4** illustrate these distributions by grade.

Figure B3. Grade 2 Scores on Initial Sound Identification at Endline by Province

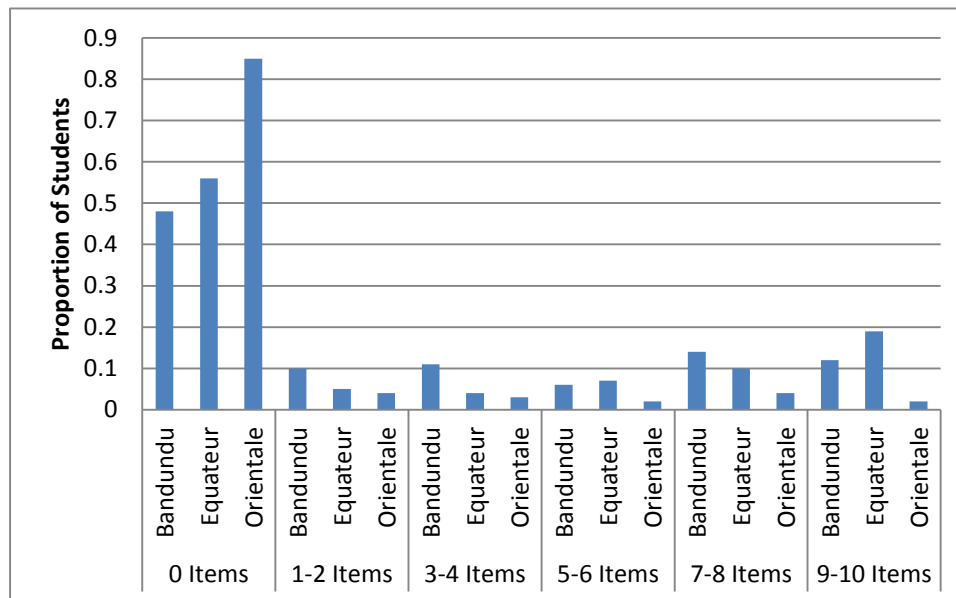
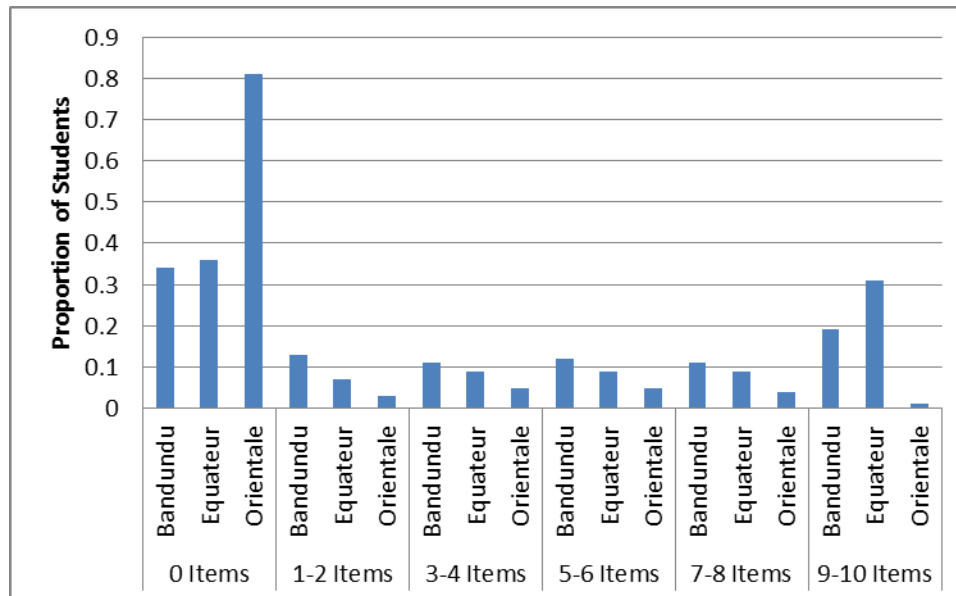


Figure B4. Grade 4 Scores on *Initial Sound Identification* at Endline by Province



As figures **B3** and **B4** illustrate, large percentages of students in both grades scored zero on this subtask at endline. Scores in both grades tended to then be fairly evenly distributed across the other score categories, with a bump in the highest category of scores (9–10 items correct) for Bandundu and Equateur students in Grade 4.

Listening Comprehension

For the *Listening Comprehension* subtask, Grade 2 and Grade 4 students listened to a short passage and answered five questions that assessed their basic comprehension of the passage. Scores reported for this subtask include percentages of students able to answer the comprehension questions based upon the number of students who attempted to answer each question.

Table B18 shows the percent of zero scores and percent of items attempted for the *Listening Comprehension* subtask.

Table B18. Listening Comprehension Zero Scores and Percentage of Items Attempted by Grade and Province

Grade	Province	% Zero Scores		% Attempted		Mean Score	
		2012	2014	2012	2014	2012	2014
2	Bandundu	65%	37%	8%	32%	0.40	1.62
	Equateur	62%	37%	14%	32%	0.71	1.14
	Orientale	62%	54%	13%	21%	0.64	1.07
4	Bandundu	49%	29%	17%	39%	0.84	1.93
	Equateur	50%	46%	22%	29%	1.11	1.46
	Orientale	21%	28%	33%	37%	1.64	1.87

As might be anticipated, given relatively low performance on the *Vocabulary* subtask, students struggled with the *Listening Comprehension* subtask. At endline in Grade 2, across treatment conditions, the percentage of students with zero scores ranged from 37% to 54%. Although the percentages of zero scores in Grade 4 were lower than Grade 2, they are still high with ranges from 28% to 46%. That so many students were unable to correctly respond to even one of the questions asked demonstrates a low overall proficiency with the French language. Even in Grade 4, there was no province in which students were able to correctly respond to two comprehension questions on average.

Table B19 provides a comparison of mean scores on this subtask by grade, province, and gender.

Table B19. Comparison of Listening Comprehension Mean Scores by Grade, Province, and Sex

Grade	Subtask	Province	Sex	Baseline 2012		Endline 2014	
				Mean (SE)	p_value	Mean (SE)	p_value
2	Listening Comprehension	Bandundu	Male	0.52 (0.15)		1.41 (0.14)	
			Female	0.53 (0.12)	0.92	1.80 (0.14)	0.09
		Equateur	Male	1.05 (0.27)		0.79 (0.13)	
			Female	0.89 (0.24)	0.63	1.37 (0.16)	0.02
		Orientale	Male	0.47 (0.10)		1.09 (0.21)	
			Female	0.54 (0.22)	0.78	1.06 (0.14)	0.91
4	Listening Comprehension	Bandundu	Male	1.04 (0.17)		1.58 (0.15)	
			Female	0.98 (0.15)	0.82	2.22 (0.15)	0.01

Grade	Subtask	Province	Sex	Baseline 2012		Endline 2014	
				Mean (SE)	p_value	Mean (SE)	p_value
		Equateur	Male	1.35 (0.24)		1.01 (0.20)	
			Female	1.37 (0.33)	0.94	1.71 (0.15)	0.01
		Orientale	Male	1.36 (0.34)		2.01 (0.20)	
			Female	1.75 (0.47)	0.53	1.76 (0.16)	0.39

At the more conservative ($p < 0.0006$) threshold, no statistically significant differences between sexes emerged.

Exploring distributions of scores across the range of possible scores (0–5 items) again shows the difficulty that students had with this subtask. *Figures B5* and *B6* illustrate these distributions by grade.

Figure B5. Grade 2 Scores on *Listening Comprehension* at Endline by Province

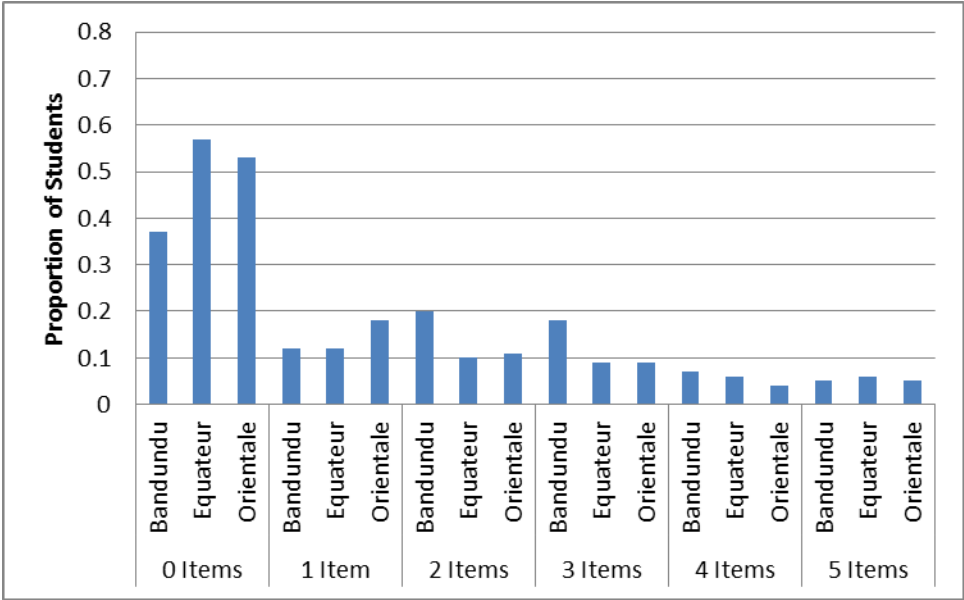
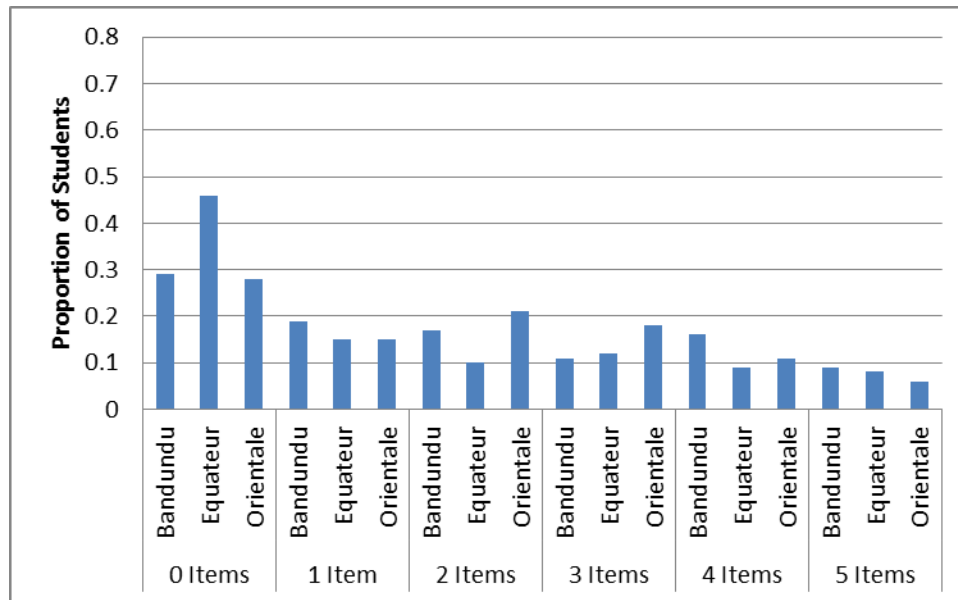


Figure B6. Grade 4 Scores on *Listening Comprehension* at Endline by Province



As illustrated in *Figures B5* and *B6*, large percentages of students scored zero on this subtask at endline in Grade 2, particularly in the Equateur and Orientale provinces. Within Grade 2, in no group did more than 20% of students score between 1 and 5 items. Within Grade 4, scores were more evenly distributed across the score point—with students in Orientale, in particular, reducing the proportion of zero scores—but still lower than needed to demonstrate proficiency with the French language.

Grapheme Recognition

In the *Grapheme Recognition* subtask, students saw a 100-item chart that presented in random order the letters of the alphabet as well as common two-letter digraphs. They were required to produce the sounds for as many graphemes as possible within one minute. This task used the early stop rule and was discontinued for students who were unable to produce any of the sounds for the 10 letters in the first row. Scores reported for this subtask include percentages of items students attempted and the number of grapheme sounds or names that students could correctly generate within one minute.

Table B20 shows percentage of zero scores and percentage of items attempted for the *Grapheme Recognition* subtask.

Table B20. Grapheme Recognition Zero Scores and Percentage of Items Attempted by Grade and Province

Grade	Province	% Zero Scores		% Attempted		Mean Score	
		2012	2014	2012	2014	2012	2014
2	Bandundu	12%	14%	33%	47%	8.64	15.40
	Equateur	16%	17%	28%	53%	6.14	18.95
	Orientale	17%	33%	34%	39%	7.31	12.31
4	Bandundu	4%	7%	57%	60%	19.87	23.90
	Equateur	6%	8%	60%	69%	20.50	30.20
	Orientale	3%	16%	67%	58%	24.90	24.15

As indicated earlier, for this subtask students were given a grid of 100 letters/graphemes for which they were to generate sounds within 60 seconds; even generating sounds at a rate of one per second would result in mean scores of 60. At endline in Grade 4, across provinces, student means ranged from 23.90 to a high of 30.20, indicating that, in general, the highest performing students generated letter sounds at a rate of one every two seconds. Grade 2 student means were lower and zero scores were high, particularly in Orientale.

Table B21 provides a comparison of mean scores on this subtask by grade, province, and sex.

Table B21. Comparison of Grapheme Recognition Mean Scores by Grade, Province, and Sex

Grade	Subtask	Province	Sex	Baseline 2012		Endline 2014	
				Mean (SE)	p_value	Mean (SE)	p_value
2	Grapheme Awareness	Bandundu	Male	11.23 (2.08)		14.82 (1.70)	
			Female	9.19 (1.50)	0.34	15.88 (1.44)	0.66
		Equateur	Male	7.80 (1.25)		16.88 (1.67)	
			Female	6.54 (1.82)	0.66	20.32 (1.63)	0.18
		Orientale	Male	6.08 (0.95)		13.75 (1.83)	
			Female	3.40 (0.62)	0.10	11.03 (1.25)	0.28
4	Grapheme Identification	Bandundu	Male	18.59 (2.11)		24.14 (1.36)	
			Female	22.78 (3.37)	0.37	23.71 (1.27)	0.84

Grade	Subtask	Province	Sex	Baseline 2012		Endline 2014	
				Mean (SE)	p_value	Mean (SE)	p_value
			Male	23.57 (3.39)		25.77 (2.39)	
		Equateur	Female	18.69 (3.76)	0.26	32.71 (1.60)	0.03
			Male	26.16 (6.32)		25.56 (2.25)	
		Orientale	Female	20.89 (5.33)	0.54	23.03 (1.88)	0.43

No statistically significant difference between sexes emerged at the more conservative ($p < 0.0006$) threshold for either grade or any of the three provinces.

Exploring distributions of scores across the range of possible scores (0–100 items) again shows the difficulty that students had with this subtask. **Figures B7** and **B8** illustrate these distributions by grade.

Figure B7. Grade 2 Scores on Grapheme Recognition at Endline by Province

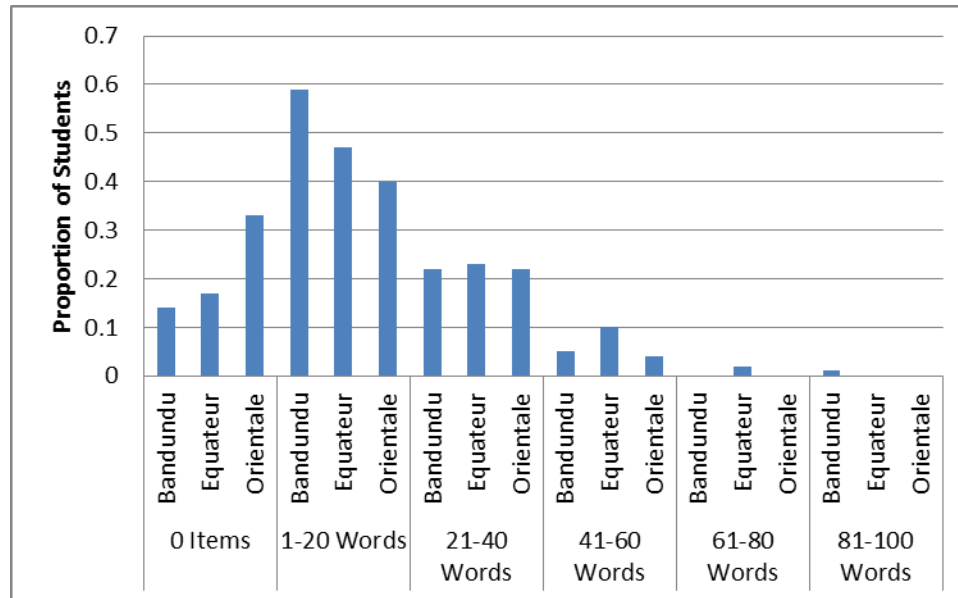
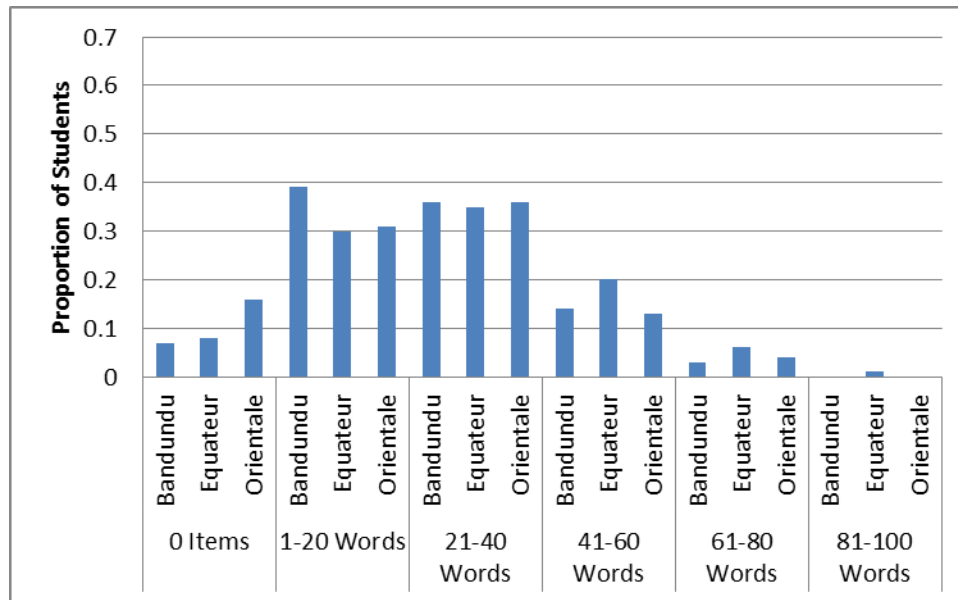


Figure B8. Grade 4 Scores on *Grapheme Recognition* at Endline by Province



As illustrated in *Figures B7* and *B8*, large percentages of students scored between 0 and 20 letters per minute on this subtask at endline in Grade 2, with few students scoring higher than 40. Within Grade 4, fewer zero scores were observed, with the majority of students scoring between 1 and 60 letters per minute. This distribution shows a promising trend, although many students appear to still struggle with grapheme sound and name knowledge and have not gained full automaticity with this skill.

Familiar Word Reading

The *Familiar Word Reading* subtask was administered only to students in Grade 4. In this subtask, students were shown a chart of 50 familiar words (e.g., *tu* and *ami*) and were required to read as many words as they could within one minute. This subtask was discontinued before the end of one minute for students who were unable to read any of the first five words. Scores reported for this subtask include percentages of students able to correctly identify words and the number of words that students could correctly identify within one minute (i.e., cwpm).

Table B22 shows the percent of zero scores and percent of items attempted for the *Familiar Word Reading* subtask.

Table B22. Familiar Word Reading Zero Scores and Percent Attempted for Grade 4 by Province

Grade	Province	% Zero Scores		% Attempted		Mean Score	
		2012	2014	2012	2014	2012	2014
4	Bandundu	37%	53%	27%	26%	6.73	7.67
	Equateur	29%	47%	34%	37%	8.51	10.86
	Orientale	39%	56%	42%	28%	12.53	8.29

As shown in **Table B22**, reading familiar words appears to be a challenging task for the students, with up to 56% of students scoring zero at endline.

Correspondingly, students across the provinces attempted fewer than half of the possible words in both 2012 and 2014. Mean scores at endline were low in all provinces, ranging from 7.67 words per minute in Bandundu to 8.29 words per minute in Orientale. An interesting trend to note is the decreased performance over time in the Orientale group, in which zero scores increased, percent attempted decreased, and mean scores decreased over time. There is no immediate rationale for this trend, although it is something to consider when planning future assessments.

Table B23 provides a comparison of mean scores on this subtask for Grade 4 by province and sex.

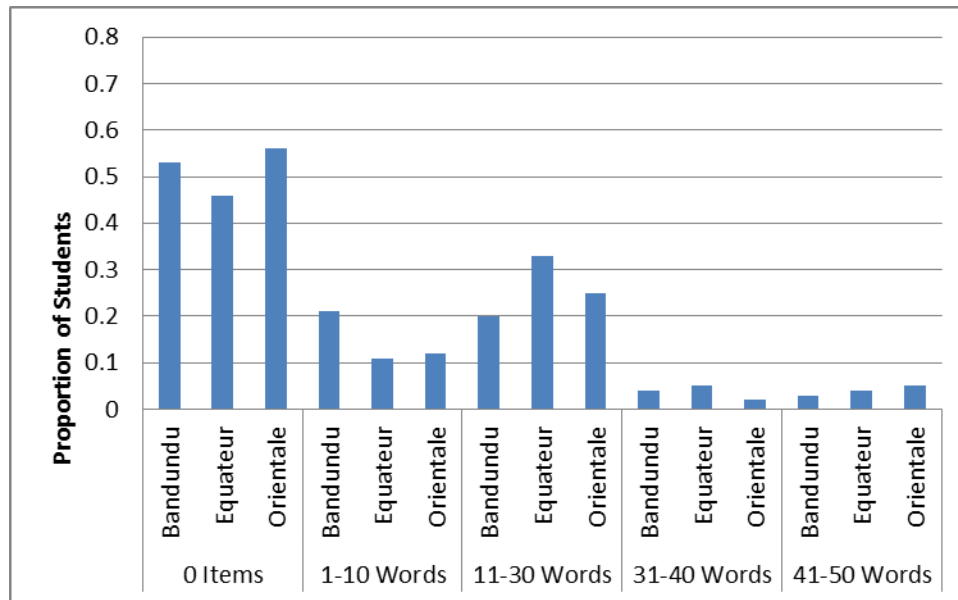
Table B23. Comparison of Familiar Word Reading Mean Scores for Grade 4 by Province and Sex

Grade	Subtask	Province	Sex	Baseline 2012		Endline 2014	
				Mean (SE)	p_value	Mean (SE)	p_value
4	Familiar Word Reading	Bandundu	Male	6.97 (0.97)		7.51 (1.07)	
			Female	7.32 (2.56)	0.90	7.80 (0.83)	0.85
		Equateur	Male	12.40 (3.08)		8.52 (1.86)	
			Female	11.09 (3.39)	0.78	12.18 (1.10)	0.12
		Orientale	Male	10.47 (4.42)		10.11 (1.45)	
			Female	7.46 (3.89)	0.59	6.84 (1.33)	0.16

No statistically significant differences between sexes in any of the provinces emerged in either 2012 or 2014.

Exploring distributions of scores across the range of possible scores (0–50 items) again shows the difficulty that students had with this subtask. **Figure B9** illustrates these distributions for Grade 4.

Figure B9. Grade 4 Scores on *Familiar Word Reading* at Endline by Province



As illustrated in **Figure B9**, large percentages (over 50% for Bandundu and Orientale provinces, and over 40% for Equateur) of students scored zero words per minute on this subtask, although scores did range between one and 30 words per minute. Very few students were able to read more than 30 words per minute, suggesting a lack of automaticity with this skill.

Invented Word Reading

The *Invented Word Reading* subtask was also administered to only the students in Grade 4. In this subtask, students were given a chart of 50 invented words (e.g., *tal* and *vor*) and were required to read as many words as they could within one minute. This subtask was discontinued before the end of the minute for students who were unable to read any of the first five words. Scores reported for this subtask include percentages of students able to correctly identify words and the number of words that students could correctly identify within one minute (correct words per minute [cnonwpm]).

Table B24 shows the percent of zero scores and percent of items attempted for the *Invented Word Reading* subtask.

Table B24. Invented Word Reading Zero Scores and Percent Attempted for Grade 4 by Province

Grade	Province	% Zero Scores		% Attempted		Mean Score	
		2012	2014	2012	2014	2012	2014
4	Bandundu	48%	58%	21%	22%	5.13	5.84
	Equateur	33%	50%	26%	34%	6.39	9.39
	Orientale	50%	63%	35%	22%	9.36	5.50

As with the *Familiar Word Reading* subtask, student scores on the *Invented Word Reading* subtask were quite low. Across the various groups, zero scores across the provinces ranged from 50% to 63% at endline (interestingly, across all provinces, proportions of zero scores were lower in 2012). In general, students in Bandundu and Orientale provinces had particularly low percentages of items attempted and mean scores. However, even in the Bandundu province, students on average were able to read fewer than 10 words in one minute at endline. This rate suggests ongoing difficulties with the skill of decoding.

Table B25 provides a comparison of mean scores on this subtask for Grade 4 by province and sex.

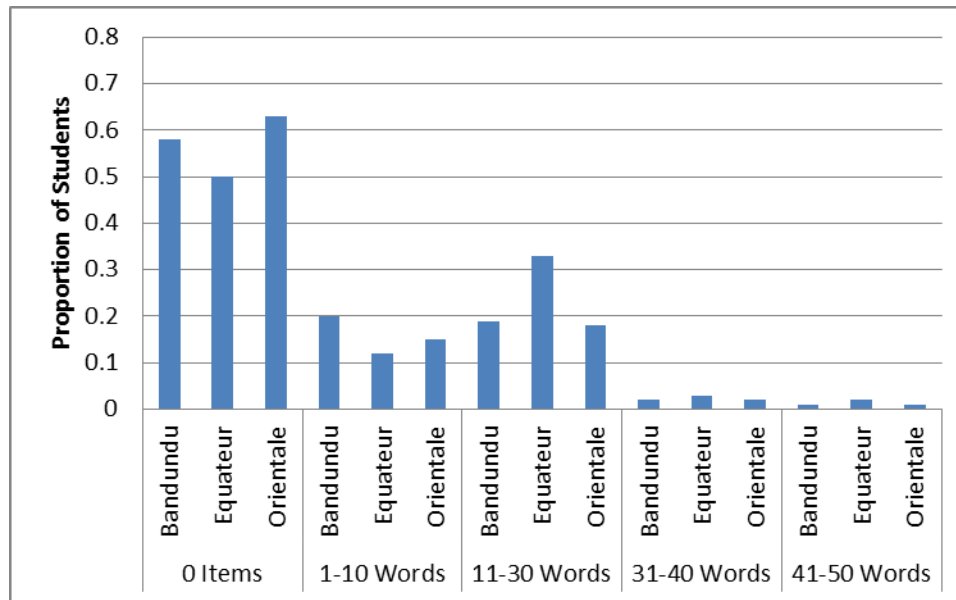
Table B25. Comparison of Invented Word Reading Mean Scores for Grade 4 by Province and Sex

Grade	Subtask	Province	Sex	Baseline 2012		Endline 2014	
				Mean (SE)	p_value	Mean (SE)	p_value
4	Unfamiliar Word Reading	Bandundu	Male	6.08 (0.90)		5.82 (0.87)	
			Female	5.10 (2.09)	0.68	5.85 (0.61)	0.98
		Equateur	Male	9.47 (2.67)		6.54 (1.40)	
			Female	7.49 (2.52)	0.56	11.00 (1.23)	0.03
		Orientale	Male	8.41 (3.47)		7.16 (1.10)	
			Female	5.77 (3.05)	0.57	4.18 (0.89)	0.08

No statistically significant differences emerged between sexes across the provinces in either 2012 or 2014 (using the more conservative $p < 0.0006$).

Exploring distributions of scores across the range of possible scores (0–50 items) again shows the difficulty that students had with this subtask. **Figure B10** illustrates these distributions for Grade 4.

Figure B10. Grade 4 Scores on *Invented Word Reading* at Endline by Province



As **Figure B10** illustrates, as with reading familiar words, large percentages of students scored zero words per minute on the invented word subtask, although just under half of the students were able to read between one and 30 invented words per minute. Very few students were able to read more than 30 words per minute, once again suggesting a lack of automaticity with this skill.

Oral Reading Fluency

The *Oral Reading Fluency* subtask was administered only to students in Grade 4. In this subtask, students were given a passage containing 50 words and were required to read as much of the passage as they could within one minute. Scores reported for this subtask include percentages of students able to read words and the number of words that students could correctly read within one minute.

Table B26 shows percent of zero scores and percent of items attempted for the *Oral Reading Fluency* subtask.

Table B26. Oral Reading Fluency Zero Scores and Percent Attempted for Grade 4 by Province

Grade	Province	% Zero Scores		% Attempted		Mean Score	
		2012	2014	2012	2014	2012	2014
4	Bandundu	50%	43%	22%	35%	8.01	12.34
	Equateur	44%	30%	28%	52%	10.36	21.34
	Orientale	51%	45%	42%	39%	16.49	15.14

As **Table B26** demonstrates, endline mean scores for this subtask ranged from 12.34 in Bandundu to 21.34 in Equateur. Generally higher scores on this subtask, compared to the *Familiar Word Reading* subtask, are expected because reading connected text is typically faster than reading words in isolation. However, even mean scores of 21 words per minute indicate that students were reading, on average, one word approximately every three seconds. This rate of reading speed suggests that attention is being given to individual word reading that should, by Grade 4, be automatic (i.e., less than one second per word). Automaticity would free up attention for comprehension.

Table B27 provides a comparison of mean scores on this subtask for Grade 4 by province and sex.

Table B27. Comparison of Oral Reading Fluency Mean Scores for Grade 4 by Province and Sex

Grade	Subtask	Province	Sex	Baseline 2012		Endline 2014	
				Mean (SE)	p_value	Mean (SE)	p_value
4	Oral Reading Fluency	Bandundu	Male	7.67 (1.42)		12.76 (1.64)	
			Female	10.73 (4.50)	0.54	11.99 (1.23)	0.73
		Equateur	Male	15.43 (4.24)		17.54 (2.44)	
			Female	13.99 (3.93)	0.80	23.49 (1.95)	0.08
		Orientale	Male	12.85 (5.78)		16.99 (2.19)	
			Female	10.12 (5.69)	0.74	13.67 (2.66)	0.38

There were no statistically significant differences that emerged within any of the three provinces in either 2012 or 2014.

Exploring distributions of scores across the range of scores (0–55 items) again shows the difficulty that students had with this subtask. **Figure B11** illustrates these distributions for Grade 4.

Figure B11. Grade 4 Scores on *Oral Reading Fluency* at Endline by Province

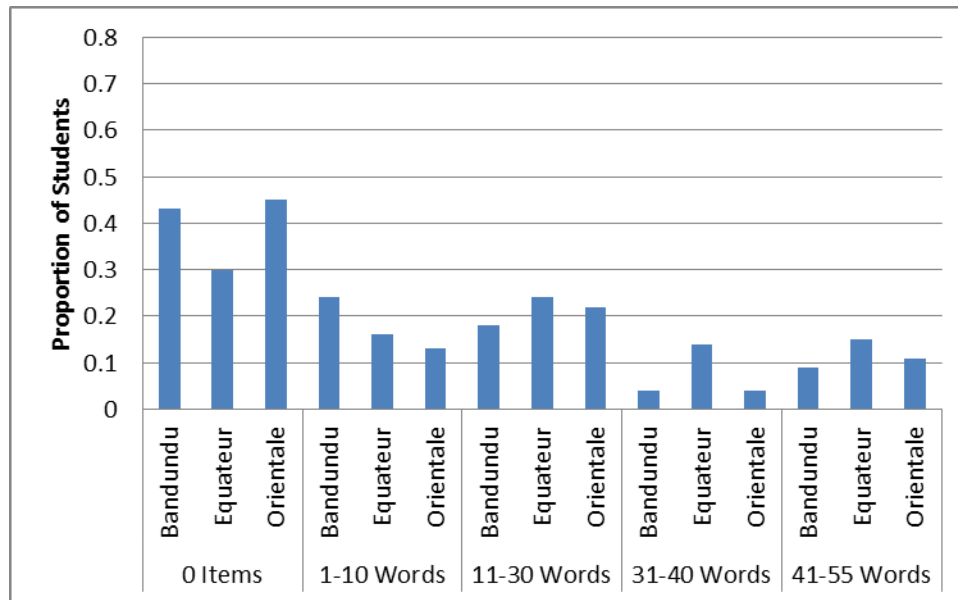


Figure 11 indicates that over 40% of students in Bandundu and Orientale and nearly 30% of students in Equateur scored zero on this subtask. Except for Equateur, fewer than a third of students could read between 11 and 40 words, indicating that they are developing their knowledge of grapheme-sound relationships. A small proportion of students in each province is reading 41-55 words, which is closer to being considered a fluent reader of grade-level text. Although the overall rate is not high enough to allow for effective comprehension, it does show a positive trend across groups.

Reading Comprehension

After reading the passage in the *Oral Reading Fluency* subtask, Grade 4 students were asked questions that assessed their basic comprehension of the passage. Students were only posed questions that aligned with the portion of the passage they had read. Scores reported for the *Reading Comprehension* subtask include percentages of students able to answer comprehension questions, based upon the number of students who attempted to answer each question.⁴¹

Table B28 shows the percent of zero scores and percent of items attempted for the *Reading Comprehension* subtask.

⁴¹ Students were only asked questions that corresponded with the lines of the passage that they were able to read within one minute.

Table B28. Reading Comprehension Zero Scores and Percent Attempted for Grade 4 by Province

Grade	Province	% Zero Scores		% Attempted		Mean Score	
		2012	2014	2012	2014	2012	2014
4	Bandundu	88%	80%	5%	11%	0.17	0.44
	Equateur	81%	69%	10%	17%	0.37	0.70
	Orientale	78%	80%	14%	11%	0.56	0.40

As shown in **Table B28**, across all groups, most students were unable to correctly attempt or respond to any comprehension questions. In both Bandundu and Orientale provinces, 80% of students had zero scores, even at endline. Percentages of attempted items and mean scores were correspondingly low, ranging from 11% to 17% at endline. As mentioned earlier, students were only administered comprehension questions that corresponded with text they were able to read. As seen in the Oral Reading Fluency section of this report, students were on average able to read only between 12 and 21 words. Therefore, on average the students were asked two questions at most. On these two questions, rates of comprehension were relatively low. This aligns with their low performance on vocabulary and listening comprehension subtasks.

Table B29 provides a comparison of mean scores on this subtask for Grade 4 by province and sex.

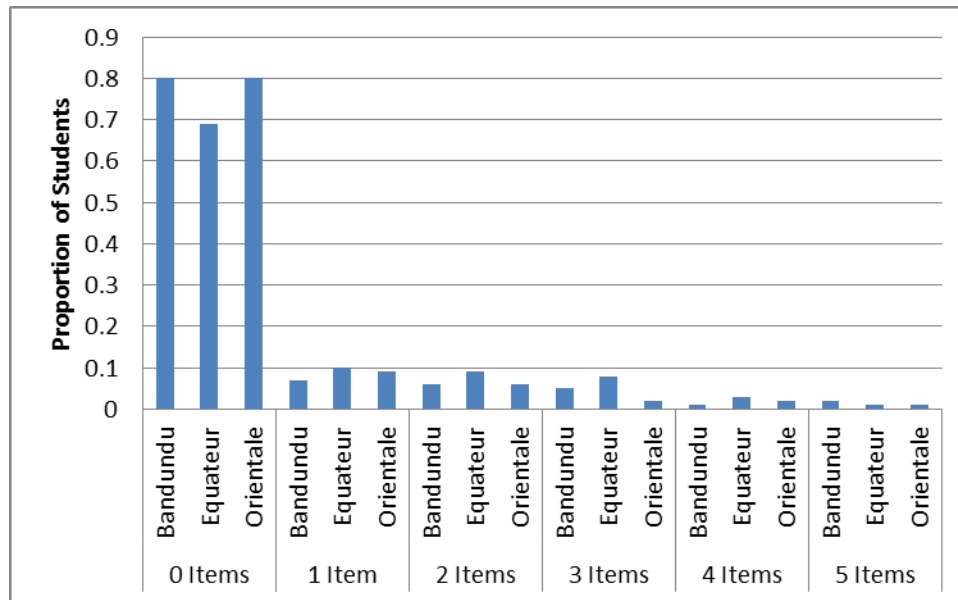
Table B29. Comparison of Reading Comprehension Mean Scores for Grade 4 by Province and Sex

Grade	Subtask	Province	Sex	Baseline 2012		Endline 2014	
				Mean (SE)	p_value	Mean (SE)	p_value
4	Reading Comprehension	Bandundu	Male	0.18 (0.14)		0.42 (0.09)	
			Female	0.21 (0.08)	0.89	0.45 (0.06)	0.80
		Equateur	Male	0.44 (0.15)		0.43 (0.15)	
			Female	0.46 (0.16)	0.93	0.86 (0.12)	0.04
		Orientale	Male	0.47 (0.26)		0.59 (0.13)	
			Female	0.43 (0.28)	0.91	0.25 (0.08)	0.06

No statistically significant difference emerged between sexes in any of the provinces using the more conservative ($p < 0.0006$) threshold.

Exploring distributions of scores across the range of scores (0–5 items) again shows the difficulty that students had with this subtask. **Figure B12** illustrates these distributions for Grade 4.

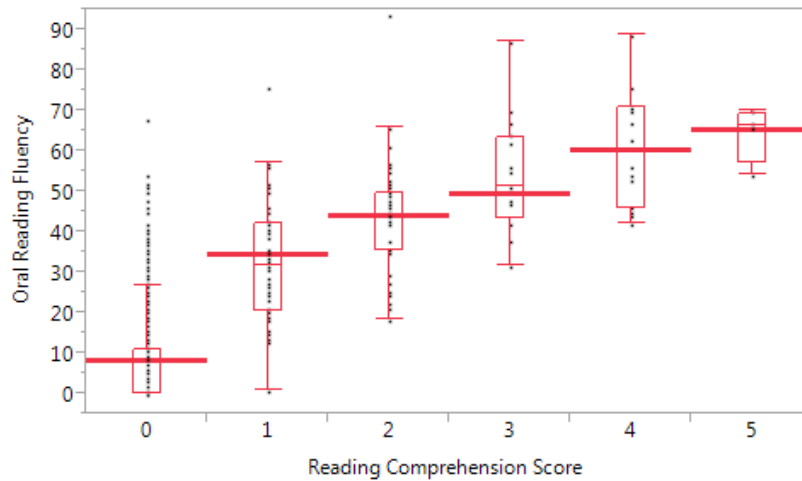
Figure B12. Grade 4 Scores on *Reading Comprehension* at Endline by Province



Given the low overall scores on the *Oral Reading Fluency* subtask, students were administered a limited number of comprehension questions to answer. Students scoring zero on the *Reading Comprehension* subtask include students who did not read far enough into the passage to receive the first question and those who received questions, but did not answer them correctly. As the figure shows, the large majority of students scored zero on this subtask, and the distribution of scores tapers down to nearly zero percent of students correctly answering all five questions.

Another informative way to look at the relationship between oral reading fluency and reading comprehension is to explore which reading fluency levels correspond with given levels of reading comprehension, as displayed in *Figure B13*.

Figure B13. Grade 4 Overall Correspondence between Oral Reading Fluency and Reading Comprehension



Box and whisker plots are good for showing differences between groups. As **Figure B13** illustrates, to be in ‘group 4’ (i.e., answered four questions correctly), the median words read correctly (indicated by the thick horizontal line) was roughly 60 wpm. The thin horizontal lines are “whiskers” and represent the range of wpm read by students in the 25th-75th percentiles in group 4; in this case, students in the 25th percentile read 42 wpm and students in the 75th percentile read 90 wpm. Together these five box and whisker plots show the median wpm increases for each group and the range of words read correctly narrows. While many outliers exist, overall, students reading between 30 and 40 words per minute tend to accurately respond to one comprehension question, whereas oral reading fluency scores of 60 to 70 wpm correspond with reading comprehension scores of 4 and 5.

Dictation

For the *Dictation* subtask, Grade 4 students attempted to write a short sentence (“*Mon ami s’appelle Bola.*”) read to them three times by the assessors. The three target words that were scored were *mon*, *ami*, and *appelle*. Scores reported for this subtask include the number of words spelled correctly.

Table B30. Dictation Zero Scores and Percent Attempted, for Grade 4 by Province

Grade	Province	% Zero Scores		% Attempted		Mean Score	
		2012	2014	2012	2014	2012	2014
4	Bandundu	53%	49%	26%	30%	0.79	0.89
	Equateur	21%	30%	41%	43%	1.22	1.29
	Orientale	46%	32%	41%	37%	1.24	1.09

Table B30 shows the percent of zero scores and percent of items attempted for the *Dictation* subtask. Zero scores on this subtask were not as pervasive as on other subtasks, suggesting that students overall have some ability to apply their grapheme knowledge to spell correctly some common words. The highest percentages of zero scores were observed in Bandundu, with 49% of students scoring zero at endline. Not surprisingly, percentages of items attempted were also lowest in this group relative to the other provinces. For the most part, students were able to write approximately one word correctly, with mean scores ranging from 0.89 in Bandundu to 1.29 in Equateur.

Table B31 provides a comparison of mean scores on this subtask for Grade 4 by province and sex.

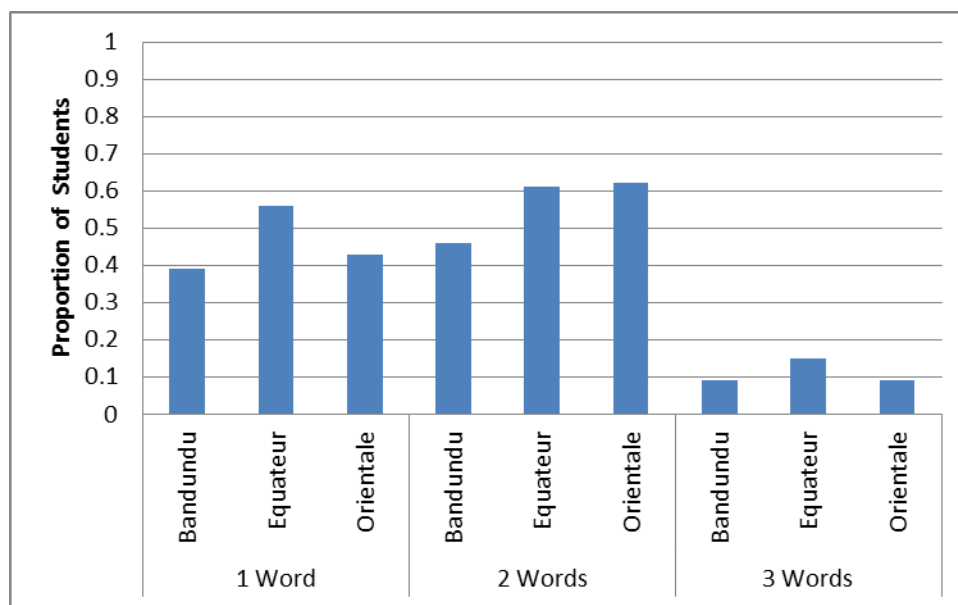
Table B31. Comparison of Dictation Mean Scores for Grade 4 by Province and Sex

Grade	Subtask	Province	Sex	Baseline 2012		Endline 2014	
				Mean (SE)	p_value	Mean (SE)	p_value
4	Dictation	Bandundu	Male	0.83 (0.21)		0.92 (0.11)	
			Female	0.80 (0.24)	0.92	0.88 (0.24)	0.80
		Equateur	Male	1.74 (0.25)		1.02 (0.14)	
			Female	1.20 (0.17)	0.06	1.43 (0.10)	0.03
		Orientale	Male	0.93 (0.29)		1.11 (0.13)	
			Female	1.03 (0.29)	0.66	1.08 (0.12)	0.84

No statistically significant difference emerged between sexes in any of the provinces using the more conservative ($p < 0.0006$) threshold.

Exploring distributions of scores across the range of scores (0–3 items) again shows the difficulty that students had with this subtask. **Figure B14** illustrates these distributions for Grade 4.

Figure B14. Grade 4 Scores on *Dictation* at Endline by Province



Within the *Dictation* subtask, students were scored on fully correct dictation of three words in a longer sentence. As illustrated in **Figure B14**, student scores were relatively evenly distributed across one and two correct words. Overall, less than a quarter of students were able to correctly write all three words.

Student Characteristics Analyses

A series of chi-square tests were run to determine the extent to which relevant student characteristics are correlated with student performance. Because the greatest student literacy gains appeared to be on the *Grapheme Recognition* subtask, that skill, as measured at endline, was used as an indicator of student competency. In addition, given relatively low levels of performance even on the *Grapheme Recognition* subtask, only high-performing students (i.e., those scoring in the top quintile on this subtask) were included in the following analyses. **Table B32** shows weighted percentages, chi-square statistics, and p-values for each of the student characteristics identified earlier in this report as being of theoretical interest for this purpose.

Table B32. Chi-Squared Analyses of Grade 2 Student Characteristics with Student High Performance on the Grapheme Recognition Subtask

Student Characteristic	Weighted Percentages		
Student has reading book in class			
	Not high- Performing Students	High- performing Students	Total
No	80%	67%	77%
Yes	20%	33%	23%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 17.8137$			
Design-based $F(1.00, 38.00) = 6.8924$ $p = 0.012$			
Someone in the student's home is able to read			
No	21%	4%	18%
Yes	79%	96%	83%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 38.8329$			
Design-based $F(1.00, 38.00) = 31.3480$ $p = 0.000$			
Student has at least one book at home			
No	79%	76%	79%
Yes	21%	24%	22%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 1.2495$			
Design-based $F(1.00, 38.00) = 0.5243$ $p = 0.473$			
Student attended kindergarten			
No	67%	54%	65%
Yes	33%	46%	35%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 13.7297$			
Design-based $F(1.00, 38.00) = 6.0163$ $p = 0.019$			

Student Characteristic	Weighted Percentages		
If teacher assigns homework, student has someone at home to help with it			
No	49%	46%	48%
Yes	51%	54%	52%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 0.6986$			
Design-based $F(1.00, 38.00) = 0.1559$ $p = 0.695$			
Student speaks French at home			
No	84%	73%	82%
Yes	16%	27%	18%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 15.9614$			
Design-based $F(1.00, 38.00) = 7.0746$ $p = 0.011$			

Table B32 illustrates that one characteristic of Grade 2 students had a significant correlation (at $p < 0.002$) with high performance on the *Grapheme Recognition* subtask:

- Having someone in the home able to read correlated significantly with high performance on the subtask.

Logistic regression analyses were also run on these student characteristics to further explore their relationship with student performance, when holding province, gender, and student SES constant. The statistically significant relationships that emerged from these analyses are presented in **Table B33**.

Table B33. Grade 2 Student Characteristics Logistic Regressions

Characteristic	p-value
Someone in the student's home is able to read	0.0000*

* significant at $p < 0.002$

When holding province, gender, and student SES constant, this student characteristic remained statistically significant in Grade 2: having access to someone at home who can read correlates with improve performance on *Grapheme Recognition*.

Table B34 shows Grade 4 weighted percentages, chi-square statistics, and p-values for each of the student characteristics identified earlier in this report as being of theoretical interest for this purpose.

Table B34. Chi-Squared Analyses of Grade 4 Student Characteristics with Student High Performance on the Grapheme Recognition Subtask

Student Characteristic	Weighted Percentages		
Student has reading book in class			
	Not high- Performing Students	High- performing Students	Total
No	74%	73%	74%
Yes	26%	27%	26%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 0.1431$			
Design-based $F(1.00, 38.00) = 0.0594$ $p = 0.809$			
Someone in the student's home is able to read			
No	14%	6%	13%
Yes	86%	95%	87%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 12.3023$			
Design-based $F(1.00, 38.00) = 4.3741$ $p = 0.043$			
Student has at least one book at home			
No	85%	64%	80%
Yes	15%	36%	20%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 47.4511$			
Design-based $F(1.00, 38.00) = 20.2986$ $p = 0.000$			
Student attended kindergarten			
No	63%	48%	60%
Yes	37%	52%	40%

Student Characteristic	Weighted Percentages		
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 16.1145$			
Design-based $F(1.00, 38.00) = 6.8436$ $p = 0.013$			
If teacher assigns homework, student has someone at home to help with it			
No	54%	44%	52%
Yes	46%	56%	48%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 5.8808$			
Design-based $F(1.00, 38.00) = 1.9575$ $p = 0.170$			
Student speaks French at home			
No	83%	56%	78%
Yes	17%	44%	22%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 73.1886$			
Design-based $F(1.00, 38.00) = 31.1533$ $p = 0.000$			

Overall, percentages in each cell for each characteristic were similar to those in Grade 2, with some notable exceptions.

- In Grade 4, having a book at home correlated significantly with high performance on the subtask.
- In Grade 4, speaking French at home correlated significantly with high performance on the subtask.

A possible interpretation of the fact that having access to a reading book in class was correlated with high performance for Grade 2 students but not for Grade 4 students involves considering the other supports in place. In Reading Program schools, while all Grade 1 – 6 students received books, it was only the teachers in Grades 1-2 that received coaching support and improved instructional materials. The absence of a correlation between access to books in Grade 4 and high performance may be a result of the Grade 4 teachers' lacking the training and instructional materials necessary to help their students take advantage of the books available to them.

Logistic regression analyses were also run on these student characteristics to further explore their relationship with student performance, when holding province, gender, and student SES constant (*Table B35*). These analyses showed the following statistically significant relationships, which largely correspond with the results of the chi-square tests.

Table B35. Grade 4 Student Characteristics Logistic Regressions

Characteristic	p-value
Student has at least one book at home	0.0001*
Student speaks French at home	0.0000*

* significant at $p < 0.002$

Teacher Characteristics Analyses

Two principal component factor indices were established to explore the relationship between teacher characteristics and student achievement on the *Grapheme Recognition* subtask, which generated the following composite factors:

1. **Teacher participation.** This includes frequency of reported visits by PAQUED personnel, teacher participation in exchange forums at the cluster level, and teacher participation in exchange forums at the school level; and
2. **Teacher access to materials.** This includes resources the teacher received from PAQUED, the number of radios received from PAQUED, and the number of PAQUED kits the teacher reported using.

Regression analyses on these two composites showed that teacher participation did have a substantial impact on student performance ($p = 0.0387$). Conversely, the teacher access to materials composite did not significantly contribute to student performance, at $p < 0.05$.

Logistic regression analyses were also run on these teacher characteristics to further explore their relationship with student performance, when holding province, gender, and student SES constant (*Table B36*).⁴² These analyses showed the following statistically significant relationships.

⁴² Due to the large number of tests conducted for this section (*Chapter B*) of the report, the Bonferroni correction was used to determine the threshold of significance for these analyses. This section contains 8 tests; applying the Bonferroni correction to these analyses results in $p < 0.05 / (8) = 0.0063$.

Table B36. Teacher Characteristics Logistic Regressions

Characteristic	p-value
Speaking and writing best in Kikongo	0.0000*
Having D6 as the highest level of education	0.0007*
Having G3 as the highest level of education	0.0000*
Having participated in non-PAQUED in-service training more than one time	0.0057*
Considering student competencies in French as average	0.0000*
Considering student competencies in French as good	0.0016*
Considering student competencies in mathematics as average	0.0000*
Considering student competencies in mathematics as good	0.0000*

* significant at $p < 0.0063$

These relationships are not immediately intuitive. Further exploration is warranted to determine exactly what characteristics teachers possess that most impact student progress.

Head Teacher Characteristics Analyses

It was hypothesized that the sex of the head teacher might impact the extent to which teachers participated in exchange forums and followed the IAI interactive lessons. However, this was not the case. No significant relationship between the sex of the head teacher and these factors emerged.

3. Summary and Conclusions

There were no consistent, statistically significant trends that emerged from the analyses presented here on the Reading Program schools when disaggregating data by region or by sex. However, some promising indications can be seen. Overall, student performance at Grade 4 exceeded that at Grade 2, although further investigation is warranted in cases where performance decreased over time, whether due to testing error or other factors in the classroom. While not statistically significant at the more conservative ($p < 0.0006$) p-value, a trend for girls to outperform boys at endline—where no difference emerged at baseline (2012)—suggests that the type of intervention used in these schools may either encourage greater participation by girls or improve the efficacy of the learning experience for girls and should be further explored in subsequent implementations.

With this in mind, overall student performance, even at endline, is below what is needed to meet national benchmarks across all subtasks. Even for oral skills such as vocabulary,

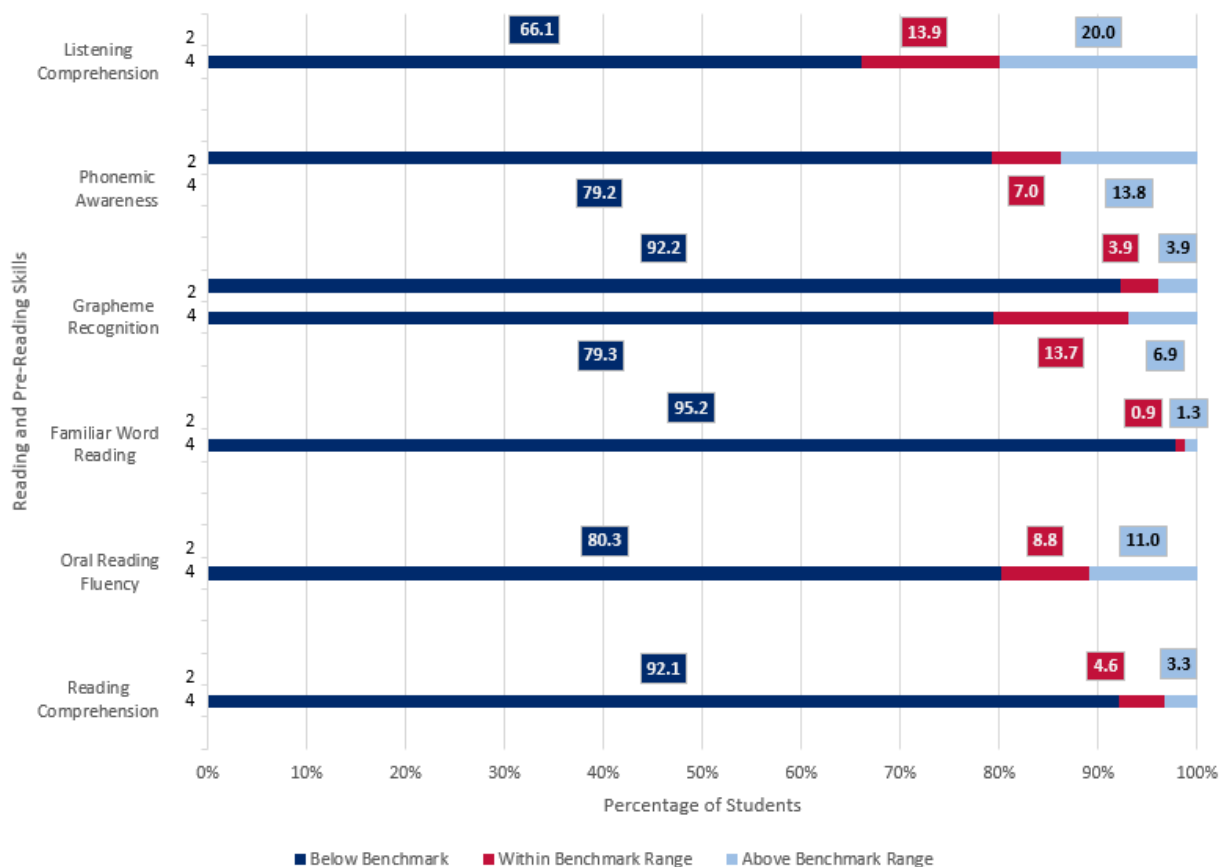
phonemic awareness, and listening comprehension, students failed to demonstrate French oral skills required to effectively read in French. Student mean scores on grapheme and word recognition, as well as connected text reading, were also lower than necessary for reading with fluency and comprehension. *Table B37* and *Figure B15* illustrate these results.

Table B37. Student Performance in Reading Program Schools Relative to National Benchmarks⁴³

Grade	Reading or Pre-reading Skill	Below benchmark		Within benchmark range		Above benchmark range	
		<i>n</i>	(%)	<i>N</i>	(%)	<i>n</i>	(%)
2	Listening Comprehension	-	-	-	-	-	-
	Phonemic Awareness	437	(79.2%)	39	(7%)	78	(13.8%)
	Graphemes	510	(92.2%)	21	(3.9%)	23	(3.9%)
	Familiar Words	-	-	-	-	-	-
	Oral Reading Fluency	-	-	-	-	-	-
	Comprehension	-	-	-	-	-	-
4	Listening Comprehension	368	(66.1%)	72	(13.9%)	109	(20%)
	Phonemic Awareness	-	-	-	-	-	-
	Graphemes	433	(79.3%)	72	(13.7%)	44	(6.9%)
	Familiar Words	521	(95.2%)	6	(0.9%)	8	(1.3%)
	Oral Reading Fluency	440	(80.3%)	46	(8.8%)	63	(11%)
	Comprehension	504	(92.1%)	25	(4.6%)	20	(3.3%)

⁴³ The *n* for each grade presented in this table is as follows: Grade 2 = 554; Grade 4 = 549. Each *n* presented is unweighted, and each percentage presented is weighted.

Figure B15. Performance of Students in Reading Program Schools Relative to DRC Benchmarks



Even if significant differences in student performance did not emerge from these analyses, indications that elements of the intervention were correlated with improved learning outcomes are promising. For example, being taught by teachers who participated in the program (as measured by engagement in forums, receiving training, and receiving visits by program staff) enhanced student performance on the *Grapheme Recognition* subtask. Similarly, having access to books in the classroom significantly correlated with student performance in Grades 2 and 4. It is also promising that other forms of support to students—albeit ones over which schools cannot often exert direct influence—such as attending kindergarten, having books at home, and having someone at home who is able to read, correlate with student performance.

These analyses show areas where further exploration and student improvement are required, but they also identify types of supports that appear to be beneficial and that should be continued and strengthened in future education implementations.

Chapter C: Results and Analysis of Student Reading Performance in Accessible PAQUED and Accessible Control Schools

Chapter C presents the results and analysis of student performance on EGRA measures for the two larger populations of schools in which PAQUED intervened. Since PAQUED's expectation was that the schools which were most accessible would be most likely to receive robust program support, and therefore demonstrate significant gains, the analysis of student performance in the Accessible PAQUED schools (and their comparison group, the Accessible Control schools) are presented in Section 1. Section 2 presents the analysis of student performance in the far larger group of PAQUED intervention schools (using the associated Control schools as a comparison).

1. Student Performance on EGRA Measures in Accessible PAQUED and Accessible Control Schools

Descriptives

To draw the sample of Accessible PAQUED schools, RTI randomly selected six subdivisions in each province from among the subdivisions that PAQUED had identified as being eligible using the 2012 criteria. Again, the 2012 criteria for subdivision eligibility required that the subdivision contain at least six "accessible" schools, where school accessibility was defined as being located within 20 km of an urban center. RTI randomly selected four Accessible PAQUED schools in each of the six subdivisions. Twelve schools (half of those selected) served as the sample school. The remaining half were intended to serve as replacement schools should reaching the sampled schools be impossible or the sampled schools be nonexistent or unwilling to participate. The schools that served as Accessible Control schools in 2012 were included in the same capacity again in 2014.

Table C1 displays the intended school sample size by province and grade. A total of 75 schools were intended to be sampled: 39 PAQUED and 36 Control. *Table C2* shows the intended student sample by province and grade.

Table C1. Intended School Sample by Province

Province	Accessible Control Schools	Accessible PAQUED Schools	Total
Bandundu	12	12	24
Equateur	12	12	24
Orientale	12	15 ^{44[22]}	27
Total	36	39	75

Table C2. Intended Student Sample by Province and Grade

Province/Grade	Accessible Control Students	Accessible PAQUED Students	Total
Bandundu	312	312	624
Grade 2	156	156	312
Grade 4	156	156	312
Equateur	312	312	624
Grade 2	156	156	312
Grade 4	156	156	312
Orientale	312	390	702
Grade 2	156	195	351
Grade 4	156	195	351
Total	936	1,014	1,950
Grade 2	468	507	975
Grade 4	468	507	975

⁴⁴ There were only five subdivisions in Orientale that met the eligibility requirement of containing at least six “accessible” schools. In order to meet the necessary sample size, three schools were sampled per subdivision in Orientale rather than two schools per subdivision, as in the other provinces.

All 75 schools were tested and are present in the following analyses. Overall, across schools and provinces, an additional 15 students were tested, resulting in an actual sample of 1,965 students.

Table C3 displays the actual student sample used in the subsequent analyses in this report, by province and grade.

Table C3. Actual Sample by Province and Grade

	Grade 2			Grade 4		
	Accessible Control	Accessible PAQUED	Total Grade 2 Students	Accessible Control	Accessible PAQUED	Total Grade 4 Students
Bandundu	145	136	281	145	134	279
Equateur	150	153	303	144	151	295
Orientale	180	230	410	174	223	397

Table C4 displays the proportion of schools in each group by school management type. As this table indicates, Protestant-managed schools were the most prominent type within this sample (overall $n = 20$), followed by Catholic-managed schools ($n = 19$). Approximately 21% of the schools were government-managed.

Table C4. School Management Type, by Province ($n = 82$)

Province	Government -Managed	Catholic	Islamic	Kimbanguist	Protestant	Other	Missing	Total Schools
Bandundu	5	6	-	2	5	4	2	24
Equateur	7	6	1	-	4	8	0	26
Orientale	5	7	-	1	11	7	1	32
Total	17	19	1	3	20	19	3	82

2014 Sample

As indicated, the 2014 sample included 1,965 students sampled from 75 schools in three provinces. While most students (41%) came from the Orientale province, and the fewest (26%) from the Equateur province, the student sample was approximately even across grade level and sex. **Table C5** describes the general characteristics of the student sample.

Table C5. General Characteristics of the Overall Student Sample (n = 1,965)

Variable	Number of Students	Percent
Province		
Bandundu	560	29%
Equateur	598	30%
Orientale	807	41%
Grade		
2	994	51%
4	971	49%
Sex		
Female	970	49%
Male	995	51%

Tables C6 and C7 show how students responded to a series of questions targeting SES indicators regarding possessions in the home. Although students were asked more questions than those listed here, the following questions were determined to be of greatest theoretical interest and the most likely to impact student performance. Because student SES has frequently been shown to impact student performance, regression analyses reported later in this report include an SES composite in their models.⁴⁵

Table C6. Student SES Indicators (n = 1,965)

SES Item	Number of Students	Percentage
Radio	1,520	78%
Telephone	1,439	73%
Electricity in the home	487	25%
Television	615	32%

⁴⁵ Having several highly correlated independent variables (such as the SES-related questions shown in *Table C6*), can produce unstable estimates. Reducing independent variables to a smaller number of principal components will yield a better test and/or more stable estimates of regression coefficients. Therefore, for the purposes of regression analyses incorporating SES, the students' answers to the questions regarding possessions in the home were collapsed into a single variable using PCA.

SES Item	Number of Students	Percentage
Refrigerator	167	9%
Indoor toilets	189	10%
Bicycle	1,223	63%
Motorcycle	688	35%
Canoe	255	13%
Motor vehicle	128	7%

Table C7. Select Student Responses to Demographic Questions (n = 1,097)

Province	Number of Students	Total in Province	Percent of Province
Student has reading book in class			
Bandundu	82	558	15%
Equateur	79	592	13%
Orientale	123	806	15%
Someone in the student's home is able to read			
Bandundu	485	558	87%
Equateur	481	592	81%
Orientale	662	806	82%
Student has at least one book at home			
Bandundu	121	558	22%
Equateur	75	592	13%
Orientale	203	806	25%
Student attended kindergarten			
Bandundu	136	558	24%
Equateur	128	591	22%
Orientale	149	806	19%

Province	Number of Students	Total in Province	Percent of Province
If teacher assigns homework, student has someone at home to help with it			
Bandundu	118	558	21%
Equateur	113	591	19%
Orientale	306	803	38%

As indicated in *Table C7*, students across the three provinces were roughly comparable on whether they had a reading book in class and whether someone in the home was able to read. However, students in Equateur appear to be somewhat less likely than students in other provinces to report having a book at home. Students in Orientale appear to be somewhat less likely to have attended kindergarten, although more likely to be helped with homework at home.

Table C8 indicates student self-reports on what language is used in the home. Students had the option of indicating more than one language. Therefore, student responses exceed the total number of students in the sample.

Table C8. Student Indication of Language Spoken in the Home

Province	French Number / %	Kikongo Number / %	Lingala Number / %	Kiswahili Number / %	Other Number / %	Total Student Reports
Bandundu	37 / 4%	420 / 44 %	110 / 11%	2 / 0%	392 / 41%	961
Equateur	82 / 8%	2 / 0%	458 / 47%	4 / 0%	424 / 44%	970
Orientale	53 / 4%	0 / 0%	429 / 33%	316 / 24%	498 / 38%	1,296
Total	172	422	997	322	1,314	3,227

Perhaps not surprisingly, given differences in predominant languages across provinces, more students in Bandundu reported speaking Kikongo at home than any other language, while more students in Equateur and Orientale reported speaking Lingala at home. Use of French in the home ranged from 4% (Bandundu and Orientale) to 8% (Equateur) of student reports.

Teachers were also asked a series of questions at the time of the student testing, and *Tables C9 and C10* describe the general characteristics and relevant survey responses (responses that will be included in subsequent regression analyses) of the teacher sample.

Table C9. General Characteristics of the Teacher Sample (n = 154)

Variable	Number of Teachers	Percent
Province (n = 154)		
Bandundu	44	29%
Equateur	48	31%
Orientale	62	40%
Grade (n = 153)		
1	2	1%
2	73	48%
3	3	2%
4	74	48%
5	1	0%
6	0	0%
Sex (n = 154)		
Female	69	45%
Male	85	55%

As seen in *Table C9*, the teacher sample included more men (55%) than women (45%).

Table C10. Select Teacher Responses to Survey Questions (n = 86)⁴⁶

Province	Number / %	Number / %	Number / %	Number / %	Number / %	Total in Province
What language do you speak and write in the best?	Other/No Response	French	Kikongo	Lingala	Kiswahili	
Bandundu	3 / 7%	27 / 61%	14 / 32%	0 / 0%	0 / 0%	44
Equateur	0 / 0%	31 / 65%	0 / 0%	17 / 35%	0 / 0%	48
Orientale	8 / 13%	31 / 50%	0 / 0%	13 / 21%	10 / 16%	62

⁴⁶Throughout this report, sums reported in tables may not equal 100% due to rounding.

Province	Number / %	Number / %	Number / %	Number / %	Number / %	Total in Province
What is your highest level of education?⁴⁷	D4	D6	PP5	CSP or CAP	G3	
Bandundu	5 / 11%	36 / 82%	0 / 0%	0 / 0%	3 / 7%	44
Equateur	9 / 19%	37 / 77%	1 / 2%	0 / 0%	1 / 2%	48
Orientale	20 / 32%	32 / 52%	0 / 0%	2 / 3%	8 / 13%	62
Other than PAQUED training, over the past two years how often did you receive in-service training in how to teach the French language?	Never	1 Time	2+ Times	No Response		
Bandundu	3 / 7%	8 / 18%	13 / 30%	20 / 46%		44
Equateur	3 / 6%	9 / 18%	10 / 21%	26 / 54%		48
Orientale	5 / 8%	10 / 16%	11 / 18%	36 / 63%		62
How do you characterize your students' competence in French?	Weak	Average	Strong	No Response		
Bandundu	8 / 19%	24 / 55%	10 / 23%	2 / 5%		44
Equateur	10 / 21%	20 / 42%	16 / 36%	2 / 5%		48
Orientale	9 / 15%	34 / 55%	12 / 19%	7 / 11%		62
How do you characterize your students' competence in mathematics?	Weak	Average	Strong	No Response		
Bandundu	3 / 7%	15 / 34%	24 / 55%	2 / 5%		44
Equateur	4 / 8%	19 / 40%	22 / 46%	3 / 6%		48
Orientale	1 / 2%	25 / 40%	29 / 47%	7 / 11%		62

⁴⁷ D4 = 4 years of post-primary education

PP5 = 5 years of post-primary education; a sort of specialized vocational degree, but not a completion of secondary education

D6 = 6 years of post-primary education; completion of secondary education

CSP or CAP = *Cycle spécialisation professionnelle*

G3 = 3 years of post-secondary education; completion of the first half of a course of study in an *institut supérieure*

Province	Number / %	Number / %	Number / %	Number / %	Number / %	Total in Province
How often did you receive a visit by PAQUED personnel this past school year?	1	2-3	4-5	6+	No Response	
Bandundu	2 / 5%	2 / 5%	0 / 0%	2 / 5%	38 / 86%	44
Equateur	1 / 2%	2 / 4%	1 / 2%	2 / 4%	42 / 86%	48
Orientale	4 / 6%	6 / 10%	0 / 0%	4 / 6%	48 / 77%	62
How often did you participate in a teacher exchange forum at the cluster level?	At Least one Time per Trimester	At Least one Time per Month	Other	No Response		
Bandundu	7 / 16%	5 / 11%	1 / 2%	31 / 70%		44
Equateur	4 / 8%	2 / 4%	1 / 2%	41 / 85%		48
Orientale	4 / 6%	6 / 10%	3 / 5%	49 / 79%		62
How often did you participate in a teacher exchange forum at the school level?	At Least one Time per Trimester	At Least one Time per Month	Other	No Response		
Bandundu	4 / 9%	7 / 16%	0 / 0%	33 / 75%		44
Equateur	2 / 4%	10 / 21%	0 / 0%	36 / 75%		48
Orientale	10 / 16%	7 / 11%	1 / 2%	44 / 71%		62
If you participated in teacher exchange forums, what video modules were used? *	IAI Lessons	Teaching Materials	No Modules Used			
Bandundu	8 / 18%	8 / 18%	8 / 18%	20 / 46%		44
Equateur	6 / 13%	4 / 8%	5 / 10%	33 / 69%		48
Orientale	11 / 18%	8 / 13%	6 / 10%	37 / 60%		62
What resources did you receive from the PAQUED project? *	IAI Guide	Reading Activities Guide	Read Aloud Books	Student Texts	Chalk	
Bandundu	18 / 25%	15 / 21%	9 / 13%	9 / 13%	20 / 28%	71
Equateur	10 / 22%	11 / 24%	8 / 17%	5 / 11%	12 / 26%	46
Orientale	23 / 22%	26 / 25%	16 / 15%	15 / 14%	26 / 25%	106

Province	Number / %	Number / %	Number / %	Number / %	Number / %	Total in Province
How many radios did your school receive from the PAQUED project?						
	1	2	3	Other	No Response	
Bandundu	17 / 39%	3 / 7%	0 / 0%	0 / 0%	24 / 55%	44
Equateur	7 / 15%	5 / 10%	0 / 0%	0 / 0%	36 / 13%	48
Orientale	12 / 19%	13 / 21%	1 / 2%	0 / 0%	36 / 58%	62
If you used a PAQUED kit, which one did you use? *						
	Materials Fabrication	School Kit	Class Kit	No Response		
Bandundu	10 / 23%	15 / 34%	14 / 32%	5 / 11%		44
Equateur	5 / 10%	7 / 15%	10 / 21%	26 / 54%		48
Orientale	12 / 19%	18 / 29%	8 / 13%	24 / 39%		62
Did you follow the interactive IAI lessons at your school?						
	Yes	No	No Response			
Bandundu	16 / 36%	5 / 11%	23 / 52%			44
Equateur	16 / 33%	10 / 21%	22 / 46%			48
Orientale	27 / 44%	9 / 15%	26 / 42%			62

*Because multiple responses were allowed per teacher, percentages do not necessarily sum to 100.

As seen in **Table C10**, most teachers across provinces reported speaking and writing best in French; relatively fewer teachers stated that Kikongo, Lingala, or Kiswahili were their strongest languages, although patterns in language preference emerged by province. Most teachers reported D6 as their highest level of education attained, with fewer teachers stating D4. Of the teachers who provided responses to these questions, most reported having attended at least two in-service trainings (other than PAQUED trainings) and having received 2 to 3 visits from PAQUED personnel. Interestingly, teachers were fairly consistent in categorizing their students' proficiencies in French and in mathematics. Among teachers who responded to this question, teachers primarily reported attending exchange forums once per trimester or once per month.

Use of video modules in exchange forums was a bit mixed. Teachers in Bandundu were more likely than their peers in other provinces to report using any video modules, with the exception of teachers in Orientale, 18% of whom reported using IAI lessons. Teachers across the provinces had fairly equal frequency using the various types of PAQUED materials. Teachers in Bandundu reported using each of the kits more than

other teachers. Finally, among teachers who responded to this question, most teachers reported using the IAI interactive lessons.

Tables C11 and *C12* describe the general characteristics and relevant survey responses (responses that will be included in subsequent regression analyses) of the head teacher sample.

Table C11. General Characteristics of the Head Teacher Sample (n = 81)

Variable	Number of Head Teachers	Percent
Province		
Bandundu	24	30%
Equateur	26	32%
Orientale	31	38%
Sex		
Female	12	15%
Male	69	85%

As shown above in *Table C11*, head teachers were predominantly male (85% compared to 15% females). This is consistent across provinces, as shown in *Table C12* below.

Table C12. Select Head Teacher Responses to Survey Questions (n = 42)

Province	Number / %	Number / %	Number / %	Number / %	Number / %	Total in Province
What is the sex of the head teacher?	Female	Male				
Bandundu	5 / 21%	19 / 79%				24
Equateur	2 / 8%	24 / 92%				26
Orientale	5 / 16%	26 / 84%				31

Province	Number / %	Number / %	Number / %	Number / %	Number / %	Total in Province
How often did you receive a visit by PAQUED personnel this past school year?	1	2-3	4-5	6+	No Response	
Bandundu	3 / 13%	4 / 17%	2 / 8%	1 / 4%	14 / 58%	24
Equateur	2 / 8%	2 / 8%	0 / 0%	2 / 8%	20 / 77%	26
Orientale	3 / 10%	3 / 10%	1 / 3%	3 / 10%	21 / 68%	31

No clear pattern regarding PAQUED visits emerged, with the majority of head teachers not responding to this question.

2. EGRA Subtest Outcomes

Table C13 shows overall zero scores and means scores—both including and excluding students with zero scores—at endline in 2014 for each subtask. A student receives a score of zero on a subtask if that student is (1) unable to attempt even one item on the task, or (2) attempts items but does not get any correct. On subtasks where relatively few students scored zero, the difference between means that include these zero scores and those that exclude them is not large. However, on subtasks where a large proportion of students had zero scores, the difference can be substantial, and it is often useful to consider both means when attempting to determine student performance.

Table C13. Overall Percent Zero Scores and Mean Scores by Subtask, Grade and Group at Endline

Subtask	Grade	Group	% Zero Scores	Mean Including Zero Scores	Mean Excluding Zero Scores
Vocabulary	2	Accessible PAQUED	1%	9.43	9.52
		Accessible Control	2%	8.29	8.43
	4	Accessible PAQUED	0%	12.09	12.15
		Accessible Control	0%	11.38	11.39
Listening Comprehension	2	Accessible PAQUED	52%	1.06	2.23
		Accessible Control	58%	0.88	2.12

Subtask	Grade	Group	% Zero Scores	Mean Including Zero Scores	Mean Excluding Zero Scores
Initial Sound Identification	4	Accessible PAQUED	35%	1.83	2.80
		Accessible Control	46%	1.35	2.49
	2	Accessible PAQUED	70%	1.32	4.47
		Accessible Control	76%	1.23	5.03
	4	Accessible PAQUED	56%	2.65	5.98
		Accessible Control	64%	2.11	5.84
Grapheme Identification	2	Accessible PAQUED	30%	9.11	13.06
		Accessible Control	39%	7.58	12.49
	4	Accessible PAQUED	9%	26.16	28.63
		Accessible Control	14%	21.90	25.36
Familiar Word Reading	4	Accessible PAQUED	50%	9.28	18.62
		Accessible Control	58%	7.58	17.95
Invented Word Reading	4	Accessible PAQUED	55%	7.26	16.13
		Accessible Control	57%	7.61	17.54
Oral Reading Fluency	4	Accessible PAQUED	38%	15.68	25.30
		Accessible Control	50%	12.54	24.99
Reading Comprehension	4	Accessible PAQUED	80%	0.45	2.27

Subtask	Grade	Group	% Zero Scores	Mean Including Zero Scores	Mean Excluding Zero Scores
		Accessible Control	82%	0.31	1.67
Dictation	4	Accessible PAQUED	40%	1.02	1.69
		Accessible Control	40%	1.04	1.73

On subtasks with small percentages of zero scores—such as *Vocabulary* and, to a lesser degree, *Grapheme Sound Knowledge*—differences between means are not great. Differences between means that include and exclude students with zero scores are notable on subtasks such as *Listening Comprehension*, *Initial Sound Identification*, the two word reading tasks, and *Reading Comprehension*. Overall, as anticipated, student mean scores are higher in Grade 4 than in Grade 2, for subtasks that were administered in both grades. However, the relative amount of increase suggests an ongoing deficiency in skills even at the higher grade. The presence of substantial proportions of zero scores, even in Grade 4, further indicates student performance that is lower than required to achieve Grade 4 benchmarks.

The following subsections of this chapter provide additional details about each of the EGRA subtasks represented in this report.

Vocabulary

The *Vocabulary* subtask presented children with 20 vocabulary words. Data collectors asked children to identify several body parts and objects as well as to move objects in a variety of directions. As such, this subtask is an assessment of basic French vocabulary, focused on the types of words and concepts found in the environment of students. The type of vocabulary assessed is in the DRC curriculum for French in Grade 1. The curriculum specifies that teaching should include common words and that students should be able to perform a gesture or action based on instructions given by the teacher in French.

Table C14 shows the percent of zero scores and percent of items attempted for the Vocabulary subtask.

Table C14. Vocabulary Zero Scores and Percent Attempted by Grade and Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score		
			2012	2014	2012	2014	2012	2014	
2	Bandundu	Accessible Control	0%	1%	34%	37%	6.90	7.34	
		Accessible PAQUED	1%	0%	39%	46%	7.75	9.23	
	Equateur	Accessible Control	1%	1%	44%	45%	8.76	9.00	
		Accessible PAQUED	1%	4%	44%	44%	8.72	8.74	
	Orientale	Accessible Control	2%	2%	34%	42%	6.73	8.32	
		Accessible PAQUED	1%	0%	46%	50%	9.21	10.09	
	4	Bandundu	Accessible Control	0%	0%	47%	50%	9.36	10.03
			Accessible PAQUED	0%	0%	54%	61%	10.70	12.23
Equateur		Accessible Control	0%	0%	56%	57%	11.24	11.43	
		Accessible PAQUED	0%	2%	52%	57%	10.35	11.38	
Orientale		Accessible Control	0%	0%	50%	62%	10.00	12.31	
		Accessible PAQUED	0%	0%	62%	62%	12.44	12.37	

As shown in *Table C14*, even in 2012, very few students (the largest proportion being 2% in Grade 2 for Equateur) scored zero on this subtask, suggesting that students had at least a minimal level of oral competence in French. Similarly, across grades and provinces, student on average were able to attempt close to, or over, 50% of vocabulary at endline (with the exception of Grade 2 Accessible Control students in Bandundu). Zero scores notwithstanding, mean scores even at endline and for Grade 4 students remain low (even in the highest group, correctly responding to only 12.42, or 62%, of items), suggesting continuing deficiencies in French oral language ability. Certainly by Grade 4, students are expected to have acquired the basic level of oral language competence in French that would allow them to respond to simple vocabulary.

Table C15 reflects a comparison of differences from 2012 to endline (2014) across the Accessible Control and Accessible PAQUED groups. No statistically significant differences emerged for any of the difference-in-differences (D-in-D) results at the $p < 0.008$ threshold, indicating that both groups grew at comparable rates over time.⁴⁸

Table C15. Vocabulary Difference-in-Differences Analyses by Grade, Region, and Group

Grade 2		Baseline 2012		Endline 2014		D-in-D	Effect Size (ES)
Region	Group	Mean	Standard Error (SE)	Mean	SE		
Bandundu	Accessible Control	6.90	0.33	7.34	0.37		
	Accessible PAQUED	7.75	0.50	9.23	0.66	1.04	0.21
Equateur	Accessible Control	8.76	0.68	9.00	0.91		
	Accessible PAQUED	8.72	0.33	8.74	0.46	-0.23	-0.04
Orientale	Accessible Control	6.73	0.39	8.32	0.45		
	Accessible PAQUED	9.21*	0.47	10.09	0.45	-0.70	-0.15
Grade 4		Baseline 2012		Endline 2014		D-in-D	ES
Region	Group	Mean	SE	Mean	SE		
Bandundu	Accessible Control	9.36	0.37	10.03	0.39		
	Accessible PAQUED	10.70	0.41	12.23	0.73	0.86	0.16
Equateur	Accessible Control	11.24	0.59	11.43	0.64		
	Accessible PAQUED	10.35	0.38	11.38	0.46	0.84	0.14

⁴⁸ Type 1 errors in statistics occur when a difference is thought to exist where one does not. (Put another way, a type 1 error is the rejection of the null hypothesis when it is actually true.) Due to the large number of difference-in-differences comparisons conducted for this section (Chapter C, section 1) of the report, the Bonferroni correction was used to determine the threshold of significance for the various analyses. This section contains 6 tests of D-in-D for the subtasks administered to both grades and 3 tests of D-in-D for the subtasks administered to Grade 4 alone. Applying the Bonferroni correction to these analyses results in $p < 0.05 / (6) = 0.008$ for Vocabulary, Initial Sound Identification, Listening Comprehension, and Grapheme Recognition subtasks and $p < (0.05) / (3) = 0.017$ for the Familiar Word, Invented Word, Oral Reading Fluency, Reading Comprehension, and Dictation subtasks. Please see *Annex 1* for a fuller discussion of the Bonferroni correction and the analytical families used in this report.

Grade 4		Baseline 2012		Endline 2014			
Region	Group	Mean	SE	Mean	SE	D-in-D	ES
Orientale	Accessible Control	10.00	0.62	12.31	0.41		
	Accessible PAQUED	12.44	0.55	12.37	0.35	-2.38	-0.43

Given that students, overall, have not mastered this level of French vocabulary, it is interesting to explore levels of competence between girls and boys. *Table C16* provides a comparison of mean scores on this subtask by grade, province, and sex.

Table C16. Comparison of Vocabulary Mean Scores by Grade, Province, Group, and Sex

Grade 2		Baseline 2012		Endline 2014			
Region	Group	Sex	Mean	SE	Mean	SE	
Bandundu	Accessible Control	Male	6.66	0.42	8.35	0.52	
		Female	7.14	0.35	6.46	0.42	
	Accessible PAQUED	Male	7.75	0.46	9.16	0.45	
		Female	7.76	0.60	9.34	1.10	
Equateur	Accessible Control	Male	8.47	0.62	9.38	1.01	
		Female	9.13	0.85	8.63	0.91	
	Accessible PAQUED	Male	8.73	0.42	9.41	0.42	
		Female	8.72	0.35	8.22	0.53	
Orientale	Accessible Control	Male	6.95	0.45	8.06	0.49	
		Female	6.53	0.41	8.58	0.57	
	Accessible PAQUED	Male	9.11	0.51	10.18	0.57	
		Female	9.30	0.48	9.99	0.59	
Grade 4		Baseline 2012		Endline 2014			
Region	Group	Sex	Mean	SE	Mean	SE	
Bandundu	Accessible Control	Male	9.66	0.70	10.65	0.38	
		Female	9.09	0.25	9.43	0.55	
		Male	10.90	0.45	12.90	0.51	

Grade 4		Baseline 2012			Endline 2014	
Region	Group	Sex	Mean	SE	Mean	SE
Equateur	Accessible PAQUED	Female	10.47	0.53	11.62	0.92
		Male	11.72	0.62	11.03	0.58
	Accessible Control	Female	10.59	0.65	11.92	0.85
		Male	10.17	0.47	11.83	0.60
	Accessible PAQUED	Female	10.51	0.41	10.98	0.61
		Male	10.84	0.64	12.82	0.42
Orientale	Accessible Control	Female	9.22	0.65	11.83	0.54
		Male	12.66	0.52	12.54	0.46
	Accessible PAQUED	Female	12.20	0.64	12.18	0.40
		Male				

The number of hypotheses tests present in this chapter require a more conservative threshold to avoid type 1 errors.⁴⁹ At the $p < 0.002$ level no statistically significant differences emerged between girls and boys.

Exploring distributions of scores across the range of possible scores (0–20 items) again shows that relatively few students scored zero, but it also demonstrates that relatively few achieved perfect scores on this subtask. *Figures C1* and *C2* illustrate these distributions by grade.

⁴⁹ Due to the large number of means comparisons conducted for this section (Chapter C, section 1) of the report, the Bonferroni correction was used to determine the threshold of significance for these analyses. This section contains 24 tests for the comparison of means for the subtasks administered to both grades and 12 tests for the comparison of means for the subtasks administered to Grade 4 alone. Applying the Bonferroni correction to these analyses results in $p < 0.05 / (24) = 0.002$ for Vocabulary, Initial Sound Identification, Listening Comprehension, and Grapheme Recognition subtasks and $p < (0.05) / (12) = 0.004$ for the Familiar Word, Invented Word, Oral Reading Fluency, Reading Comprehension, and Dictation subtasks.

Figure C1. Grade 2 Scores on Vocabulary at Endline by Province

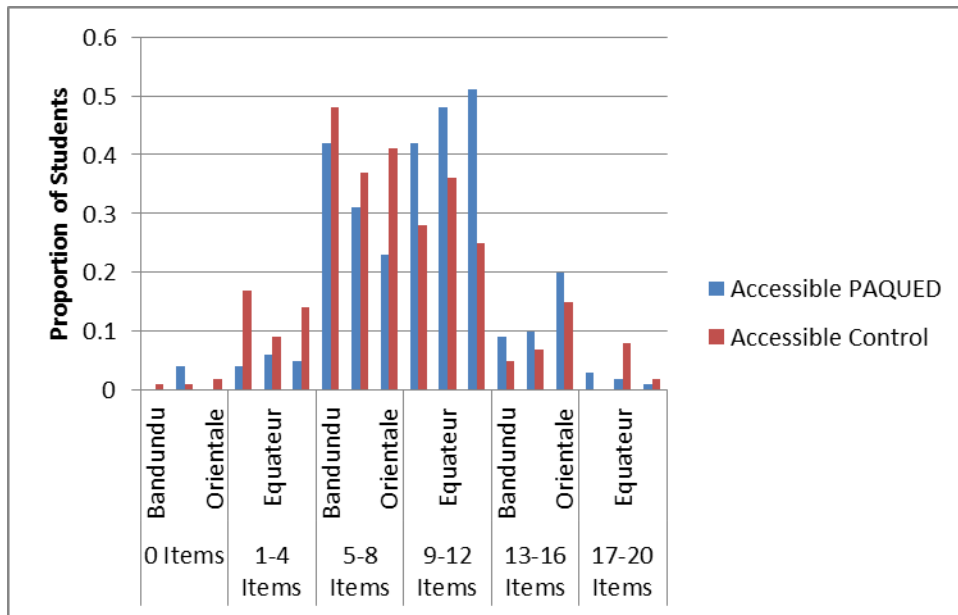
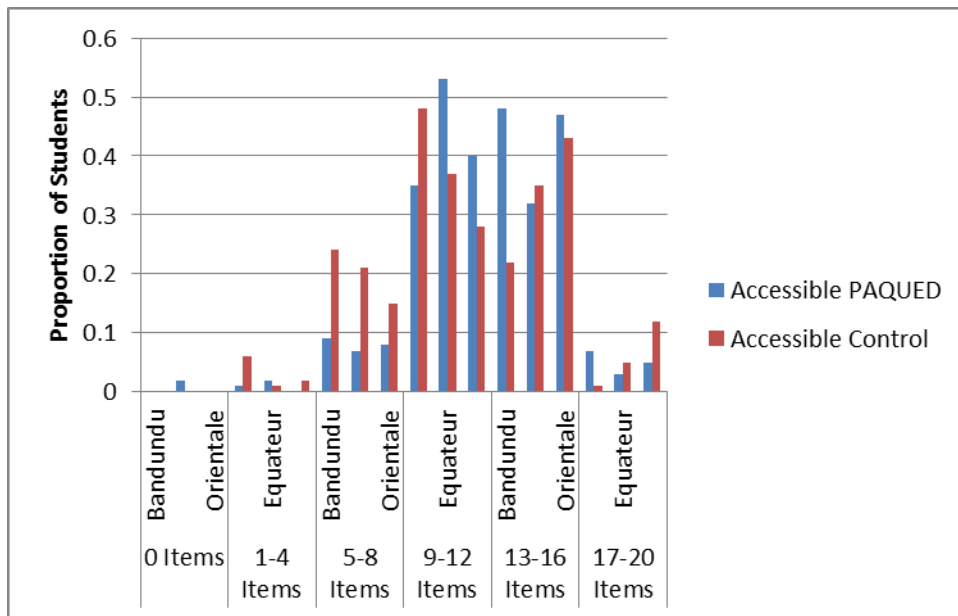


Figure C2. Grade 4 Scores on Vocabulary at Endline by Province



What is encouraging is that although the distribution for Grade 2 shows most students scored in the five to 12 item range, the distribution is shifted up for Grade 4, where most students scored between nine and 16 items.

Initial Sound Identification

In the Initial Sound Identification subtask, students listened to individual words, such as “*sack*,” and were asked to identify the first sound, or phoneme, of that word (in this case, /s/). This subtask was comprised of 10 items, for a maximum possible score of 10. The first five items (*sac*, *jour*, *date*, *lac*, and *car*) were presented to all students. Students who were able to identify the first sound of at least one of these words were presented the remaining five items (*balle*, *tour*, *par*, *vol*, and *fil*).

Table C17 shows the percent of zero scores and percent of items attempted for the *Initial Sound Identification* subtask.

Table C17. Initial Sound Identification Zero Scores and Percent Attempted by Grade and Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2012	2014	2012	2014	2012	2014
2	Bandundu	Accessible Control	78%	75%	12%	12%	1.03	1.20
		Accessible PAQUED	76%	68%	11%	13%	1.09	1.32
	Equateur	Accessible Control	64%	53%	17%	27%	1.66	2.73
		Accessible PAQUED	60%	54%	17%	28%	1.66	2.79
	Orientale	Accessible Control	65%	93%	7%	1%	0.70	0.11
		Accessible PAQUED	65%	83%	11%	5%	1.03	0.50
4	Bandundu	Accessible Control	62%	59%	20%	22%	2.00	2.22
		Accessible PAQUED	47%	45%	29%	32%	2.84	3.20
	Equateur	Accessible Control	37%	48%	35%	34%	3.43	3.40
		Accessible PAQUED	42%	39%	36%	43%	3.54	4.31
	Orientale	Accessible Control	54%	81%	15%	10%	1.50	0.95
		Accessible PAQUED	46%	81%	19%	8%	1.81	0.83

As shown in *Table C17*, the Initial Sound Identification subtask was difficult for students in both 2012 and 2014. Even in Grade 4, percentages of zero scores at endline ranged from 39% in Equateur to 81% in Orientale. Similarly, students, in general, attempted relatively few items, particularly in Orientale (% attempted in Grade 2 = 1%, and in Grade 4 = 8%). Mean scores were also low. The highest level of competence on this task was among Grade 4 students at endline in Orientale, who identified, on average, less than one item.

Table C18 reflects a comparison of differences from 2012 to 2014 across the Accessible Control and Accessible PAQUED groups. No statistically significant differences emerged for any of the D-in-D results at the $p < 0.008$ threshold, indicating that both groups grew at comparable rates over time.

Table C18. Initial Sound Identification Difference-in-Differences Analyses by Grade, Region, and Group

Grade 2 Region	Group	Baseline 2012		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
Bandundu	Accessible Control	1.03	0.29	1.20	0.24		
	Accessible PAQUED	1.09	0.20	1.32	0.37	0.06	0.01
Equateur	Accessible Control	1.66	0.35	2.73	0.52		
	Accessible PAQUED	1.66	0.25	2.79	0.61	0.05	0.01
Orientale	Accessible Control	0.70	0.12	0.11	0.05		
	Accessible PAQUED	1.03	0.19	0.50	0.25	0.05	0.03
Grade 4 Region	Group	Mean	SE	Mean	SE	D-in-D	ES
Bandundu	Accessible Control	2.00	0.34	2.22	0.31		
	Accessible PAQUED	2.84	0.37	3.20	0.40	0.14	0.03
Equateur	Accessible Control	3.43	0.45	3.40	0.42		
	Accessible PAQUED	3.54	0.51	4.31	0.40	0.80	0.12
Orientale	Accessible Control	1.50	0.23	0.95	0.22		

Grade 4 Region	Group	Baseline 2012		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
	Accessible PAQUED	1.81	0.23	0.83	0.28	-0.44	-0.12

Table C19 provides a comparison of mean scores on this subtask by grade, province, group and sex.

Table C19. Comparison of Initial Sound Identification Mean Scores by Grade, Province, Group, and Sex

Grade 2 Region	Group	Sex	Baseline 2012		Endline 2014	
			Mean	SE	Mean	SE
Bandundu	Accessible Control	Male	1.18	0.43	1.64	0.43
		Female	0.89	0.22	0.81	0.15
	Accessible PAQUED	Male	1.03	0.21	1.34	0.23
		Female	1.15	0.34	1.28	0.73
Equateur	Accessible Control	Male	1.52	0.43	2.80	0.64
		Female	1.84	0.57	2.66	0.57
	Accessible PAQUED	Male	1.81	0.36	2.59	0.50
		Female	1.52	0.30	2.94	0.77
Orientale	Accessible Control	Male	0.49	0.17	0.11	0.08
		Female	0.90	0.15	0.12	0.04
	Accessible PAQUED	Male	1.15	0.26	0.82	0.41
		Female	0.93	0.18	0.11	0.06
Grade 4 Region	Group	Sex	Baseline 2012		Endline 2014	
Mean			SE	Mean	SE	
Bandundu	Accessible Control	Male	2.27	0.44	2.96	0.47
		Female	1.75	0.36	1.50	0.34
	Accessible PAQUED	Male	2.81	0.43	3.93	0.45
		Female	2.88	0.48	2.53	0.47
Equateur		Male	3.44	0.53	3.61	0.45

Grade 4 Region	Group	Sex	Baseline 2012 Mean	Baseline 2012 SE	Endline 2014 Mean	Endline 2014 SE
Orientale	Accessible Control	Female	3.43	0.47	3.14	0.61
		Male	3.51	0.64	4.74	0.61
	Accessible PAQUED	Female	3.58	0.47	3.93	0.36
		Male	1.62	0.28	0.99	0.31
	Accessible Control	Female	1.38	0.27	0.91	0.25
		Male	2.00	0.31	0.89	0.38
Accessible PAQUED	Female	1.61	0.27	0.75	0.20	

No statistically significant differences between sexes emerged using the $p < 0.002$ threshold.

Exploring distributions of scores across the range of possible scores (0–10 items) again shows the difficulty that students had with this subtask. *Figures C3* and *C4* illustrate these distributions by grade.

Figure C3. Grade 2 Scores on Initial Sound Identification at Endline by Province

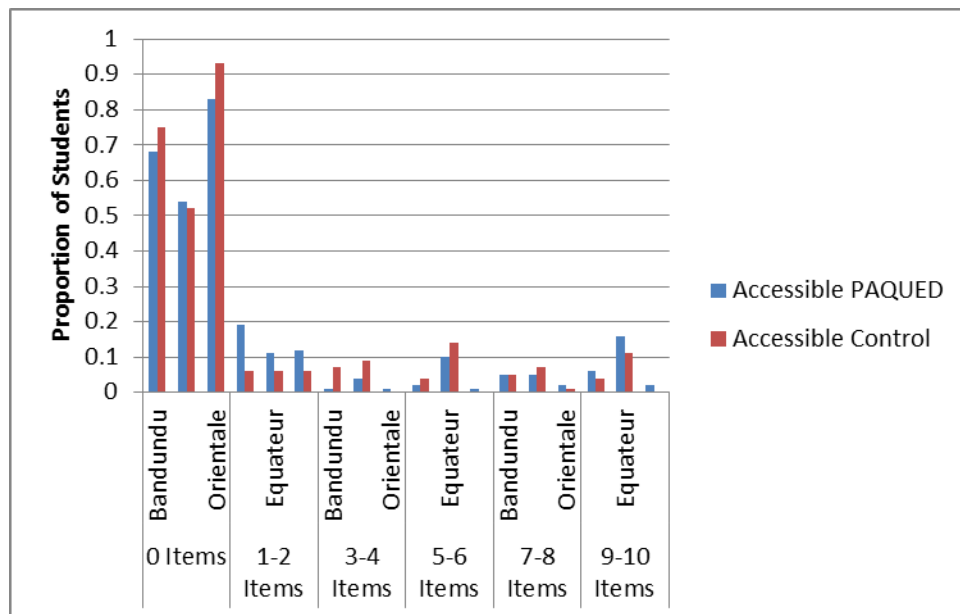
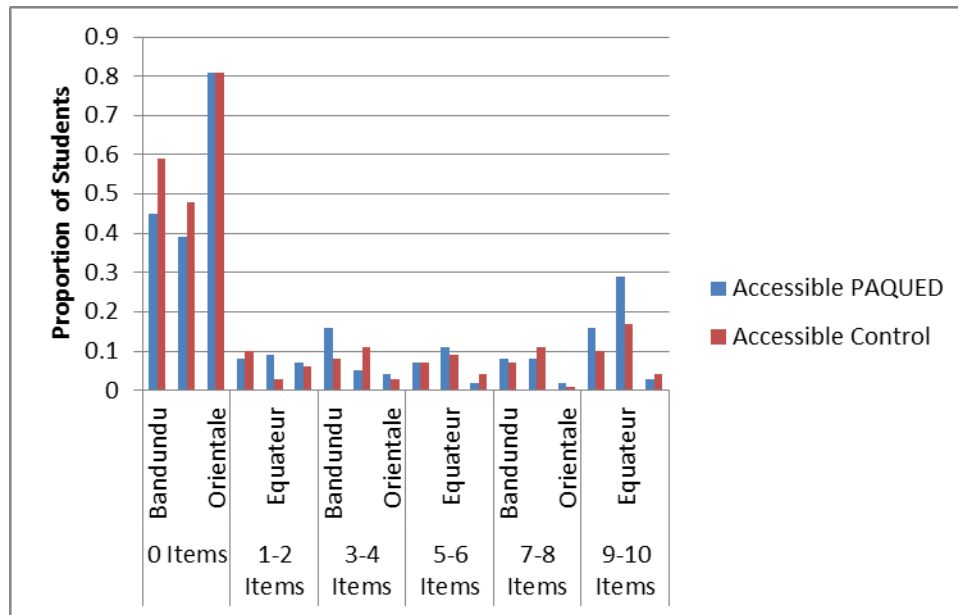


Figure C4. Grade 4 Scores on *Initial Sound Identification* at Endline by Province



As illustrated, large percentages of students in both grades scored zero on this subtask at endline. Scores in both grades tended to be fairly evenly distributed across the other score categories, with a bump in the highest category of scores (9–10 items correct) for Bandundu and Equateur students in Grade 4.

Listening Comprehension

For the *Listening Comprehension* subtask, Grade 2 and Grade 4 students listened to a short passage and were asked five questions that assessed their basic comprehension of that passage. Scores reported for this subtask include percentages of students able to answer the comprehension questions, based upon the number of students who attempted to answer each question.

Table C20 shows the percent of zero scores and percent of items attempted for the *Listening Comprehension* subtask.

Table C20. Listening Comprehension Zero Scores and Percent Attempted by Grade and Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2012	2014	2012	2014	2012	2014
2	Bandundu	Accessible Control	84%	74%	7%	13%	0.37	0.64
		Accessible PAQUED	78%	48%	8%	27%	0.41	1.37
	Equateur	Accessible Control	75%	52%	10%	21%	0.49	1.03
		Accessible PAQUED	71%	67%	14%	12%	0.72	0.61
	Orientale	Accessible Control	83%	53%	4%	18%	0.21	0.92
		Accessible PAQUED	59%	51%	14%	17%	0.71	0.87
4	Bandundu	Accessible Control	70%	57%	12%	19%	0.61	0.95
		Accessible PAQUED	57%	34%	17%	39%	0.85	1.96
	Equateur	Accessible Control	49%	48%	30%	24%	1.49	1.18
		Accessible PAQUED	53%	50%	22%	21%	1.09	1.05
	Orientale	Accessible Control	55%	35%	16%	36%	0.78	1.80
		Accessible PAQUED	26%	25%	35%	43%	1.75	2.17

As might be anticipated, given relatively low performance on the *Vocabulary* subtask, students appeared to have struggled with the Listening Comprehension task. In Grade 2 at endline, across treatment groups, the percentage of students with zero scores ranges from 48% to 74%. Percentages of zero scores in Grade 4 are lower, but still high, ranging from 25% to 57%. That so many students were unable to correctly respond to even one question asked demonstrates a low overall proficiency with the French language. Even in Grade 4, there was no group of students that were able to correctly respond to two comprehension questions.

Table C21 reflects a comparison of differences from 2012 to 2014 across the Accessible Control and Accessible PAQUED groups. No statistically significant differences emerged

for any of the D-in-D results using the $p < 0.008$ threshold, indicating that both groups grew at comparable rates over time.

Table C21. Listening Comprehension Difference-in-Differences Analyses by Grade, Region, and Group

Grade 2		Baseline 2012		Endline 2014			
Region	Group	Mean	SE	Mean	SE	D-in-D	ES
Bandundu	Accessible Control	0.37	0.10	0.64	0.12		
	Accessible PAQUED	0.41	0.09	1.37	0.27	0.69	0.36
Equateur	Accessible Control	0.49	0.15	1.03	0.28		
	Accessible PAQUED	0.72	0.16	0.61	0.06	-0.66	-0.33
Orientale	Accessible Control	0.21	0.06	0.92	0.16		
	Accessible PAQUED	0.71	0.15	0.87	0.14	-0.54	-0.36
Grade 4		Baseline 2012		Endline 2014			
Region	Group	Mean	SE	Mean	SE	D-in-D	ES
Bandundu	Accessible Control	0.61	0.12	0.95	0.12		
	Accessible PAQUED	0.85	0.10	1.96	0.30	0.78	0.39
Equateur	Accessible Control	1.49	0.41	1.18	0.30		
	Accessible PAQUED	1.09	0.21	1.05	0.11	0.27	0.09
Orientale	Accessible Control	0.78	0.12	1.80	0.21		
	Accessible PAQUED	1.75	0.27	2.17	0.38	-0.60	-0.25

Table C22 provides a comparison of mean scores on this subtask by grade, province, group, and sex.

Table C22. Comparison of Listening Comprehension Mean Scores by Grade, Province, Group, and Sex

Grade 2 Region	Group	Sex	Baseline 2012		Endline 2014	
			Mean	SE	Mean	SE
Bandundu	Accessible Control	Male	0.46	0.12	0.76	0.18
		Female	0.28	0.10	0.53	0.10
	Accessible PAQUED	Male	0.38	0.09	1.46	0.26
		Female	0.43	0.11	1.23	0.31
Equateur	Accessible Control	Male	0.46	0.11	1.11	0.38
		Female	0.54	0.23	0.95	0.23
	Accessible PAQUED	Male	0.77	0.22	0.72	0.10
		Female	0.68	0.14	0.52	0.13
Orientale	Accessible Control	Male	0.28	0.09	0.78	0.15
		Female	0.15	0.05	1.06	0.22
	Accessible PAQUED	Male	0.75	0.17	0.94	0.15
		Female	0.67	0.16	0.79	0.16
Grade 4 Region	Group	Sex	Baseline 2012 Mean	SE	Endline 2014 Mean	SE
Bandundu	Accessible Control	Male	0.75	0.15	1.01	0.18
		Female	0.48	0.13	0.89	0.15
	Accessible PAQUED	Male	0.96	0.10	1.91	0.24
		Female	0.72	0.12	2.02	0.38
Equateur	Accessible Control	Male	1.54	0.47	0.89	0.23
		Female	1.42	0.36	1.52	0.40
	Accessible PAQUED	Male	1.12	0.22	0.95	0.16
		Female	1.07	0.23	1.13	0.17

Grade 4 Region	Group	Sex	Baseline 2012 Mean	Baseline 2012 SE	Endline 2014 Mean	Endline 2014 SE
Orientale	Accessible Control	Male	0.94	0.16	2.13	0.23
		Female	0.62	0.14	1.47	0.30
	Accessible PAQUED	Male	1.79	0.28	2.30	0.41
		Female	1.70	0.29	2.02	0.37

No statistically significant differences between sexes emerged using the $p < 0.002$ threshold.

Exploring distributions of scores across the range of possible scores (0–5 items) again shows the difficulty that students had with this subtask. **Figures C5** and **C6** illustrate these distributions by grade.

Figure C5. Grade 2 Scores on *Listening Comprehension* at Endline by Province

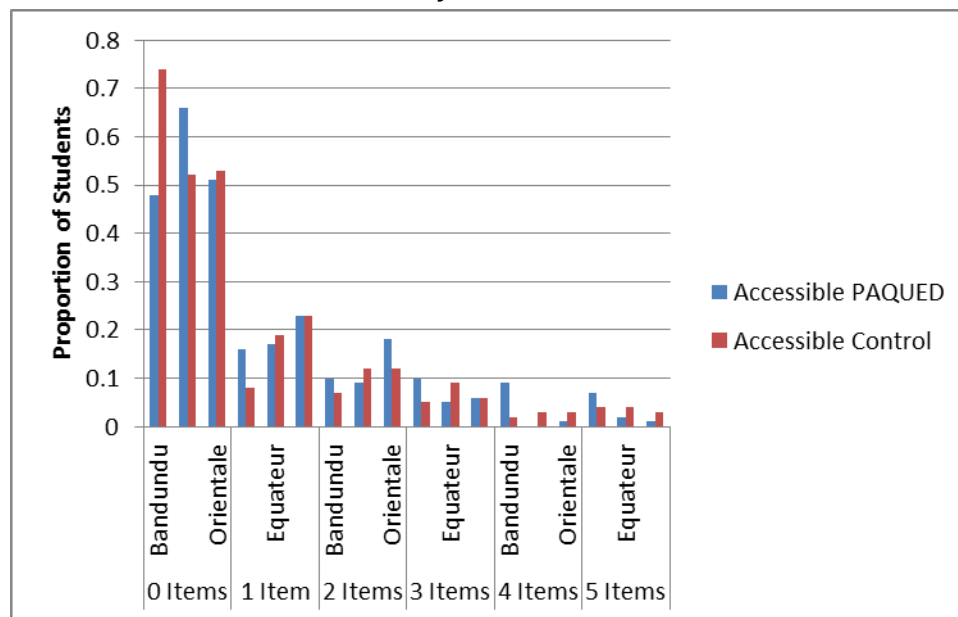
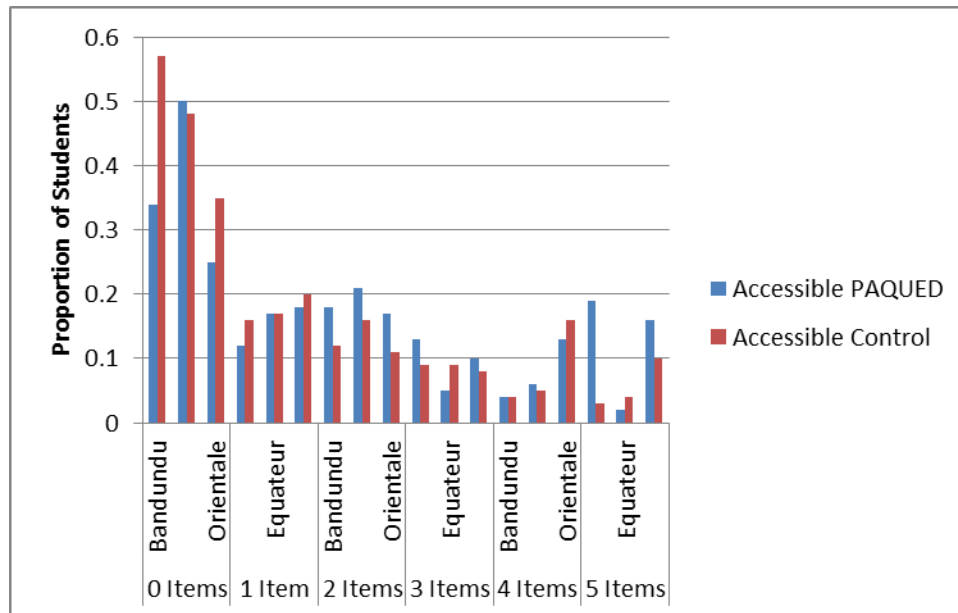


Figure C6. Grade 4 Scores on *Listening Comprehension* at Endline by Province



As illustrated in *Figures C5* and *C6*, a large percentage of students scored zero on this subtask at endline in Grade 2. Within Grade 4, scores were more evenly distributed across the score point, but were still lower than needed to demonstrate proficiency with the French language.

Grapheme Recognition

In the *Grapheme Recognition* subtask, students were presented with a 100-item chart containing the letters of the alphabet as well as common two-letter graphemes in random order and were required to produce the sounds for as many as possible within one minute. This task was discontinued prior to the end of the minute for students who were unable to produce any of the sounds for the 10 letters in the first row. Scores reported for this subtask include percentages of students able to correctly generate numbers of graphemes and the number of grapheme sounds that students could correctly generate within one minute.

Table C23 shows the percent of zero scores and percent of items attempted for the *Grapheme Recognition* subtask.

Table C23. Grapheme Recognition Zero Scores, Percent Attempted, and Mean Scores by Grade and Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score		
			2012	2014	2012	2014	2012	2014	
2	Bandundu	Accessible Control	24%	33%	27%	22%	6.45	5.42	
		Accessible PAQUED	17%	19%	34%	30%	8.73	9.26	
	Equateur	Accessible Control	32%	39%	27%	32%	6.74	11.34	
		Accessible PAQUED	29%	55%	28%	22%	6.10	8.65	
	Orientale	Accessible Control	18%	43%	30%	25%	6.64	6.07	
		Accessible PAQUED	10%	33%	35%	30%	7.41	9.16	
	4	Bandundu	Accessible Control	6%	19%	53%	45%	17.60	13.80
			Accessible PAQUED	5%	5%	57%	58%	19.98	26.09
Equateur		Accessible Control	8%	15%	65%	54%	24.79	23.17	
		Accessible PAQUED	8%	18%	60%	55%	20.29	24.42	
Orientale		Accessible Control	5%	8%	63%	62%	22.65	26.77	
		Accessible PAQUED	3%	8%	67%	62%	25.18	27.39	

As indicated before, for this subtask, students were given a grid of 100 letters/graphemes for which they were to generate either sounds or names within 60 seconds. Even generating sounds at a rate of one sound per second would result in mean score of 60. At endline in Grade 4, across groups, student means ranged from 13.80 to a high of 27.39, indicating that the highest performing students generated letter sounds at a rate of one sound every two seconds. Grade 2 student means were lower. Additionally, the zero scores were quite high, particularly in Equateur.

Table C24 reflects a comparison of differences from 2012 to 2014 across the Accessible Control and Accessible PAQUED groups. No statistically significant differences emerged

for any of the D-in-D results using the $p < 0.008$ threshold, indicating that both groups grew at comparable rates over time.

Table C24. Grapheme Recognition Difference-in-Differences Analyses by Grade, Region, and Group

Grade 2 Region	Group	Baseline 2012		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
Bandundu	Accessible Control	6.45	0.76	5.42	0.87		
	Accessible PAQUED	8.73	1.07	9.26	1.63	1.56	0.13
Equateur	Accessible Control	6.74	0.95	11.34	2.98		
	Accessible PAQUED	6.10	0.74	8.65	3.44	-2.05	-0.10
Orientale	Accessible Control	6.64	0.88	6.07	0.98		
	Accessible PAQUED	7.41	0.87	9.16	1.54	2.33	0.20
Grade 4 Region	Group	Baseline 2012 Mean	SE	Endline 2014 Mean	SE	D-in-D	ES
Bandundu	Accessible Control	17.60	0.82	13.80	0.95		
	Accessible PAQUED	19.98	1.47	26.09	4.50	9.91	0.50
Equateur	Accessible Control	24.79	2.67	23.17	3.26		
	Accessible PAQUED	20.29	1.61	24.42	3.45	5.75	0.18
Orientale	Accessible Control	22.65	1.27	26.77	1.58		
	Accessible PAQUED	25.18	2.13	27.39	2.45	-1.90	-0.07

Table C25 provides a comparison of mean scores on this subtask by grade, province, and sex.

Table C25. Comparison of Grapheme Recognition Mean Scores by Grade, Province, and Sex

Grade 2 Region	Group	Sex	Baseline 2012		Endline 2014	
			Mean	SE	Mean	SE
Bandundu	Accessible Control	Male	6.40	1.08	7.27	1.41
		Female	6.50	0.84	3.80	0.55
	Accessible PAQUED	Male	8.38	1.35	10.89	1.68
		Female	9.10	1.17	6.73	2.34
Equateur	Accessible Control	Male	6.83	0.93	12.24	2.99
		Female	6.62	1.73	10.44	3.46
	Accessible PAQUED	Male	6.81	0.95	9.76	3.52
		Female	5.40	0.92	7.77	3.50
Orientale	Accessible Control	Male	8.66	1.25	5.93	1.03
		Female	4.77	0.66	6.20	1.27
	Accessible PAQUED	Male	8.22	1.19	13.35 *	1.93
		Female	6.67	0.82	4.17	0.75
Grade 4 Region	Group	Sex	Baseline 2012 Mean	SE	Endline 2014 Mean	SE
Bandundu	Accessible Control	Male	20.18	1.48	16.83	1.48
		Female	15.24	1.20	10.84	1.23
	Accessible PAQUED	Male	20.34	1.76	27.35	2.37
		Female	19.54	2.23	24.93	7.09
Equateur	Accessible Control	Male	25.49	3.12	23.09	2.86
		Female	23.87	2.78	23.27	4.47
	Accessible PAQUED	Male	22.11	1.86	29.50	4.91
		Female	18.70	1.92	19.89	2.75
Orientale	Accessible Control	Male	26.40	1.99	29.90	2.25
		Female	19.14	1.63	23.75	1.75
		Male	27.64	2.12	29.72	4.16

Grade 4 Region	Group	Sex	Baseline 2012 Mean	Baseline 2012 SE	Endline 2014 Mean	Endline 2014 SE
	Accessible PAQUED	Female	22.57	3.17	24.77	1.75

* significant at $p < 0.002$

One statistically-significant difference between the sexes emerged using the $p < 0.002$ threshold; boys in Orientale’s Accessible PAQUED schools outperformed girls at endline in 2014. (There was no such difference at baseline in 2012.)

Exploring distributions of scores across the range of possible scores (0–100 items) again shows the difficulty that students had with this subtask. **Figures C7** and **C8** illustrate these distributions by grade.

Figure C7. Grade 2 Scores on Grapheme Recognition at Endline by Province

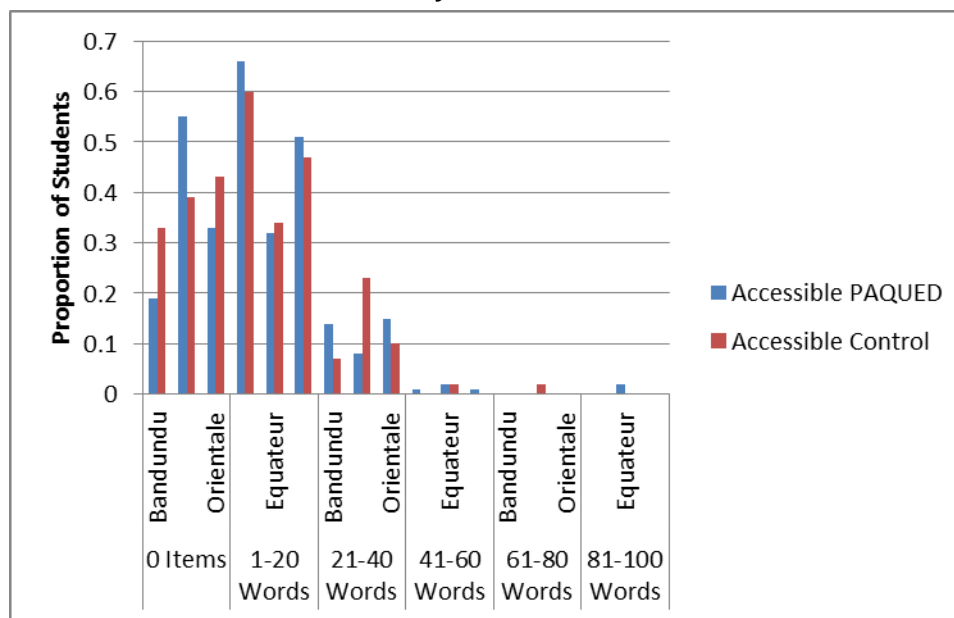
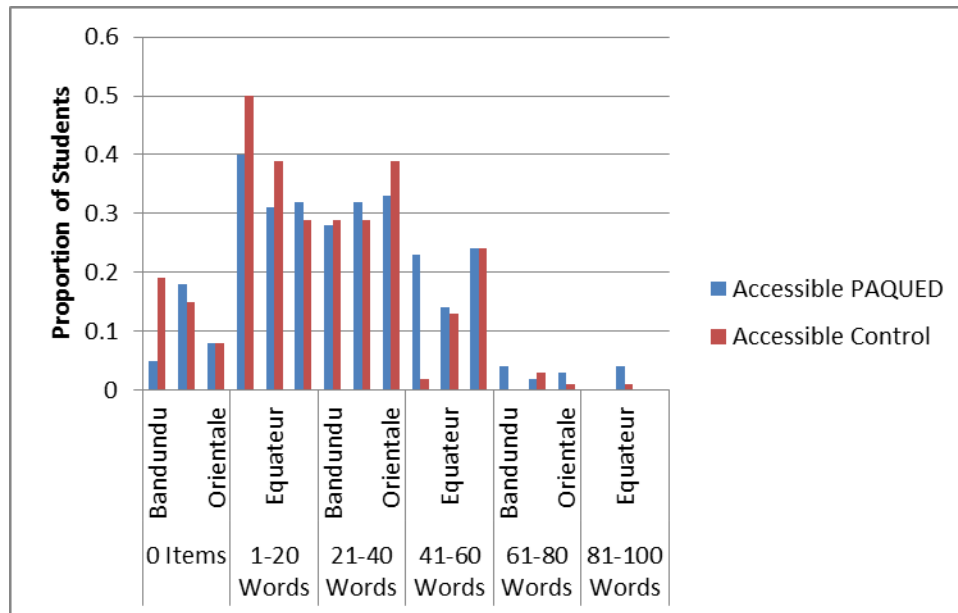


Figure C8. Grade 4 Scores on *Grapheme Recognition* at Endline by Province



As illustrated in *Figures C7* and *C8*, a large percentage of Grade 2 students scored between zero and 20 letters per minute on this subtask at endline, with only a few students scoring higher than 40. Within Grade 4, fewer zero scores were observed, with the majority of students scoring between one and 60 letters per minute. This distribution shows a promising trend, although many students appear to still struggle with letter knowledge and have not gained full automaticity with this skill.

Familiar Word Reading

The *Familiar Word Reading* subtask was administered to only students in Grade 4. In this task, students were shown a chart of 50 familiar words (e.g., *tu* and *ami*) and were required to read as many words as they could within one minute. This subtask was discontinued before the end of one minute for students who were unable to read any of the first five words. Scores reported for this subtask include percentages of students able to correctly identify words and the number of words that students could correctly identify within one minute (i.e., cwpm).

Table C26 shows the percent of zero scores and percent of items attempted for the *Familiar Word Reading* subtask.

Table C26. Familiar Word Reading Zero Scores and Percent Attempted for Grade 4 by Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2012	2014	2012	2014	2012	2014
4	Bandundu	Accessible Control	56%	74%	19%	13%	4.38	2.56
		Accessible PAQUED	42%	53%	28%	24%	6.84	7.25
	Equateur	Accessible Control	34%	57%	35%	26%	8.67	7.90
		Accessible PAQUED	38%	59%	33%	24%	8.50	7.23
	Orientale	Accessible Control	28%	47%	41%	38%	10.93	11.00
		Accessible PAQUED	28%	40%	43%	40%	12.73	13.36

As shown in **Table C26**, reading familiar words appeared to be a challenging task for these students, with up to 74% of students at endline scoring zero. Correspondingly, students across the provinces attempted fewer than half of the possible words in both 2012 and 2014. Mean scores at endline were also low, ranging from 2.56 words per minute in Bandundu to 13.36 words per minute in Orientale.

Table C27 reflects a comparison of differences from 2012 to 2014 across the Accessible Control and Accessible PAQUED groups. No statistically significant differences emerged for any of the D-in-D results using the $p < 0.017$ threshold, indicating that both groups grew at comparable rates over time.

Table C27. Familiar Word Reading Difference-in-Differences Analyses by Region and Group

Region	Group	Baseline 2012		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
Bandundu	Accessible Control	4.38	0.50	2.56	0.55		
	Accessible PAQUED	6.84	1.15	7.25	2.51	2.22	0.19
Equateur	Accessible Control	8.67	1.41	7.90	2.68		
	Accessible PAQUED	8.50	1.28	7.23	1.59	-0.49	-0.02
Orientale	Accessible Control	10.93	1.41	11.00	1.25		
	Accessible PAQUED	12.73	2.16	13.36	2.91	0.56	0.03

Table C28 provides a comparison of mean scores on this subtask for Grade 4 by province, group, and sex.

Table C28. Comparison of Familiar Word Reading Mean Scores for Grade 4 by Province, Group, and Sex

Region	Group	Sex	Baseline 2012		Endline 2014	
			Mean	SE	Mean	SE
Bandundu	Accessible Control	Male	5.98	0.93	4.11	1.05
		Female	2.92	0.60	1.05	0.37
	Accessible PAQUED	Male	7.35	1.33	7.09	1.71
		Female	6.23	1.56	7.38	4.03
Equateur	Accessible Control	Male	9.12	1.50	7.11	2.17
		Female	8.08	1.58	8.85	3.66
	Accessible PAQUED	Male	9.29	1.56	10.00	2.42
		Female	7.81	1.76	4.77	1.07
Orientale	Accessible Control	Male	14.43 *	1.97	12.69	1.69
		Female	7.65	1.32	9.38	1.50

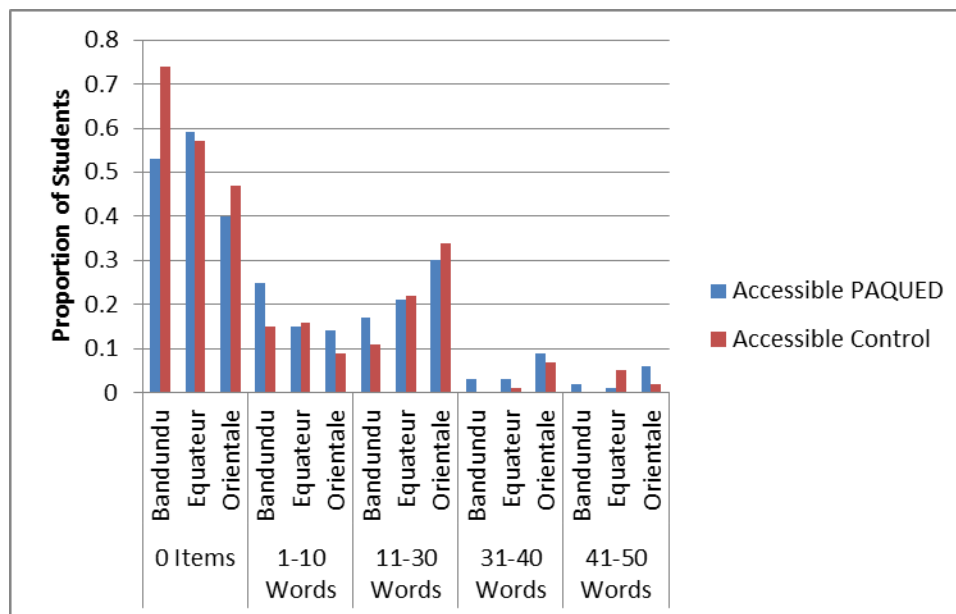
Region	Group	Sex	Baseline 2012		Endline 2014	
			Mean	SE	Mean	SE
	Accessible PAQUED	Male	14.06	1.98	16.80	3.93
		Female	11.32	3.09	9.49	2.47

* significant at $p < 0.004$ threshold

One statistically significant difference between the sexes emerged using the $p < 0.004$ threshold; boys in Orientale Accessible Control schools outperformed girls at baseline in 2012. (No such difference remained at endline in 2014.)

Exploring distributions of scores across the range of possible scores (0–50 items) again shows the difficulty that students had with this subtask. **Figure C9** illustrates these distributions for Grade 4.

Figure C9. Grade 4 Scores on *Familiar Word Reading* at Endline by Province



As illustrated in **Figure C9**, a large percentage (over 50% for Bandundu and Orientale provinces, and over 40% for Equateur) of students scored zero words per minute on this subtask, although scores did range between one and 30 words per minute. Very few students were able to read more than 30 words per minute, suggesting a lack of automaticity with this skill.

Invented Word Reading

The *Invented Word Reading* subtask was also administered to only students in Grade 4. In this subtask, students were given a chart of 50 invented words (e.g., *tal* and *vor*), and were required to read as many words as they could within one minute. This subtask was discontinued before the end of the minute for students who were unable to read any of the first five words. Scores reported for this subtask include percentages of students able to correctly identify non-words and the number of non-words that students could correctly identify within one minute (correct non-words per minute [cnonwpm]).

Table C29 shows the percent of zero scores and percent of items attempted for the *Invented Word Reading* subtask.

Table C29. Invented Word Reading Zero Scores and Percent Attempted for Grade 4 by Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2012	2014	2012	2014	2012	2014
4	Bandundu	Accessible Control	64%	77%	15%	11%	3.39	1.99
		Accessible PAQUED	56%	59%	21%	19%	5.20	5.55
	Equateur	Accessible Control	39%	54%	32%	27%	7.53	10.65
		Accessible PAQUED	51%	65%	25%	20%	6.33	5.90
	Orientale	Accessible Control	35%	44%	34%	38%	8.38	9.21
		Accessible PAQUED	38%	42%	35%	38%	9.48	10.47

As with the *Familiar Word Reading* subtask, student scores on the *Invented Word Reading* subtask were quite low. Across the various groups, zero scores across the provinces ranged from 42% to 77% at endline (interestingly, across all provinces, proportions of zero scores were lower in 2012). In general, students in Bandundu and Equateur provinces had particularly low percentages of items attempted and mean scores. However, even at endline in Bandundu province students, on average, were able to read fewer than 10 words in one minute. This rate suggests ongoing difficulties with the skill of decoding.

Table C30 reflects a comparison of differences from 2012 to 2014 across the Accessible Control and Accessible PAQUED groups. No statistically significant differences emerged for any of the D-in-D results using the $p < 0.017$ threshold, indicating that both groups grew at comparable rates over time.

Table C30. Invented Word Reading Difference-in-Differences Analyses by Region and Group

Region	Group	Baseline 2012		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
Bandundu	Accessible Control	3.39	0.45	1.99	0.40		
	Accessible PAQUED	5.20	0.86	5.55	1.96	1.74	0.19
Equateur	Accessible Control	7.53	1.28	10.65	4.09		
	Accessible PAQUED	6.33	0.99	5.90	1.29	-3.55	-0.13
Orientale	Accessible Control	8.38	1.06	9.21	1.00		
	Accessible PAQUED	9.48	1.49	10.47	2.07	0.16	0.01

Table C31 provides a comparison of mean scores on this subtask for Grade 4 by province, group, and sex.

Table C31. Comparison of Invented Word Reading Mean Scores for Grade 4 by Province, Group, and Sex

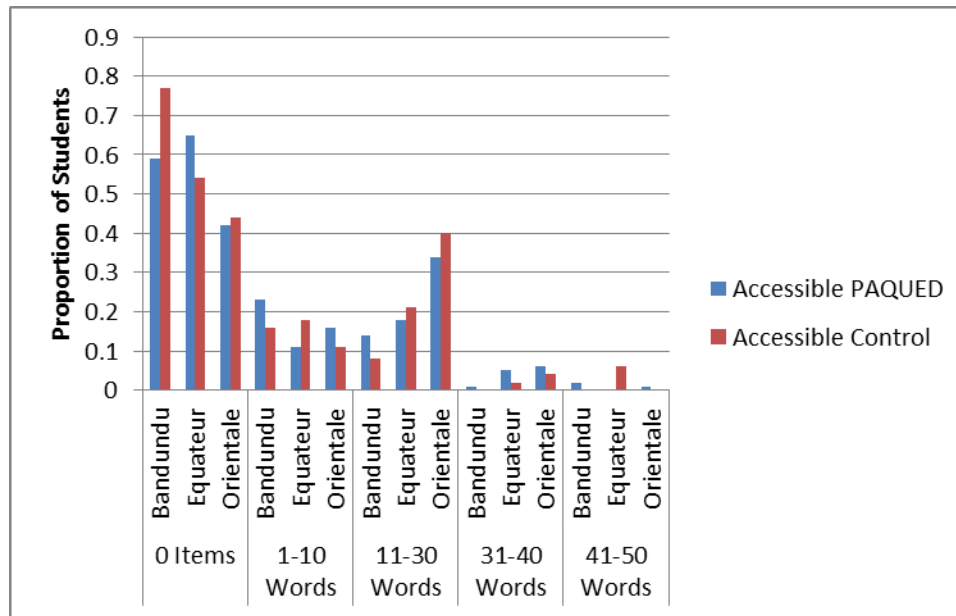
Region	Group	Sex	Baseline 2012		Endline 2014	
			Mean	SE	Mean	SE
Bandundu	Accessible Control	Male	4.76	0.75	3.22	0.78
		Female	2.13	0.46	0.78	0.28
	Accessible PAQUED	Male	5.90	0.97	5.01	1.13
		Female	4.37	1.27	6.04	3.62
Equateur	Accessible Control	Male	8.12	1.41	12.71	6.75
		Female	6.75	1.43	8.16	3.30
	Accessible PAQUED	Male	7.45	1.44	7.45	1.78
		Female	5.36	1.47	4.54	0.98
Orientale	Accessible Control	Male	11.20	1.61	10.97	1.32
		Female	5.75	0.98	7.52	1.21

Region	Group	Sex	Baseline 2012		Endline 2014	
			Mean	SE	Mean	SE
	Accessible PAQUED	Male	10.80	1.47	13.17	3.01
		Female	8.09	2.14	7.42	1.49

No statistically significant differences emerged between sexes across the provinces in either 2012 or 2014 using the $p < 0.004$ threshold.

Exploring distributions of scores across the range of possible scores (0–50 items) again shows the difficulty that students had with this subtask. **Figure C10** illustrates these distributions for Grade 4.

Figure C10. Grade 4 Scores on *Invented Word Reading* at Endline by Province



As illustrated in **Figure C10**, as with reading familiar words, large percentages of students scored zero words per minute on this subtask, although just under half of students were able to read between one and 30 words per minute. Very few students were able to read more than 30 words per minute, again suggesting a lack of automaticity with this skill.

Oral Reading Fluency

The *Oral Reading Fluency* subtask was administered to only students in Grade 4. In this task, students were given a passage containing 54 words and were required to read as much of the passage as they could within one minute. Scores reported for this subtask include percentages of students able to orally read words and the number of words that students could correctly read within one minute.

Table C32 shows percent of zero scores and percent of items attempted for the *Oral Reading Fluency* subtask.

Table C32. Oral Reading Fluency Zero Scores and Percent Attempted for Grade 4 by Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2012	2014	2012	2014	2012	2014
4	Bandundu	Accessible Control	68%	71%	14%	17%	3.86	4.28
		Accessible PAQUED	55%	40%	22%	38%	8.20	12.88
	Equateur	Accessible Control	43%	53%	36%	30%	15.72	12.73
		Accessible PAQUED	57%	52%	28%	33%	10.09	12.68
	Orientale	Accessible Control	41%	31%	38%	52%	13.25	18.42
		Accessible PAQUED	38%	26%	42%	55%	16.89	21.42

As seen in **Table C32**, mean scores at endline on this subtask range from 4.28 in Bandundu to 21.42 in Orientale. Higher scores overall on this subtask than on the *Familiar Word Reading* subtask are not unusual, because reading words in connected text is typically easier than reading words in isolation. That said, even mean scores of 21 words per minute indicate that students, on average, were reading one word approximately every three seconds. This rate of reading speed suggests that attention is being given to individual word reading that should, by Grade 4, be freed up to focus on comprehension of the text being read.

Table C33 reflects a comparison of differences from 2012 to 2014 across the Accessible Control and Accessible PAQUED groups. No statistically significant differences emerged for any of the D-in-D results using the $p < 0.017$ threshold, indicating that both groups grew at comparable rates over time.

Table C33. Oral Reading Fluency Difference-in-Differences Analyses by Region and Group

Region	Group	Baseline 2012		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
Bandundu	Accessible Control	3.86	0.78	4.28	0.89		
	Accessible PAQUED	8.20	1.62	12.88	4.31	4.27	0.28
Equateur	Accessible Control	15.72	2.93	12.73	4.27		
	Accessible PAQUED	10.09	1.78	12.68	2.06	5.57	0.12
Orientale	Accessible Control	13.25	1.80	18.42	1.88		
	Accessible PAQUED	16.89	3.03	21.42	3.75	-0.64	-0.02

Table C34 provides a comparison of mean scores on this subtask for Grade 4 by province, group, and sex.

Table C34. Comparison of Oral Reading Fluency Mean Scores for Grade 4 by Province, Group, and Sex

Region	Group	Sex	Baseline 2012		Endline 2014	
			Mean	SE	Mean	SE
Bandundu	Accessible Control	Male	6.16	1.30	6.92	1.66
		Female	1.78	0.52	1.70	0.55
	Accessible PAQUED	Male	8.41	1.69	12.74	2.44
		Female	7.94	2.50	13.01	6.50
Equateur	Accessible Control	Male	13.42	2.98	12.17	3.66
		Female	18.78	3.68	13.41	5.65
	Accessible PAQUED	Male	10.27	1.93	15.17	2.89
		Female	9.94	2.69	10.46	1.76
Orientale	Accessible Control	Male	17.53 *	2.39	21.51	2.30
		Female	9.28	1.73	15.45	2.26

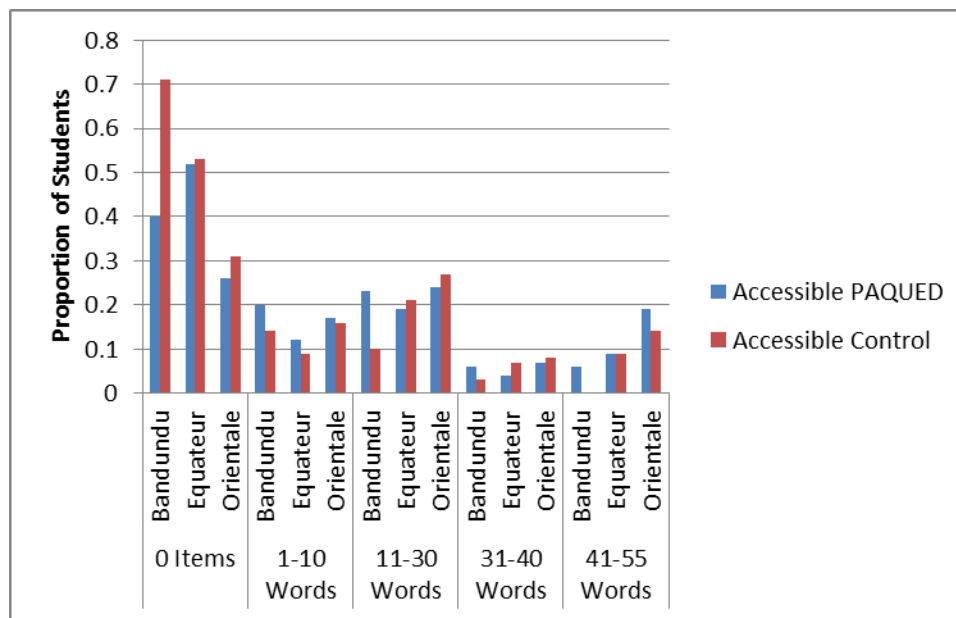
Region	Group	Sex	Baseline 2012		Endline 2014	
			Mean	SE	Mean	SE
	Accessible PAQUED	Male	18.13	2.70	25.97	4.62
		Female	15.57	4.57	16.29	3.78

* significant at $p < 0.004$

One statistically significant difference emerged between sexes using the $p < 0.004$ threshold; boys in Orientale Accessible Control schools outperformed girls at baseline in 2012. (No such difference remained at endline in 2014.)

Exploring distributions of scores across the range of scores (0–55 items) again shows the difficulty that students had with this subtask. *Figure C11* illustrates these distributions for Grade 4.

Figure C11. Grade 4 Scores on Oral Reading Fluency at Endline by Province



As indicated earlier, reading words in connected text is typically easier than reading words in isolation, which is reflected in *Figure C11*. Although more than 40% of students in Bandundu and Orientale, and nearly 30% of students in Equateur, scored zero on this subtask, more than half of students in each province were able to read between one and 55 words. While the overall rate is not high enough to allow for effective comprehension, it does show a positive trend across groups.

Reading Comprehension

After reading the passage in the *Oral Reading Fluency* subtask, Grade 4 students were asked five questions that assessed their basic comprehension of the passage they had read. Scores reported for this subtask include percentages of students able to answer comprehension questions, based upon the number of students who attempted to answer each question.⁵⁰

Table C35 shows the percent of zero scores and percent of items attempted for the *Reading Comprehension* subtask.

Table C35. Reading Comprehension Zero Scores and Percent Attempted for Grade 4 by Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2012	2014	2012	2014	2012	2014
4	Bandundu	Accessible Control	89%	95%	19%	2%	0.13	0.06
		Accessible PAQUED	89%	84%	27%	8%	0.17	0.34
	Equateur	Accessible Control	73%	77%	37%	10%	0.63	0.41
		Accessible PAQUED	83%	84%	41%	7%	0.35	0.30
	Orientale	Accessible Control	78%	76%	43%	10%	0.34	0.39
		Accessible PAQUED	67%	72%	41%	16%	0.59	0.68

As shown in **Table C35**, most students across groups were unable to correctly attempt or respond to any comprehension questions. Across provinces, nearly 80% of students had zero scores—even at endline. Percentages of attempted items and mean scores were correspondingly low, ranging from 2% to 16% at endline. As mentioned earlier, students are only administered comprehension questions that correspond with text they were able to read, and as seen in the Oral Reading Fluency section of this report, students were on average able to read only between four and 21 words. Therefore, the students were administered, on average, two questions at most. Even among the two questions, however, rates of comprehension were relatively low.

Table C36 reflects a comparison of differences from 2012 to 2014 across the Accessible Control and Accessible PAQUED groups. No statistically significant differences emerged

⁵⁰ Students were only asked questions that corresponded with the lines of the passage that they were able to read within one minute.

for any of the D-in-D results using the $p < 0.017$ threshold, indicating that both groups grew at comparable rates over time.

Table C36. Reading Comprehension Difference-in-Differences Analyses by Region and Group

Region	Group	Baseline 2012		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
Bandundu	Accessible Control	0.13	0.04	0.06	0.02		
	Accessible PAQUED	0.17	0.05	0.34	0.18	0.25	0.37
Equateur	Accessible Control	0.63	0.21	0.41	0.19		
	Accessible PAQUED	0.35	0.06	0.30	0.08	0.17	0.08
Orientale	Accessible Control	0.34	0.06	0.39	0.10		
	Accessible PAQUED	0.59	0.14	0.68	0.32	0.03	0.03

Table C37 provides a comparison of mean scores on this subtask for Grade 4 by province, group, and sex.

Table C37. Comparison of Reading Comprehension Mean Scores for Grade 4 by Province, Group, and Sex

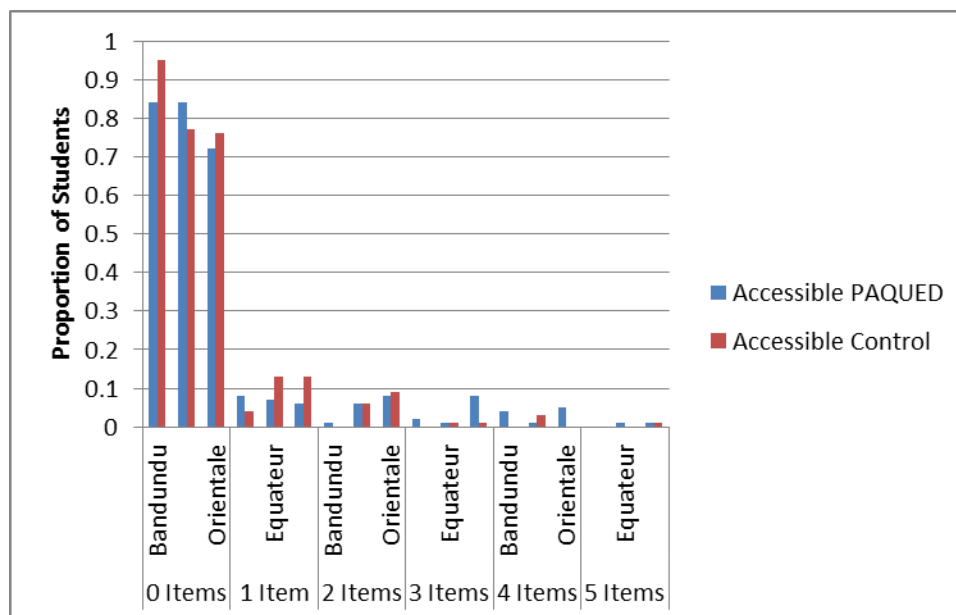
Region	Group	Sex	Baseline 2012		Endline 2014	
			Mean	SE	Mean	SE
Bandundu	Accessible Control	Male	0.17	0.05	0.10	0.04
		Female	0.10	0.04	0.03	0.02
	Accessible PAQUED	Male	0.16	0.06	0.26	0.14
		Female	0.18	0.06	0.42	0.29
Equateur	Accessible Control	Male	0.60	0.25	0.43	0.19
		Female	0.68	0.21	0.40	0.21
	Accessible PAQUED	Male	0.38	0.09	0.30	0.13
		Female	0.33	0.09	0.31	0.08
Orientale		Male	0.58	0.10	0.52	0.11

Region	Group	Sex	Baseline 2012		Endline 2014	
			Mean	SE	Mean	SE
	Accessible Control	Female	0.13	0.06	0.28	0.12
	Accessible PAQUED	Male	0.64	0.14	0.85	0.35
		Female	0.54	0.20	0.49	0.34

No statistically significant difference emerged between sexes in any of the provinces in either 2012 or 2014 using the $p < 0.004$ threshold.

Exploring distributions of scores across the range of scores (0–5 items) again shows the difficulty that students had with this subtask. *Figure C12* illustrates these distributions for Grade 4.

Figure C12. Grade 4 Scores on *Reading Comprehension* at Endline by Province



Because of the low overall scores on the *Oral Reading Fluency* subtask, students were administered a limited number of comprehension questions to answer. Students scoring zero on the *Reading Comprehension* subtask include those who did not read far enough into the passage to receive the first question and those who received questions, but did not answer them correctly. As seen in *Figure C12*, the large majority of students scored zero on this subtask, and the distribution of scores tapers down to nearly 0% of students correctly answering all five questions.

Another informative way to look at the relationship between oral reading fluency and reading comprehension is to explore which reading fluency levels correspond with each level of reading comprehension, as displayed in *Figure C13*.

Figure C13. Grade 4 Overall Correspondence between Oral Reading Fluency and Reading Comprehension

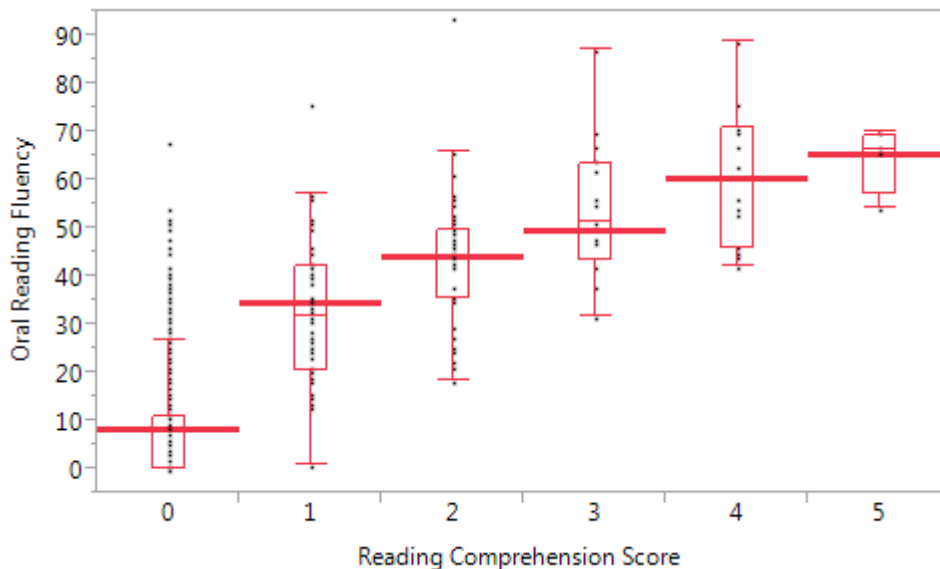


Figure C13 clearly illustrates the relationship between oral reading fluency and reading comprehension. While several outliers exist, students reading between 30 and 40 words per minute tend to accurately respond to one comprehension question, whereas oral reading fluency scores of 60 to 70 words per minute correspond with reading comprehension scores of 4 and 5.

Dictation

For the *Dictation* subtask, examiners read a short sentence to Grade 4 students (“*Mon ami s’appelle Bola.*”), and students attempted to write the sentence. The three target words that were scored were “*mon*”, “*ami*”, and “*appelle*”. Scores reported for this subtask include the number of words spelled correctly.

Table C38. Dictation Zero Scores and Percent Attempted for Grade 4 by Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2012	2014	2012	2014	2012	2014
4	Bandundu	Accessible Control	65%	49%	19%	27%	0.57	0.79
		Accessible PAQUED	53%	40%	27%	32%	0.80	0.94
	Equateur	Accessible Control	39%	47%	37%	31%	1.09	0.92
		Accessible PAQUED	34%	44%	41%	34%	1.23	0.99
	Orientale	Accessible Control	29%	28%	43%	44%	1.30	1.30
		Accessible PAQUED	32%	36%	41%	38%	1.23	1.14

Table C38 shows percent of zero scores and percent of items attempted for the *Dictation* subtask. Interestingly, zero scores on this subtask were not as pervasive as on the *Reading Comprehension* subtask, suggesting that students were generally able to correctly write at least some parts of some words. The highest percentages of zero scores were observed in Bandundu, with 49% of students scoring zero at endline. Not surprisingly, percentages of items attempted were also lowest in this group relative to the other provinces. For the most part, students were able to write approximately one word correctly, with mean scores ranging from 0.79 in Bandundu to 1.30 in Orientale.

Table C39 reflects a comparison of differences from 2012 to 2014 across the Accessible Control and Accessible PAQUED groups. No statistically significant differences emerged for any of the D-in-D results using the $p < 0.017$ threshold, indicating that both groups grew at comparable rates over time.

Table C39. Dictation Difference-in-Differences Analyses by Region and Group

Region	Group	Baseline 2012		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
Bandundu	Accessible Control	0.57	0.07	0.79	0.09		
	Accessible PAQUED	0.80	0.10	0.94	0.20	-0.08	-0.05
Equateur	Accessible Control	1.09	0.10	0.92	0.16		
	Accessible PAQUED	1.23	0.12	0.99	0.13	-0.07	-0.04
Orientale	Accessible Control	1.30	0.13	1.30	0.09		
	Accessible PAQUED	1.23	0.13	1.14	0.24	-0.11	-0.07

Table C40 provides a comparison of mean scores on this subtask for Grade 4 by province and sex.

Table C40. Comparison of Dictation Mean Scores for Grade 4 by Province and Sex

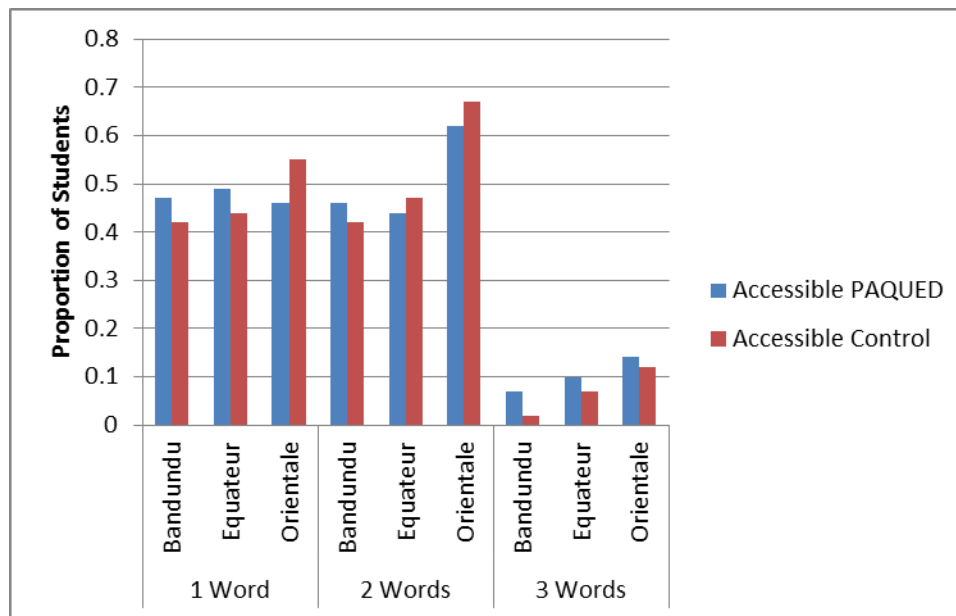
Region	Group	Sex	Baseline 2012		Endline 2014	
			Mean	SE	Mean	SE
Bandundu	Accessible Control	Male	0.79	0.10	0.99	0.13
		Female	0.36	0.07	0.56	0.12
	Accessible PAQUED	Male	0.93	0.10	1.01	0.11
		Female	0.66	0.16	0.89	0.30
Equateur	Accessible Control	Male	1.08	0.11	0.92	0.15
		Female	1.10	0.13	0.91	0.23
	Accessible PAQUED	Male	1.46	0.17	1.11	0.17
		Female	1.01	0.12	0.87	0.12
Orientale	Accessible Control	Male	1.30	0.12	1.39	0.10
		Female	1.29	0.21	1.20	0.15
		Male	1.28	0.14	1.22	0.32

Region	Group	Sex	Baseline 2012		Endline 2014	
			Mean	SE	Mean	SE
	Accessible PAQUED	Female	1.18	0.15	1.03	0.21

One statistically significant difference emerged between sexes using the $p < 0.004$ threshold; boys in Bandundu Accessible Control schools outperformed girls at baseline in 2012. (No such difference remained at endline in 2014.)

Exploring distributions of scores across the range of scores (0–3 items) again shows the difficulty that students had with this subtask. *Figure C14* illustrates these distributions for Grade 4.

Figure C14. Grade 4 Scores on *Dictation* at Endline by Province



Within the *Dictation* subtask, students were scored on fully correct dictation of three words in a longer sentence. As illustrated in *Figure C14*, student scores were relatively evenly distributed across correct encoding of one and two words. Less than a quarter of students overall were able to correctly write all three words.

Student Characteristics Analyses

A series of chi-square tests were run to determine the extent to which relevant student characteristics are correlated with student performance. Because the greatest student literacy gains appear to be on the *Grapheme Recognition* subtask, that skill, as measured at endline, was used as an indicator of student competency. In addition, given that even

this subtask resulted in relatively low levels of performance, only high-performing students (i.e., those scoring in the top quintile on this subtask) were included in the following analyses. *Table C41* shows weighted percentages, chi-square statistics, and p-values for each of the student characteristics identified earlier in this chapter as being of theoretical interest for this purpose.

Table C41. Chi-Squared Analyses of Grade 2 Student Characteristics with Student High Performance on the Grapheme Sound Knowledge Subtask, Accessible PAQUED and Accessible Control Schools

Student Characteristic	Weighted Percentages		
Student has reading book in class			
	Not high-Performing Students	High-performing Students	Total
No	84%	87%	85%
Yes	16%	13%	15%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 1.8317$			
Design-based $F(1.00, 18.00) = 0.1608$ $p = 0.693$			
Someone in the student's home is able to read			
No	15%	9%	13%
Yes	85%	91%	87%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 13.2008$			
Design-based $F(1.00, 18.00) = 1.6525$ $p = 0.215$			
Student has at least one book at home			
No	81%	82%	81%
Yes	19%	19%	19%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 0.0035$			
Design-based $F(1.00, 18.00) = 0.0007$ $p = 0.979$			
Student attended kindergarten			

Student Characteristic	Weighted Percentages		
No	64%	53%	61%
Yes	36%	47%	40%
Total	100%	100%	100%

Pearson: Uncorrected $\chi^2(1) = 18.9831$

Design-based $F(1.00, 18.00) = 4.4247$ $p = 0.050$

If teacher assigns homework, student has someone at home to help with it			
No	52%	59%	55%
Yes	48%	41%	45%
Total	100%	100%	100%

Pearson: Uncorrected $\chi^2(1) = 6.6378$

Design-based $F(1.00, 18.00) = 0.5957$ $p = 0.450$

Student speaks French at home			
No	93%	85%	91%
Yes	7%	15%	9%
Total	100%	100%	100%

Pearson: Uncorrected $\chi^2(1) = 31.9373$

Design-based $F(1.00, 18.00) = 6.1600$ $p = 0.023$

Table C41 shows that no characteristics of Grade 2 students had a significant correlation (at $p < 0.0025$) with high performance on the *Grapheme Sound Knowledge* subtask.⁵¹

Logistic regression analyses were also run on these student characteristics to further explore their relationship with student performance, when province, sex, and student SES were held constant. None of these analyses showed statistically significant relationships.

Table C42 shows weighted percentages, chi-square statistics, and p-values for each of the student characteristics identified earlier in this report as being of theoretical interest for this purpose for Grade 4.

⁵¹ Due to the large number of tests conducted for this section (Chapter C, section 1) of the report, the Bonferroni correction was used to determine the threshold of significance for these analyses. This section contains 20 tests; applying the Bonferroni correction to these analyses results in $p < 0.05 / (20) = 0.0025$.

Table C42. Chi-Squared Analyses of Grade 4 Student Characteristics with Student High Performance on the Grapheme Sound Knowledge Subtask, Accessible PAQUED and Accessible Control Schools

Student Characteristic	Weighted Percentages		
Student has reading book in class			
	Not high-Performing Students	High-performing Students	Total
No	81%	90%	84%
Yes	19%	10%	16%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 22.6288$			
Design-based $F(1.00, 18.00) = 3.6761$ $p = 0.071$			
Someone in the student's home is able to read			
No	10%	8%	9%
Yes	90%	92%	91%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 1.8168$			
Design-based $F(1.00, 18.00) = 0.2176$ $p = 0.646$			
Student has at least one book at home			
No	78%	80%	78%
Yes	22%	20%	22%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 1.7250$			
Design-based $F(1.00, 18.00) = 0.4550$ $p = 0.509$			
Student attended kindergarten			
No	70%	41%	61%
Yes	30%	59%	39%
Total	100%	100%	100%

Student Characteristic	Weighted Percentages		
Pearson: Uncorrected $\chi^2(1) = 137.1445$			
Design-based $F(1.00, 18.00) = 33.9531$ $p = 0.000$			
If teacher assigns homework, student has someone at home to help with it			
No	60%	61%	60%
Yes	40%	39%	40%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 0.4736$			
Design-based $F(1.00, 18.00) = 0.0605$ $p = 0.809$			
Student speaks French at home			
No	88%	80%	86%
Yes	12%	20%	14%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 23.8792$			
Design-based $F(1.00, 18.00) = 1.6559$ $p = 0.214$			

Overall, only one student characteristic was significantly correlated with being a high-performer on this subtask using the $p < 0.0025$ threshold: attendance in kindergarten. Students who reported attending kindergarten were more likely to be high-performers on this subtask.

Logistic regression analyses were also run on these student characteristics to further explore their relationship with student performance, when province, gender, and student SES were held constant. These analyses showed the following (*Table C43*) statistically significant relationship.

Table C43. Grade 4 Student Characteristics Logistic Regressions

Characteristic	p-value
Student attended kindergarten	0.0000*

* significant at $p < 0.0025$

Teacher Characteristics Analyses

Two principal component factor indices were established to explore the relationship between teacher characteristics and student achievement on the *Grapheme Recognition* subtask, which generated the following composite factors:

1. Teacher participation—which includes frequency of reported visits by PAQUED personnel, teacher participation in exchange forums at the cluster level, and teacher participation in exchange forums at the school level; and
2. Teacher access to materials—which includes resources the teacher received from PAQUED, number of radios received from PAQUED, and the number of PAQUED kits the teacher reported using.

Regression analyses on these two composites showed that neither had a substantial impact on student performance.

Logistic regression analyses were also run on these teacher characteristics to further explore their relationship with student performance, when province, gender, and student SES were held constant.⁵² These analyses showed the following (**Table C44**) statistically significant relationships.

Table C44. Teacher Characteristics Logistic Regressions

Characteristic	p-value
Speaking and writing best in Kikongo	0.0006*
Considering student competencies in French as good	0.0022*
Considering student competencies in mathematics as average	0.0022*
Considering student competencies in mathematics as good	0.0012*
Receiving 3–5 visits from PAQUED staff	0.0000*
Participating in cluster-based <i>forums d'échange</i> at a frequency other than 1x/term or 1x/month	0.0070*
Receiving three radios	0.0050*

* significant at $p < 0.0071$

These relationships are not immediately intuitive, and further exploration to determine exactly which characteristics of teachers most impact student progress is warranted.

Head Teacher Characteristics Analyses

It was hypothesized that the sex of the head teacher might impact the extent to which teachers participated in exchange forums and followed the IAI interactive lessons. A

⁵² Due to the large number of tests conducted for this section (Chapter C, section 1) of the report, the Bonferroni correction was used to determine the threshold of significance for these analyses. This section contains 7 tests; applying the Bonferroni correction to these analyses results in $p < 0.05 / (7) = 0.0071$.

significant relationship between the sex of the head teacher and participation in exchange forums at the cluster level ($p < 0.001$) and the school level ($p < 0.05$) did emerge. No other significant relationship was observed.

3. Summary and Conclusions

No consistent, statistically significant trends emerged from the analyses presented in *Chapter C*, which focuses on the Accessible PAQUED schools, either when disaggregating data by region or by sex. However, promising trends can be seen. Overall, student performance at Grade 4 exceeded that of Grade 2, although further investigation is warranted in cases where performance decreased over time, whether due to testing error or other factors in the classroom. Only one significant difference between girls and boys emerged at endline in 2014—the stronger performance of Grade 2 boys on *Grapheme Recognition* for Accessible PAQUED schools in Orientale. While this represents a slight reduction in the number of subtasks where boys outperformed girls at baseline (Grade 4 Accessible Control in Bandundu on *Dictation*, Grade 4 Accessible Control on *Familiar Words* and *Oral Reading Fluency*), it is probably best not to read too much into it given the dozens of tests run which showed no consistent trend.

With this in mind, overall student performance, even at endline, is below what is needed to meet national benchmarks, across all subtasks. Even for oral vocabulary skills, such as vocabulary, phonemic awareness, and listening comprehension, students failed to demonstrate French oral skills required to effectively read in French. Student mean scores on grapheme and word recognition, as well as connected text reading, were also lower than necessary for reading with fluency and comprehension. *Table C45* and *Figure C15* illustrate these results.

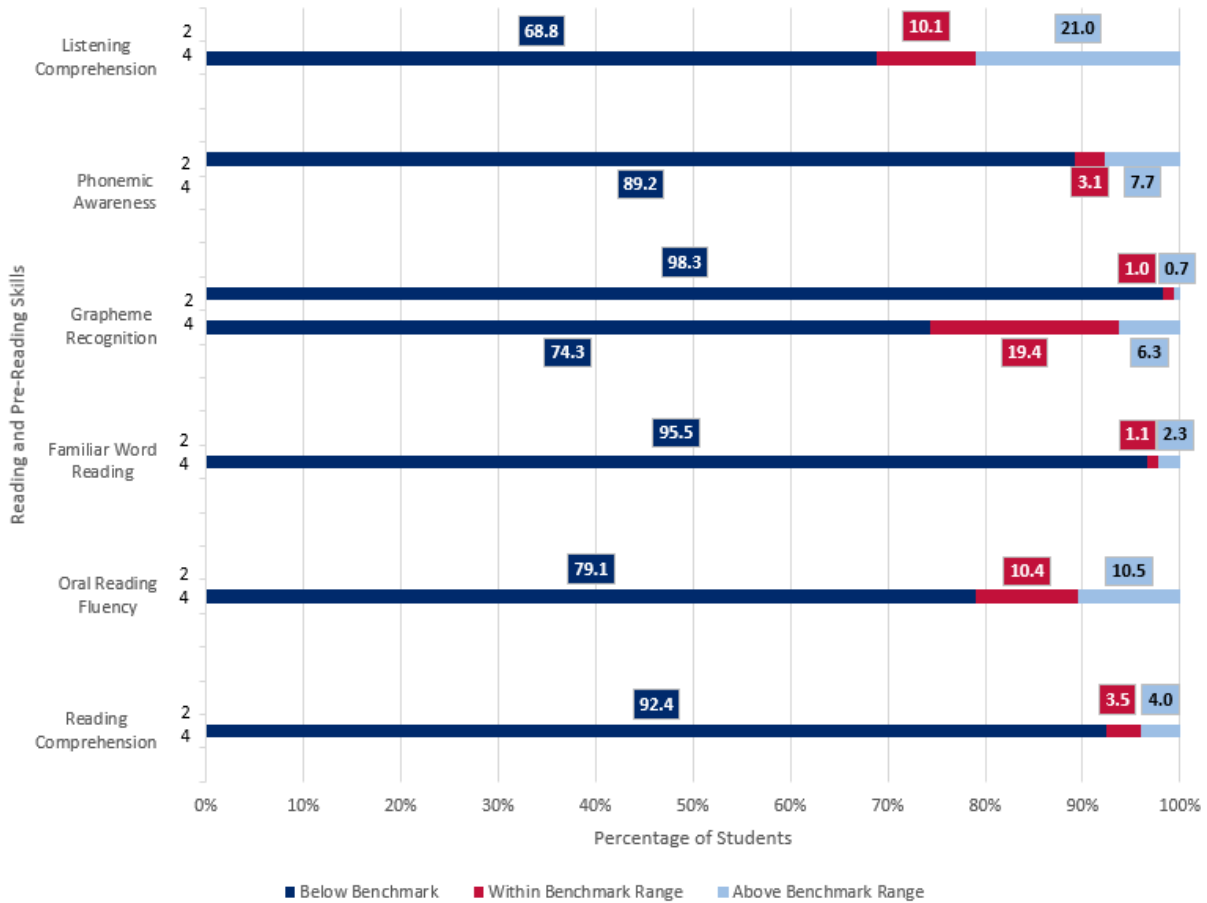
Table C45. Student Performance in Accessible PAQUED Schools Relative to National Benchmarks⁵³

Grade	Reading or Pre-reading Skill	Below benchmark		Within benchmark range		Above benchmark range	
		N	(%)	N	(%)	N	(%)
	Listening Comprehension	-	-	-	-	-	-
2	Phonemic Awareness	897	(89.2%)	36	(3.1%)	61	(7.7%)
	Graphemes	978	(98.3%)	6	(1%)	10	(0.7%)
	Familiar Words	-	-	-	-	-	-
	Oral Reading Fluency	-	-	-	-	-	-

⁵³ The n for each grade presented in this table is as follows: Grade 2 = 994; Grade 4 = 971. Each n presented is unweighted, and each percentage presented is weighted.

Grade	Reading or Pre-reading Skill	Below benchmark		Within benchmark range		Above benchmark range	
		N	(%)	N	(%)	N	(%)
4	Comprehension	-	-	-	-	-	-
	Listening Comprehension	769	(68.8%)	70	(10.1%)	132	(21%)
	Phonemic Awareness	-	-	-	-	-	-
	Graphemes	812	(74.3%)	113	(19.4%)	46	(6.3%)
	Familiar Words	940	(95.5%)	9	(1.1%)	12	(2.3%)
	Oral Reading Fluency	813	(79.1%)	81	(10.4%)	77	(10.5%)
	Comprehension	935	(92.4%)	16	(3.5%)	20	(4%)

Figure C15. Performance of Students in Accessible PAQUED Schools Relative to DRC Benchmarks



These analyses show areas where further exploration and student improvement are required, but they also identify types of supports that appear to be beneficial and that should be continued and strengthened in future education implementations.

Chapter D: Results and Analysis of Student Reading Performance in PAQUED and Control Schools

1. Student Performance on EGRA Measures in PAQUED and Control Schools

Descriptives

The study design enables PAQUED to identify learning gains in PAQUED intervention schools between 2010 and 2014. The analysis presented in the current study includes these comparisons for students in Grades 2, 4, and 6. To draw this sample, RTI randomly selected six subdivisions in each province from among the subdivisions that PAQUED had identified as being eligible using the 2010 criteria. Again, the 2010 criteria for subdivision eligibility required that the subdivision not be in a flood zone, not be more than three days from the center of the province, and must not pose a high security risk. RTI randomly selected six schools in each of the six subdivisions and then identified three to four schools per subdivision with characteristics similar to those of PAQUED schools. RTI selected Control schools at random from this list.

Table D1 displays the intended school sample size by province and grade. A total of 144 schools were intended to be sampled: 108 PAQUED and 36 Control. *Table D2* shows the intended student sample by province and grade. A total of 5,616 students were intended to be sampled and tested (1,872 per grade).

Table D1. Intended School Sample by Province and Grade

Province	Control Schools	PAQUED Schools	Total
Bandundu	12	36	48
Equateur	12	36	48
Orientale	12	36	48
Total	36	108	144

Table D2. Intended Student Sample by Province and Grade ($n = 5,616$)

Province/Grade	Control Students	PAQUED Students	Total
Bandundu			
Grade 2	156	468	624
Grade 4	156	468	624
Grade 6	156	468	624
Equateur			
Grade 2	156	468	624
Grade 4	156	468	624
Grade 6	156	468	624
Orientale			
Grade 2	156	468	624
Grade 4	156	468	624
Grade 6	156	468	624
Total			
Grade 2	468	1404	1872
Grade 4	468	1404	1872
Grade 6	468	1404	1872

Due to challenges faced during testing—such as student absenteeism on the scheduled day—361 fewer students (from the intended 5,616) were tested overall. **Table D3** displays the actual school sample, and **Table D4** displays the actual student sample used in the subsequent analyses in this report by province and grade.

Table D3. Actual School Sample by Province and Grade

Province	Control Schools	PAQUED Schools	Total
Bandundu	12	36	48
Equateur	13	36	49
Orientale	9	39	48
Total	34	111	145

Table D4. Actual Student Sample by Province and Grade ($n = 5,255$)

	Grade 2		Grade 4		Grade 6		Total Students
	Control	PAQUED	Control	PAQUED	Control	PAQUED	
Bandundu	139	413	118	390	121	398	1,579
Equateur	141	479	134	475	130	472	1,831
Orientale	118	505	120	508	117	477	1,845

Table D5 displays the proportion of schools in each group by school management type. As this table indicates, Catholic-managed schools were the most prominent type within this sample (overall $n = 50$), followed by unregistered schools ($n = 37$).

Table D5. School Type, by Group ($n = 145$)

Province	Government-managed	Catholic	Kimbanguist	Islamic	Protestant	Missing	Total Schools
Bandundu	13	13	2	-	15	5	48
Equateur	15	23	4	2	5	-	39
Orientale	14	20	1		12	1	48
Total	42	56	7	2	32	6	145

2014 Sample

As indicated, the 2014 sample included 5,255 students sampled from 124 schools in three provinces. Equal percentages (35%) of students came from Equateur and Orientale provinces, while 30% of students came from Bandundu. Students were relatively equally distributed across grades and sex. *Table D6* describes the general characteristics of the student sample.

Table D6. General Characteristics of the Overall Student Sample (n = 5,255)

Variable	Number of Students	Percent
Province		
Bandundu	1,579	30%
Equateur	1,831	35%
Orientale	1,845	35%
Grade		
2	1,795	34%
4	1,745	33%
6	1,715	33%
Sex		
Female	2,407	46%
Male	2,848	54%

Tables D7 and *D8* show how students responded to a series of questions targeting SES indicators regarding possessions in the home. While students were asked more questions than those listed here, the following questions were determined to be of greatest theoretical interest and the most likely to impact student performance. Because student SES has frequently been shown to impact student performance, regression analyses reported later in this report include an SES composite in their models.⁵⁴

⁵⁴ Having several highly correlated independent variables (such as the SES-related questions shown in *Table D7*) can produce unstable estimates; reducing those to a smaller number of principal components will yield a better test and/or more stable estimates of regression coefficients. Therefore, for the purposes of regression analyses incorporating SES, the students' answers to the questions regarding possessions in the home were collapsed into a single variable using PCA.

Table D7. Student SES Indicators (n = 5,255)

SES Item	Number of Students	Percentage
Radio	3,695	71%
Telephone	2,513	48%
Electricity in the home	976	19%
Television	1,027	20%
Refrigerator	317	6%
Indoor toilets	420	8%
Bicycle	3,610	69%
Motorcycle	1,312	25%
Canoe	710	14%
Motor vehicle	145	3%

Table D8. Select Student Responses to Demographic Questions (n = 5,204)

Province	Number of Students	Total in Province	Percent of Province
Student has reading book in class			
Bandundu	316	1,567	21%
Equateur	269	1,808	15%
Orientale	287	1,829	16%
Someone in the student's home is able to read			
Bandundu	1,319	1,567	84%
Equateur	1,324	1,808	73%
Orientale	1,616	1,829	88%
Student has at least one book at home			
Bandundu	446	1,567	29%
Equateur	275	1,808	15%
Orientale	372	1,829	20%

Province	Number of Students	Total in Province	Percent of Province
Student attended kindergarten			
Bandundu	490	1,567	31%
Equateur	307	1,808	17%
Orientale	232	1,829	13%
If teacher assigns homework, student has someone at home to help with it			
Bandundu	389	1,567	25%
Equateur	342	1,808	19%
Orientale	618	1,829	34%

As indicated in *Table D8*, students across the three provinces were generally comparable on whether they had a reading book in class, although the percentage of students in Bandundu exceeded that of the other provinces. Students in Equateur appeared to be somewhat less likely than students in other provinces to report having a book at home or someone at home able to read. Students in Bandundu were more likely to report having attended kindergarten (31% versus 17% in Equateur and 13% Orientale), although it was Orientale that had the highest percentage of students reporting having someone at home helping with homework.

Table D9 reflects the results of student self-reports on what language is used in their home. Students had the option of indicating more than one language. Therefore, student responses exceed the total number of students in the sample.

Table D9. Student Indication of Language(s) Spoken in the Home

Province	French Number / %	Kikongo Number / %	Lingala Number / %	Kiswahili Number / %	Other Number / %	Total Student Reports
Bandundu	89 / 5%	740 / 38%	619 / 32%	5 / 0%	472 / 25%	1,925
Equateur	118 / 5%	4 / 0%	1,137 / 51%	5 / 0%	961 / 43%	2,225
Orientale	101 / 4%	1 / 0%	1,124 / 40%	756 / 27%	817 / 29%	2,799
Total	308 / 4%	745 / 11%	2,880 / 41%	766 / 11%	2,250 / 32%	6,949

Perhaps not surprisingly, given differences in predominant languages across provinces, more students in Bandundu reported speaking Kikongo at home than any other language, although more students in Equateur and Orientale reported speaking Lingala at home.

Use of French in the home ranged from a total number of 89 student reports (Bandundu) to 118 (Equateur).

Teachers were also asked a series of questions at the time of the student testing. *Tables D10* and *D11* describe the general characteristics and relevant survey responses (responses that will be included in subsequent regression analyses) of the teacher sample.

Table D10. General Characteristics of the Teacher Sample (n = 389)

Variable	Number of Teachers	Percent
Province (n = 389)		
Bandundu	113	29%
Equateur	142	37%
Orientale	134	34%
Grade (n = 404)*		
1	5	1%
2	127	31%
3	9	2%
4	133	33%
5	5	1%
6	125	31%
Sex (n = 389)		
Female	96	25%
Male	293	75%

*Some teachers reported teaching multiple grades. Therefore, the total number of responses exceeds the total number of teachers in this sample.

As seen in *Table D10*, the teacher sample includes substantially more men (75%) than women (25%).

Table D11. Select Teacher Responses to Survey Questions (n = 389)

Province	Number / %	Number / %	Number / %	Number / %	Number / %	Total in Province
What language do you speak and write in the best?						
	Other/No Response	French	Kikongo	Lingala	Kiswahili	
Bandundu	4 / 4%	64 / 57%	33 / 29%	12 / 11%	0 / 0%	113
Equateur	4 / 3%	88 / 62%	0 / 0%	50 / 35%	0 / 0%	142
Orientale	8 / 6%	58 / 43%	0 / 0%	35 / 26%	33 / 25%	134
What is your highest level of education?⁵⁵						
	D4	D6	PP5	CSP or CAP	G3	
Bandundu	12 / 11%	95 / 86%	0 / 0%	3 / 3%	0 / 0%	110
Equateur	29 / 21%	107 / 76%	1 / 1%	0 / 0%	1 / 0%	138
Orientale	27 / 21%	92 / 72%	6 / 5%	3 / 2%	0 / 0%	128
Other than PAQUED training, over the past two years how often did you receive in-service training in how to teach the French language?						
	Never	1 Time	2+ Times	No Response		
Bandundu	4 / 4%	18 / 16%	33 / 29%	58 / 51%		113
Equateur	15 / 11%	19 / 13%	49 / 35%	59 / 42%		142
Orientale	10 / 8%	24 / 18%	46 / 34%	54 / 40%		134
How do you characterize your students' competence in French?						
	Weak	Average	Strong	No Response		
Bandundu	24 / 21%	63 / 56%	23 / 21%	3 / 3%		113
Equateur	22 / 16%	85 / 60%	31 / 22%	4 / 3%		142
Orientale	24 / 18%	83 / 62%	24 / 18%	3 / 2%		134

⁵⁵ D4 = 4 years of post-primary education

PP5 = 5 years of post-primary education; a sort of specialized vocational degree, but not completion of secondary education

D6 = 6 years of post-primary education; completion of secondary education

CSP or CAP = *Cycle spécialisation professionnelle*

G3 = 3 years of post-secondary education; completion of the first half of a course of study in an *institut supérieure*

Province	Number / %	Number / %	Number / %	Number / %	Number / %	Total in Province
How do you characterize your students' competence in mathematics?	Weak	Average	Strong	No Response		
Bandundu	7 / 6%	52 / 46%	51 / 45%	3 / 3%		113
Equateur	13 / 9%	85 / 60%	40 / 28%	4 / 3%		142
Orientale	9 / 7%	68 / 51%	54 / 40%	3 / 2%		134
How often did you receive a visit by PAQUED personnel this past school year?	1	2-3	4-5	6+	No Response	
Bandundu	10 / 9%	15 / 13%	1 / 1%	0 / 0%	87 / 77%	113
Equateur	14 / 10%	3 / 2%	1 / 1%	0 / 0%	124 / 87%	142
Orientale	23 / 17%	15 / 11%	2 / 2%	2 / 2%	92 / 67%	134
How often did you participate in a teacher exchange forum at the cluster level?	At Least One Time per Trimester	At Least One Time per Month	Other	No Response		
Bandundu	26 / 23%	17 / 15%	2 / 2%	68 / 60%		113
Equateur	10 / 7%	13 / 9%	3 / 2%	116 / 82%		142
Orientale	26 / 19%	9 / 7%	6 / 5%	93 / 69%		134
How often did you participate in a teacher exchange forum at the school level?	At Least One Time per Trimester	At Least One Time per Month	Other	No Response		
Bandundu	9 / 8%	18 / 16%	5 / 4%	81 / 72%		113
Equateur	6 / 4%	27 / 19%	5 / 4%	104 / 73%		142
Orientale	30 / 22%	23 / 17%	4 / 3%	77 / 58%		134
If you participated in teacher exchange forums, what video modules were used?*	IAI Lessons	Teaching Materials	No Modules Used			
Bandundu	14 / 22%	3 / 5%	47 / 73%			64
Equateur	26 / 43%	25 / 42%	9 / 15%			60
Orientale	49 / 54%	34 / 37%	8 / 9%			91

Province	Number / %	Number / %	Number / %	Number / %	Number / %	Total in Province
What resources did you receive from the PAQUED project?*	IAI Guide	Reading Activities Guide	Read Aloud Books	Student Texts	Chalk	
Bandundu	40 / 19%	52 / 25%	38 / 18%	33 / 16%	47 / 22%	210
Equateur	79 / 41%	25 / 13%	20 / 10%	17 / 9%	53 / 27%	194
Orientale	77 / 25%	68 / 22%	46 / 15%	37 / 12%	76 / 25%	304
How many radios did your school receive from the PAQUED project?	1	2	3	Other	No Response	
Bandundu	65 / 58%	3 / 3%	0 / 0%	0 / 0%	45 / 40%	113
Equateur	62 / 44%	5 / 4%	0 / 0%	0 / 0%	75 / 53%	142
Orientale	77 / 58%	15 / 11%	4 / 3%	0 / 0%	38 / 28%	134
If you used a PAQUED kit, which one did you use?*	Materials Fabrication	School Kit	Class Kit	No Response		
Bandundu	49 / 38%	45 / 35%	35 / 27%	1 / 1%		130
Equateur	33 / 30%	39 / 35%	38 / 34%	2 / 2%		112
Orientale	53 / 34%	57 / 36%	47 / 30%	0 / 0%		157
Did you follow the interactive IAI lessons at your school?	Yes	No	No Response			
Bandundu	60 / 53%	21 / 19%	32 / 28%			113
Equateur	64 / 45%	47 / 33%	31 / 22%			142
Orientale	74 / 55%	30 / 22%	30 / 22%			134

*Because multiple responses were allowed per teacher, percentages do not necessarily sum to 100.

As seen in *Table D11*, most teachers in all provinces reported speaking and writing best in French; relatively few teachers stated that Kikongo, Lingala, or Kiswahili were their strongest languages, although patterns in language preference emerged by province. Most teachers reported D6 as their highest level of education attained, with fewer teachers stating D4 and very few teachers indicating PP5, CSP/CAP, or G3. Of the teachers who provided responses on these questions, most reported having attended at least two in-service trainings (other than PAQUED trainings) and having received at least one visit from PAQUED personnel. Interestingly, teachers were fairly consistent in categorizing their students' proficiencies in French and in mathematics, with most teachers rating students as average in both subjects. Promisingly, among teachers who responded to this

question, most reported having participated in cluster- and school-level exchange forums, although the frequency varied by province.

The use of video modules in exchange forums was a bit mixed; 73% of teachers in Bandundu indicated that no modules were used, although this percentage was much lower for teachers in Equateur (15%) and Orientale (9%). Teachers across the provinces reported receiving a range of PAQUED materials, with a higher percentage of teachers in Equateur reporting receipt of IAI guides than teachers in the other provinces. Use of the different types of classroom kits was fairly evenly distributed. Finally, most teachers reported using the IAI interactive lessons.

Tables D12 and **D13** describe the general characteristics and relevant survey responses (responses that will be included in subsequent regression analyses) of the head teacher sample.

Table D12. General Characteristics of the Head Teacher Sample (n = 142)

Variable	Number of Head Teachers	Percent
Province		
Bandundu	46	32%
Equateur	49	35%
Orientale	47	33%
Sex		
Female	12	9%
Male	130	92%

As shown in **Table D12**, head teachers were predominantly male (92% compared to 9% female). This is consistent across provinces, as demonstrated in **Table D13** below.

Table D13. Select Head Teacher Responses to Survey Questions (n = 142)

Province	Number / %	Number / %	Number / %	Number / %	Total in Province
What is the sex of the head teacher?	Female	Male			
Bandundu	2 / 4%	44 / 96%			46
Equateur	3 / 6%	46 / 94%			49
Orientale	7 / 15%	40 / 85%			47

Province	Number / %	Number / %	Number / %	Number / %	Number / %	Total in Province
How often did you receive a visit by PAQUED personnel this past school year?	1	2–3	4–5	No Response		
Bandundu	5 / 11%	10 / 22%	5 / 11%	26 / 57%		46
Equateur	2 / 4%	5 / 10%	1 / 2%	41 / 84%		49
Orientale	9 / 19%	15 / 32%	3 / 6%	20 / 43%		47

Table D13 also shows that among head teachers who responded to this question, the majority in Bandundu and Orientale reported receiving two or more visits by PAQUED personnel in the prior year, although in Equateur only six head teachers report two or more visits.

2. EGRA Subtest Outcomes

Table D14 shows overall zero scores and means scores—both including and excluding students with zero scores—at endline in 2014 for each subtask. A student receives a score of zero on a subtask if that student (1) is unable to attempt even one item on the task, or (2) attempts items but does not get any correct. On subtasks where relatively few students scored zero, the difference between means that include these zero scores and those that exclude them is not large. However, on subtasks where a large proportion of students had zero scores, the difference can be substantial, and it is often useful to consider both means when attempting to determine student performance.

Table D14. Overall Percent Zero Scores and Mean Scores by Grade and Subtask at Endline

Subtask	Grade	Group	% Zero Scores	Means Excluding Zero Scores	Means Including Zero Scores
Vocabulary	2	Control	4%	6.98	7.24
		PAQUED	3%	7.57	7.81
	4	Control	0%	9.83	9.85
		PAQUED	0%	9.94	9.97
	6	Control	0%	12.73	12.76
		PAQUED	0%	12.66	12.66
Initial Sound Identification	2	Control	80%	0.96	4.78
		PAQUED	83%	0.67	3.87

Subtask	Grade	Group	% Zero Scores	Means Excluding Zero Scores	Means Including Zero Scores	
	4	Control	66%	1.93	5.72	
		PAQUED	68%	1.67	5.25	
	6	Control	47%	3.67	6.87	
		PAQUED	55%	2.70	5.95	
Listening Comprehension	2	Control	72%	0.55	1.97	
		PAQUED	73%	0.53	1.92	
	4	Control	58%	0.84	1.98	
		PAQUED	53%	0.92	1.95	
	6	Control	29%	1.71	2.40	
		PAQUED	30%	1.72	2.45	
	Grapheme Sound Knowledge	2	Control	52%	5.00	10.33
			PAQUED	53%	4.13	8.74
4		Control	14%	18.44	21.49	
		PAQUED	16%	16.36	19.39	
6		Control	1%	38.32	38.64	
		PAQUED	4%	33.83	35.37	
Familiar Word Reading	4	Control	65%	4.59	13.13	
		PAQUED	70%	3.88	12.91	
	6	Control	22%	20.14	25.82	
		PAQUED	28%	16.57	22.96	
Invented Word Reading	4	Control	71%	3.86	13.16	
		PAQUED	71%	3.38	11.62	
	6	Control	24%	15.98	21.13	
		PAQUED	33%	12.63	18.77	
Oral Reading Fluency	4	Control	54%	9.36	20.45	

Subtask	Grade	Group	% Zero Scores	Means Excluding Zero Scores	Means Including Zero Scores
Reading Comprehension	6	PAQUED	55%	8.23	18.20
		Control	14%	34.44	40.02
	4	PAQUED	21%	28.66	36.40
		Control	89%	0.24	2.07
	6	PAQUED	92%	0.14	1.80
		Control	58%	0.88	2.13
Dictation	4	PAQUED	61%	0.78	1.98
		Control	49%	0.84	1.65
	6	PAQUED	50%	0.79	1.60
		Control	19%	1.65	2.02
		PAQUED	24%	1.47	1.93

On subtasks with small percentages of zero scores (such as Vocabulary and, to a lesser degree, Grapheme Sound Knowledge in Grade 6) differences between means are not great. Differences between means that include and exclude students with zero scores are notable on the remaining subtasks. As anticipated, overall student mean scores were higher in Grade 6 than in Grade 4, and higher in Grade 4 than in Grade 2, for subtasks that were administered in all grades. The relative amount of increase, however, suggests an ongoing deficiency in skills even at the higher grade. The presence of substantial proportions of zero scores, even in Grade 6, further indicates student performance that is lower than required to achieve Grade 6 benchmarks.

The following subsections of this chapter provide additional detail about each of the EGRA subtasks represented in this chapter.

Vocabulary

The Vocabulary subtask presented students with 20 vocabulary words. Data collectors asked the students to identify several body parts and objects as well as to move objects in a variety of directions. As such, this subtask is an assessment of basic French vocabulary, focused on the types of words and concepts found in the environment of students. The type of vocabulary assessed is in the DRC curriculum for French in Grade 1. The curriculum specifies that teaching should include common words and that students should

be able to perform a gesture or action based on instructions given by the teacher in French.

Table D15 shows the percent of zero scores and percent of items attempted for the *Vocabulary* subtask.

Table D15. Vocabulary Zero Scores and Percent Attempted, by Grade and Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2010	2014	2010	2014	2010	2014
2	Bandundu	Control	3%	12%	32%	28%	6.45	5.69
		PAQUED	1%	8%	31%	31%	6.15	6.12
	Equateur	Control	1%	1%	28%	36%	5.53	7.22
		PAQUED	2%	4%	31%	42%	6.21	8.40
	Orientale	Control	1%	0%	36%	39%	7.25	7.72
		PAQUED	1%	0%	35%	39%	7.08	7.72
4	Bandundu	Control	1%	0%	43%	43%	8.59	8.58
		PAQUED	0%	1%	43%	43%	8.47	8.53
	Equateur	Control	0%	1%	42%	48%	8.43	9.67
		PAQUED	0%	1%	44%	50%	8.78	9.94
	Orientale	Control	0%	0%	55%	56%	10.91	11.20
		PAQUED	0%	0%	52%	53%	10.34	10.59
6	Bandundu	Control	0%	0%	59%	61%	11.73	12.30
		PAQUED	0%	0%	56%	59%	11.22	11.78
	Equateur	Control	0%	1%	55%	63%	10.92	12.53
		PAQUED	0%	0%	56%	60%	11.12	11.97
	Orientale	Control	0%	0%	69%	67%	13.89	13.39
		PAQUED	0%	0%	74%	69%	14.75	13.78

As shown in **Table D15**, even in 2010, very few students (the largest proportion being 3% for Grade 2 in Bandundu) scored zero on this subtask, suggesting that students had at least a minimal level of oral competence in French. That said, at endline, students in Grade 2 were able to attempt relatively few items (ranging from 28% to 42%) and on

average were only able to correctly respond to less than ten words. Performance in Grade 4 was better, but even by endline Grade 6, on average, students were not demonstrating proficiency on this subtask; the highest mean score (13.78) came from Orientale (PAQUED group). By Grade 6, students are expected to have acquired the basic level of oral language competence in French that would allow them to respond to simple vocabulary.

Table D16 reflects a comparison of differences from 2010 to 2014 across the Control and PAQUED groups. No statistically differences emerged for any of the D-in-D results using the $p < 0.006$ threshold, indicating that both groups grew at comparable rates over time.⁵⁶

Table D16. Vocabulary Difference-in-Differences Analyses by Grade, Province, and Group

Grade 2 Province	Group	Baseline 2010		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
Bandundu	Control	6.45	1.05	5.69	1.12		
	PAQUED	6.15	0.58	6.12	0.69	0.72	0.15
Equateur	Control	5.53	0.21	7.22	0.83		
	PAQUED	6.21	0.20	8.40	0.76	0.50	0.10
Orientale	Control	7.25	0.52	7.72	0.68		
	PAQUED	7.08	0.62	7.72	0.56	0.17	0.04
Grade 4 Province	Group	Baseline 2010 Mean	SE	Endline 2014 Mean	SE	D-in-D	ES
Bandundu	Control	8.59	0.91	8.58	0.72		
	PAQUED	8.47	0.41	8.53	0.66	0.08	0.01
Equateur	Control	8.43	0.39	9.67	0.77		
	PAQUED	8.78	0.31	9.94	0.67	-0.07	-0.02
Orientale	Control	10.91	0.76	11.20	0.87		

⁵⁶ Type 1 errors in statistics occur when a difference is thought to exist where one does not. (Put another way, a type 1 error is the rejection of the null hypothesis when it is actually true.) Due to the large number of difference-in-differences comparisons conducted for this section (Chapter C, section 2) of the report, the Bonferroni correction was used to determine the threshold of significance for the various analyses. This section contains 9 tests of D-in-D for the subtasks administered to all three grades and 6 tests of D-in-D for the subtasks administered to Grades 4 and 6 alone. Applying the Bonferroni correction to these analyses results in $p < 0.05 / (9) = 0.006$ for Vocabulary, Initial Sound Identification, Listening Comprehension, and Grapheme Recognition subtasks and $p < (0.05) / (9) = 0.008$ for the Familiar Word, Invented Word, Oral Reading Fluency, Reading Comprehension, and Dictation subtasks. Please see Annex 1 for a fuller discussion of the Bonferroni correction and the analytical families used in this report.

Grade 4 Province	Group	Baseline 2010		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
	PAQUED	10.34	0.85	10.59	0.51	-0.03	-0.01
Grade 6 Province	Group	Baseline 2010		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
Bandundu	Control	11.73	0.42	12.30	0.76		
	PAQUED	11.22	0.22	11.78	0.72	0.00	0.00
Equateur	Control	10.92	0.45	12.53	0.80		
	PAQUED	11.12	0.35	11.97	0.60	-0.78	-0.18
Orientale	Control	13.89	0.62	13.39	0.75		
	PAQUED	14.75	0.70	13.78	0.35	-0.47	-0.12

Given that, overall, students have not mastered this level of French vocabulary, it is interesting to explore levels of competence between girls and boys. *Table D17* provides a comparison of mean scores on this subtask by grade, province, group, and sex.⁵⁷

Table D17. Comparison of Vocabulary Mean Scores by Grade, Province, Group, and Sex

Grade	Province	Group	Sex	Baseline 2010		Endline 2014	
				Mean	SE	Mean	SE
2	Bandundu	Control	Male	6.74	1.17	5.46	0.94
			Female	5.99	0.85	5.90	1.58
		PAQUED	Male	5.96	0.31	6.14	0.72
			Female	6.29	0.85	6.09	0.71
	Equateur	Control	Male	5.82	0.37	7.29	1.00
			Female	5.20	0.17	7.16	0.76
		PAQUED	Male	6.36	0.24	8.75	0.72
			Female	6.02	0.19	8.00	0.85

⁵⁷ This section contains 36 tests of comparison of means for the subtasks administered to all three grades and 24 tests of comparison of means for the subtasks administered to Grades 4 and 6 alone. Applying the Bonferroni correction to these analyses results in $p < 0.05 / (36) = 0.001$ for Vocabulary, Initial Sound Identification, Listening Comprehension, and Grapheme Recognition subtasks and $p < (0.05) / (24) = 0.002$ for the Familiar Word, Invented Word, Oral Reading Fluency, Reading Comprehension, and Dictation subtasks. Please see Annex 1 for a fuller discussion of the Bonferroni correction and the analytical families used in this report.

Grade	Province	Group	Sex	Baseline 2010		Endline 2014		
				Mean	SE	Mean	SE	
4	Orientale	Control	Male	7.36	0.54	8.53	0.65	
			Female	7.15	0.63	6.86	0.74	
		PAQUED	Male	7.00	0.68	8.26	0.65	
			Female	7.12	0.58	7.12	0.46	
		Bandundu	Control	Male	9.19	1.11	8.30	0.65
				Female	7.97	0.75	8.83	1.10
			PAQUED	Male	8.63	0.36	8.79	0.63
				Female	8.28	0.51	8.20	0.78
	Equateur	Control	Male	9.30 *	0.40	9.40	0.91	
			Female	7.22	0.30	9.97	0.65	
		PAQUED	Male	8.69	0.23	10.34	0.79	
			Female	8.85	0.44	9.53	0.59	
	6	Orientale	Control	Male	11.05	1.07	11.47	0.85
				Female	10.75	0.69	10.81	0.99
			PAQUED	Male	11.09	0.98	10.85	0.57
				Female	9.22	0.70	10.30	0.49
Bandundu		Control	Male	12.27 *	0.45	12.27	0.80	
			Female	11.02	0.37	12.33	1.19	
		PAQUED	Male	11.46	0.39	11.93	0.46	
			Female	11.04	0.19	11.63	1.10	
Equateur	Control	Male	11.08	0.63	12.29	0.75		
		Female	10.51	0.59	12.80	0.85		
	PAQUED	Male	11.20	0.20	12.47	0.56		
		Female	11.01	0.69	11.28	0.68		

Grade	Province	Group	Sex	Baseline 2010		Endline 2014	
				Mean	SE	Mean	SE
Oriental		Control	Male	14.32	0.73	13.84	0.66
			Female	13.56	0.59	12.77	1.02
		PAQUED	Male	14.93	0.74	14.11	0.39
			Female	14.40	0.66	13.34	0.36

*Difference is statistically significant at $p < 0.001$.

In examining the levels of competence between girls and boys, only two significant differences emerged using the $p < 0.001$ threshold: (1) Grade 4 boys in Equateur Control schools outperformed girls at baseline in 2010, and (2) Grade 6 boys in Bandundu Control schools outperformed girls at endline in 2014. (In neither case did the difference remain at endline in 2014.)

Exploring distributions of scores across the range of possible scores (0–20 items) again shows that relatively few students scored zero, but also relatively few achieved perfect scores on this subtask. *Figures D1* through *D3* illustrate these distributions by grade.

Figure D1. Grade 2 Scores on Vocabulary at Endline by Province

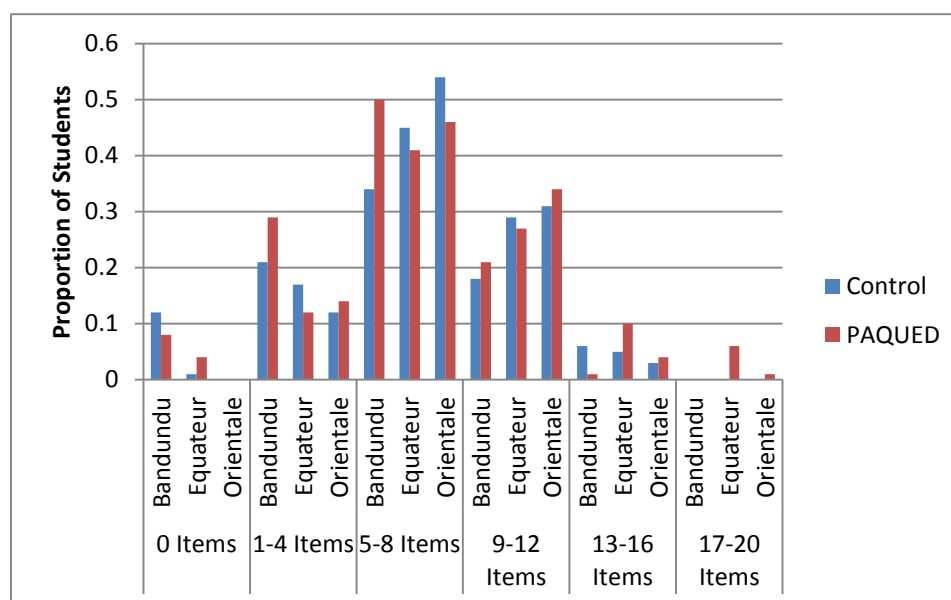


Figure D2. Grade 4 Scores on Vocabulary at Endline by Province

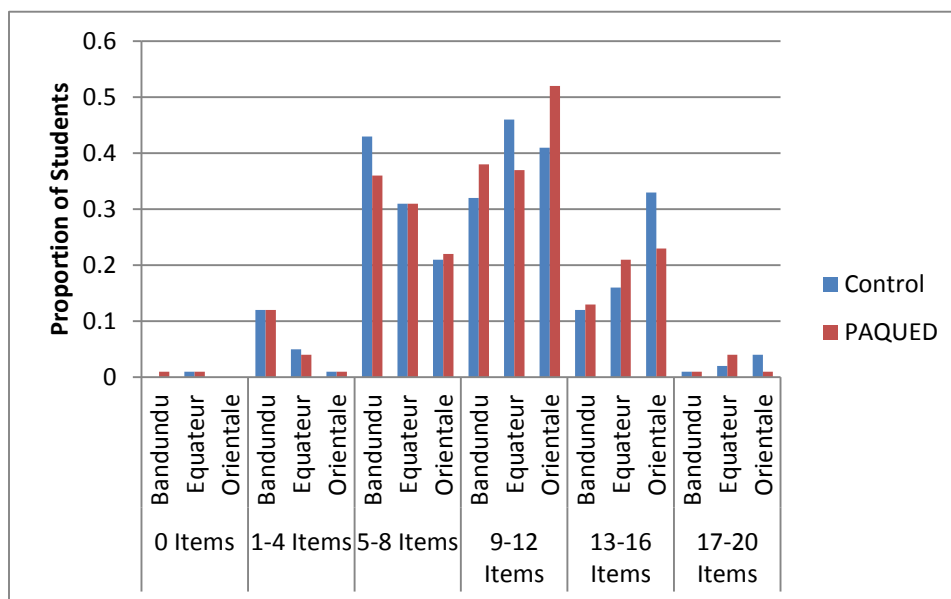
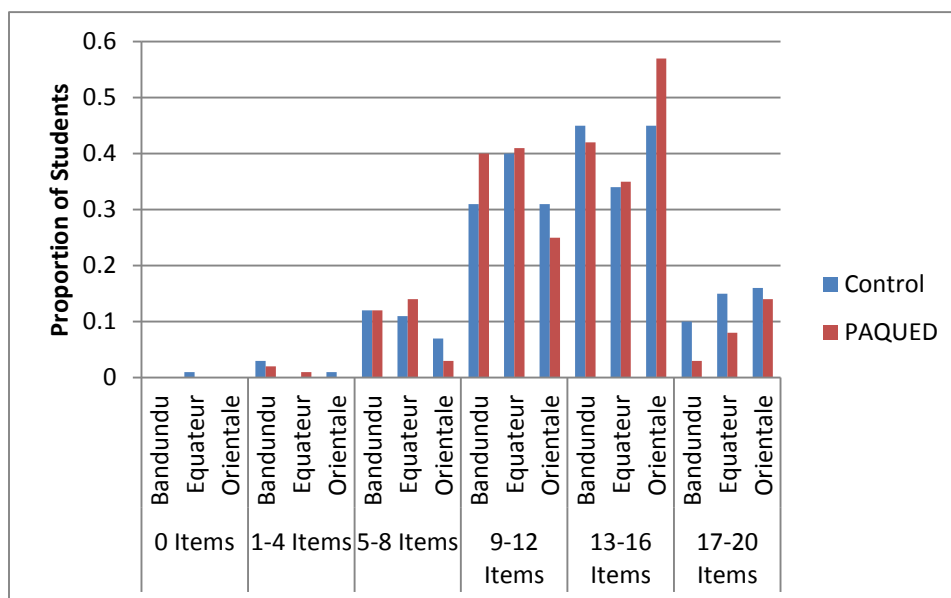


Figure D3. Grade 6 Scores on Vocabulary at Endline by Province



What is encouraging is that while the distribution for Grade 2 shows most students scoring in the five to 16 item range, this distribution is shifted up for Grade 4 and again for Grade 6, with most students in Grade 6 scoring between nine and 16 items and a growing number achieving perfect scores.

Initial Sound Identification

In the Initial Sound Identification subtask, students listened to individual words, such as “*sac*,” and were asked to identify the first sound, or phoneme, of that word (in this case, /s/). This subtask was comprised of 10 items, for a maximum possible score of 10. The first five items, *sac*, *jour*, *date*, *lac*, and *car* were presented to all students. Students who were able to identify the first sound of at least one of these words were presented the remaining five items (*balle*, *tour*, *par*, *vol*, and *fil*).

Table D18 shows the percent of zero scores and percent of items attempted for the Initial Sound Identification subtask.

Table D18. Initial Sound Identification Zero Scores and Percent Attempted, by Grade and Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score		
			2010	2014	2010	2014	2010	2014	
2	Bandundu	Control	72%	81%	9%	9%	0.82	0.90	
		PAQUED	76%	83%	7%	6%	0.64	0.64	
	Equateur	Control	87%	71%	4%	15%	0.33	1.55	
		PAQUED	75%	72%	12%	13%	1.09	1.30	
	Orientale	Control	82%	88%	4%	3%	0.37	0.34	
		PAQUED	86%	90%	4%	2%	0.40	0.24	
	4	Bandundu	Control	44%	66%	33%	20%	3.12	1.96
			PAQUED	51%	67%	24%	17%	2.23	1.70
Equateur		Control	53%	53%	30%	28%	2.94	2.84	
		PAQUED	44%	58%	37%	27%	3.55	2.68	
Orientale		Control	62%	88%	12%	4%	1.16	0.43	
		PAQUED	72%	77%	12%	8%	1.14	0.81	
6		Bandundu	Control	35%	42%	42%	42%	4.08	4.18
			PAQUED	30%	51%	39%	27%	3.66	2.74
	Equateur	Control	37%	31%	43%	50%	4.17	5.03	
		PAQUED	34%	47%	47%	38%	4.63	3.81	
	Orientale	Control	57%	75%	20%	11%	1.87	1.12	
		PAQUED	53%	64%	24%	17%	2.39	1.66	

As shown in *Table D18*, the Initial Sound Identification subtask was difficult for students, in both 2010 and 2014. Even in Grade 6, percentages of zero scores at endline ranged from 31% in Equateur to 75% in Orientale. Similarly, at endline, students attempted relatively few items, particularly in Orientale (2–3% attempted in Grade 2; 4–8% in Grade 4; and 11–17% in Grade 6). Mean scores were also low; the highest level of competence on this task was among Grade 6 control students at endline in Equateur who were able to identify five items (50%) on average.

Table D19 reflects a comparison of differences from 2010 to 2014 across the Control and PAQUED groups.

Table D19. Initial Sound Identification Difference-in-Differences Analyses by Grade, Province, and Group

Grade 2 Province	Group	Baseline 2010		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
Bandundu	Control	0.82	0.26	0.90	0.31		
	PAQUED	0.64	0.15	0.64	0.23	-0.08	-0.03
Equateur	Control	0.33	0.20	1.55	0.69		
	PAQUED	1.09	0.22	1.30	0.28	-1.00	-0.31
Orientale	Control	0.37	0.11	0.34	0.23		
	PAQUED	0.40	0.12	0.24	0.08	-0.13	-0.09
Grade 4 Province	Group	Baseline 2010 Mean	SE	Endline 2014 Mean	SE	D-in-D	ES
Bandundu	Control	3.12	1.12	1.96	0.40		
	PAQUED	2.23	0.39	1.70	0.30	0.63	0.12
Equateur	Control	2.94	0.37	2.84	1.02		
	PAQUED	3.55	0.50	2.68	0.46	-0.76	-0.14
Orientale	Control	1.16	0.26	0.43	0.18		
	PAQUED	1.14	0.22	0.81	0.16	0.40	0.14

Grade 6 Province	Group	Baseline 2010 Mean	SE	Endline 2014 Mean	SE	D-in-D	ES
Bandundu	Control	4.08	0.83	4.18	0.53		
	PAQUED	3.66	0.49	2.74	0.35	-1.03	-0.18
Equateur	Control	4.17	0.67	5.03	0.88		
	PAQUED	4.63	0.31	3.81	0.57	-1.67	-0.30
Orientale	Control	1.87	0.59	1.12	0.28		
	PAQUED	2.39	0.27	1.66	0.23	0.01	0.00

No statistically significant differences emerged for any of the D-in-D results using the $p < 0.006$ threshold, indicating that both groups grew at comparable rates over time.

To further explore possible differences between control and PAQUED groups, *Table D20* provides a comparison of mean scores on this subtask by grade, province, group, and sex.

Table D20. Comparison of Initial Sound Identification Mean Scores by Grade, Province, Group, and Sex

Grade	Province	Group	Sex	Baseline 2010 Mean	SE	Endline 2014 Mean	SE
2	Bandundu	Control	Male	0.94	0.36	0.59	0.24
			Female	0.64	0.20	1.19	0.47
		PAQUED	Male	0.76	0.24	0.61	0.23
			Female	0.55	0.09	0.68	0.29
	Equateur	Control	Male	0.31	0.18	1.17	0.70
			Female	0.34	0.27	1.91	0.85
		PAQUED	Male	1.11	0.27	1.54	0.27
			Female	1.05	0.22	1.02	0.33
Orientale	Control	Male	0.20	0.07	0.43	0.36	
		Female	0.51	0.17	0.24	0.14	
	PAQUED	Male	0.29	0.08	0.26	0.07	
		Female	0.51	0.16	0.21	0.11	
4	Bandundu	Control	Male	3.43	1.24	1.42	0.39

Grade	Province	Group	Sex	Baseline 2010		Endline 2014		
				Mean	SE	Mean	SE	
6	Equateur	PAQUED	Female	2.89	1.03	2.44	0.55	
			Male	2.17	0.38	2.05	0.36	
			Female	2.27	0.46	1.26	0.36	
			Control	Male	3.13	0.48	2.26	1.11
				Female	2.69	0.47	3.46	1.00
				Male	3.76	0.49	3.12	0.48
		PAQUED	Female	3.35	0.60	2.23	0.49	
			Control	Male	1.37	0.61	0.61	0.34
				Female	0.98	0.24	0.16	0.10
			PAQUED	Male	1.42	0.26	0.94	0.19
		Female		0.75	0.16	0.66	0.18	
		Bandundu	Control	Male	4.31	0.70	3.51	0.80
	Female			3.78	1.05	4.84	0.74	
	Male			3.69	0.53	2.79	0.43	
	PAQUED		Female	3.61	0.48	2.69	0.35	
			Control	Male	4.33	0.70	4.74	1.15
				Female	3.77	0.70	5.35	0.68
	Equateur	PAQUED	Male	4.80	0.21	4.06	0.50	
			Female	4.38	0.51	3.48	0.69	
			Control	Male	2.37	0.55	1.39	0.48
Female		1.45		0.64	0.75	0.10		
PAQUED		Male		2.43	0.27	1.75	0.25	
		Female	2.31	0.38	1.53	0.25		
Orientale	Control	Male	2.37	0.55	1.39	0.48		
		Female	1.45	0.64	0.75	0.10		
	PAQUED	Male	2.43	0.27	1.75	0.25		
		Female	2.31	0.38	1.53	0.25		

No statistically significant differences between sexes emerged on this subtask using the $p < 0.001$ threshold.

Exploring distributions of scores across the range of possible scores (0–10 items) again shows the difficulty that students had with this subtask. **Figures D4** through **D6** illustrate these distributions by grade.

Figure D4. Grade 2 Scores on *Initial Sound Identification* at Endline by Province

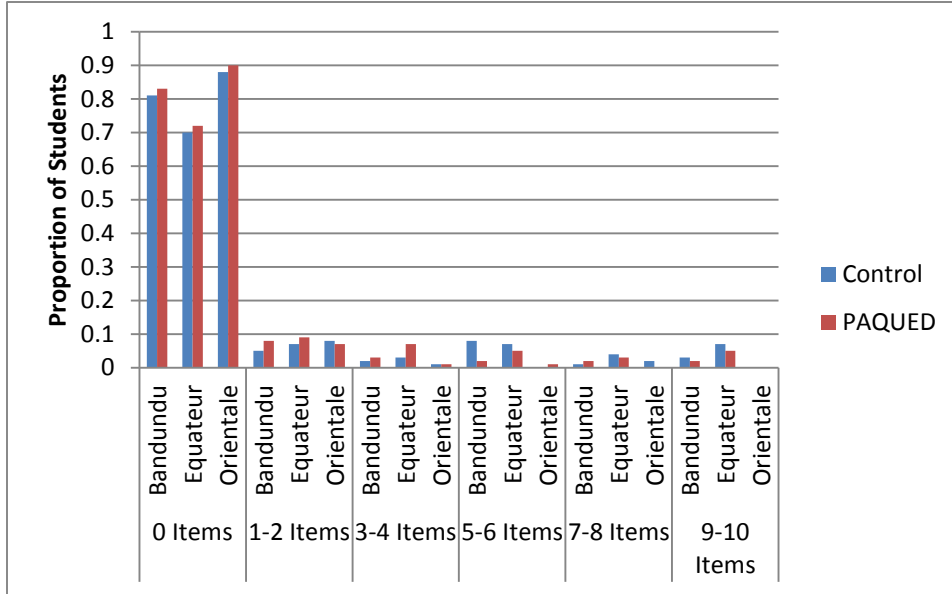


Figure D5. Grade 4 Scores on *Initial Sound Identification* at Endline by Province

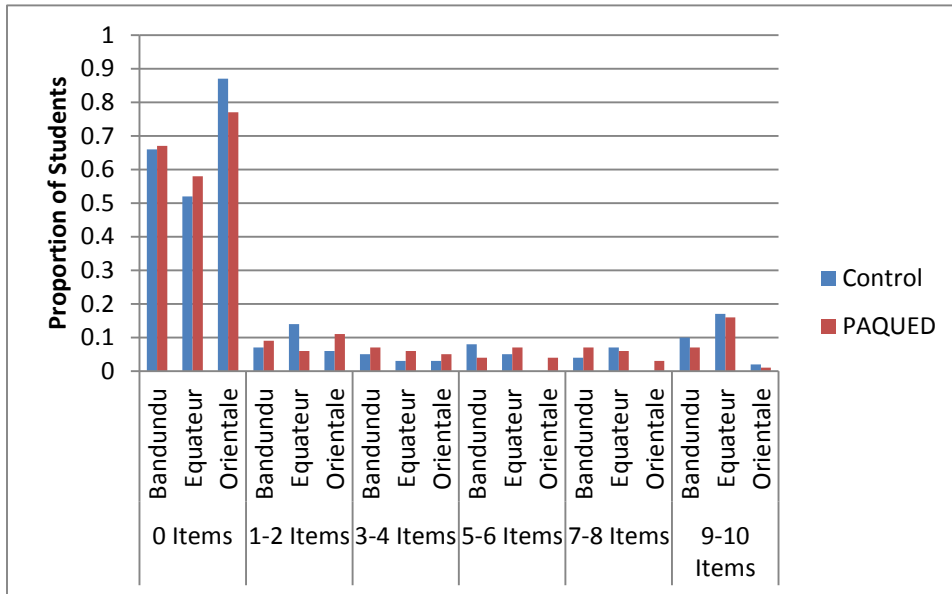
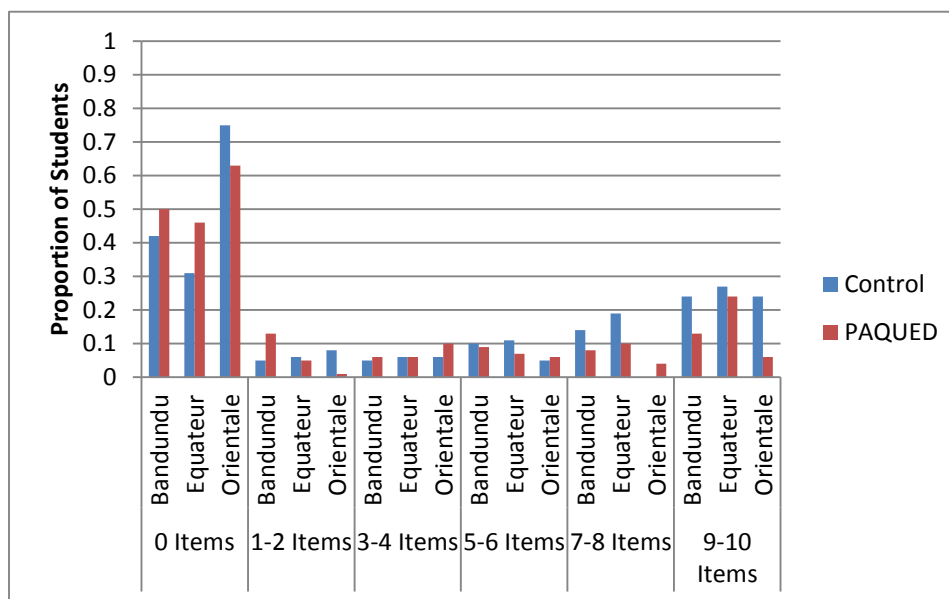


Figure D6. Grade 6 Scores on *Initial Sound Identification* at Endline by Province



As the figures illustrate, a large percentage of students scored zero on this subtask at endline in all three grades. However, by Grade 6, the distribution of scores begins to shift from zero scores to higher levels of performance. While it is apparent that students, even in Grade 6 at endline, have not yet mastered this important pre-reading skill, these figures do suggest a promising trend that can be built upon.

Listening Comprehension

For the *Listening Comprehension* subtask, students listened to a short passage and were asked five questions that assessed their basic comprehension of that passage. Scores reported for this subtask include percentages of students able to answer the comprehension questions, based upon the number of students who attempted to answer each question.

Table D21 shows the percent of zero scores and percent of items attempted for the *Listening Comprehension* subtask.

Table D21. Listening Comprehension Zero Scores and Percent Attempted by Grade and Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2010	2014	2010	2014	2010	2014
2	Bandundu	Control	60%	69%	21%	19%	1.04	0.96
		PAQUED	60%	77%	16%	7%	0.79	0.35
	Equateur	Control	77%	71%	5%	9%	0.25	0.44
		PAQUED	80%	70%	5%	14%	0.24	0.72

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score		
			2010	2014	2010	2014	2010	2014	
4	Orientale	Control	61%	75%	12%	7%	0.61	0.37	
		PAQUED	56%	72%	17%	10%	0.83	0.48	
	Bandundu	Control	55%	63%	18%	13%	0.91	0.66	
		PAQUED	46%	58%	17%	15%	0.86	0.74	
	Equateur	Control	59%	60%	9%	14%	0.45	0.68	
		PAQUED	58%	58%	10%	18%	0.52	0.90	
	Orientale	Control	43%	48%	17%	25%	0.84	1.27	
		PAQUED	34%	46%	24%	20%	1.17	1.02	
	6	Bandundu	Control	20%	30%	28%	33%	1.39	1.65
			PAQUED	19%	31%	26%	31%	1.29	1.57
Equateur		Control	29%	28%	18%	31%	0.90	1.55	
		PAQUED	32%	39%	18%	30%	0.90	1.50	
Orientale		Control	7%	28%	35%	40%	1.74	2.02	
		PAQUED	10%	20%	40%	40%	1.98	2.00	

As might be anticipated given the relatively low performance on the *Vocabulary* subtask, students struggled with the Listening Comprehension task. It is encouraging that percentages of zero scores decreased between Grade 2 and Grade 6, with percentages in Grade 6 at endline ranging from 20% (Orientale) to 39% (Equateur). Even so, these percentages, together with relatively low attempted percentages and mean scores (at endline in Grade 6, ranging from 1.50 to 2.02 questions out of the five total) reinforce the finding that students at all grades lack basic oral proficiency in French.

Table D22 reflects a comparison of differences from 2010 to 2014 across the Control and PAQUED groups. No statistically significant differences emerged for any of the D-in-D results using the $p < 0.006$ threshold, indicating that both groups grew at comparable rates over time.

Table D22. Listening Comprehension Difference-in-Differences Analyses by Grade, Province, and Group

Grade 2	Province	Group	Baseline 2010		Endline 2014		D-in-D	ES
			Mean	SE	Mean	SE		
Bandundu		Control	1.04	0.66	0.96	0.45		
		PAQUED	0.79	0.32	0.35	0.09	-0.37	-0.18
Equateur		Control	0.25	0.14	0.44	0.14		
		PAQUED	0.24	0.05	0.72	0.23	0.29	0.20
Orientale		Control	0.61	0.20	0.37	0.16		
		PAQUED	0.83	0.29	0.48	0.18	-0.10	-0.07
Grade 4	Province	Group	Baseline 2010		Endline 2014		D-in-D	ES
			Mean	SE	Mean	SE		
Bandundu		Control	0.91	0.53	0.66	0.21		
		PAQUED	0.86	0.23	0.74	0.10	0.13	0.07
Equateur		Control	0.45	0.12	0.68	0.24		
		PAQUED	0.52	0.09	0.90	0.21	0.15	0.10
Orientale		Control	0.84	0.25	1.27	0.50		
		PAQUED	1.17	0.19	1.02	0.32	-0.58	-0.36
Grade 6	Province	Group	Baseline 2010		Endline 2014		D-in-D	ES
			Mean	SE	Mean	SE		
Bandundu		Control	1.39	0.38	1.65	0.33		
		PAQUED	1.29	0.21	1.57	0.23	0.02	0.01
Equateur		Control	0.90	0.12	1.55	0.37		
		PAQUED	0.90	0.11	1.50	0.30	-0.05	-0.03
Orientale		Control	1.74	0.23	2.02	0.50		
		PAQUED	1.98	0.24	2.00	0.24	-0.26	-0.14

Table D23 provides a comparison of mean scores on this subtask by grade, province, group, and sex.

Table D23. Comparison of Listening Comprehension Mean Scores by Grade, Province, Group, and Sex

Grade	Province	Group	Sex	Baseline 2010 Mean	SE	Endline 2014 Mean	SE
2	Bandundu	Control	Male	1.33	0.85	0.56	0.17
			Female	0.55	0.28	1.31	0.69
		PAQUED	Male	0.80	0.26	0.34	0.10
			Female	0.78	0.41	0.36	0.10
	Equateur	Control	Male	0.33	0.18	0.36	0.09
			Female	0.15	0.10	0.51	0.25
		PAQUED	Male	0.23	0.05	0.75	0.24
			Female	0.24	0.05	0.69	0.23
	Orientale	Control	Male	0.64	0.23	0.49	0.16
			Female	0.59	0.20	0.24	0.17
		PAQUED	Male	0.78	0.35	0.62	0.24
			Female	0.86	0.23	0.33	0.12
4	Bandundu	Control	Male	0.93	0.61	0.54	0.12
			Female	0.91	0.47	0.76	0.37
		PAQUED	Male	0.94	0.22	0.78	0.12
			Female	0.81	0.28	0.68	0.11
	Equateur	Control	Male	0.44	0.14	0.53	0.23
			Female	0.47	0.13	0.85	0.22
		PAQUED	Male	0.53	0.10	1.05	0.23
			Female	0.51	0.12	0.74	0.20
	Orientale	Control	Male	0.98	0.24	1.15	0.41
			Female	0.71	0.29	1.43	0.64
		PAQUED	Male	1.24	0.18	1.16	0.35
			Female	1.08	0.23	0.86	0.30

Grade	Province	Group	Sex	Baseline 2010		Endline 2014		
				Mean	SE	Mean	SE	
6	Bandundu	Control	Male	1.46	0.36	1.37	0.23	
			Female	1.32	0.42	1.94	0.54	
		PAQUED	Male	1.27	0.23	1.46	0.12	
			Female	1.32	0.20	1.68	0.35	
		Equateur	Control	Male	0.89	0.14	1.63	0.44
				Female	0.94	0.15	1.46	0.33
	PAQUED		Male	0.93	0.10	1.70	0.35	
			Female	0.86	0.17	1.22	0.29	
	Orientale	Control	Male	1.75	0.32	2.03	0.44	
			Female	1.75	0.15	2.02	0.68	
		PAQUED	Male	2.09	0.24	2.17	0.25	
			Female	1.77	0.25	1.78	0.28	

Again, no statistically significant differences between sexes emerged using the $p < 0.001$ threshold.

Exploring distributions of scores across the range of possible scores (0–5 items) again shows the difficulty that students had with this subtask. *Figures D7* through *D9* illustrate these distributions by grade.

Figure D7. Grade 2 Scores on *Listening Comprehension* at Endline by Province

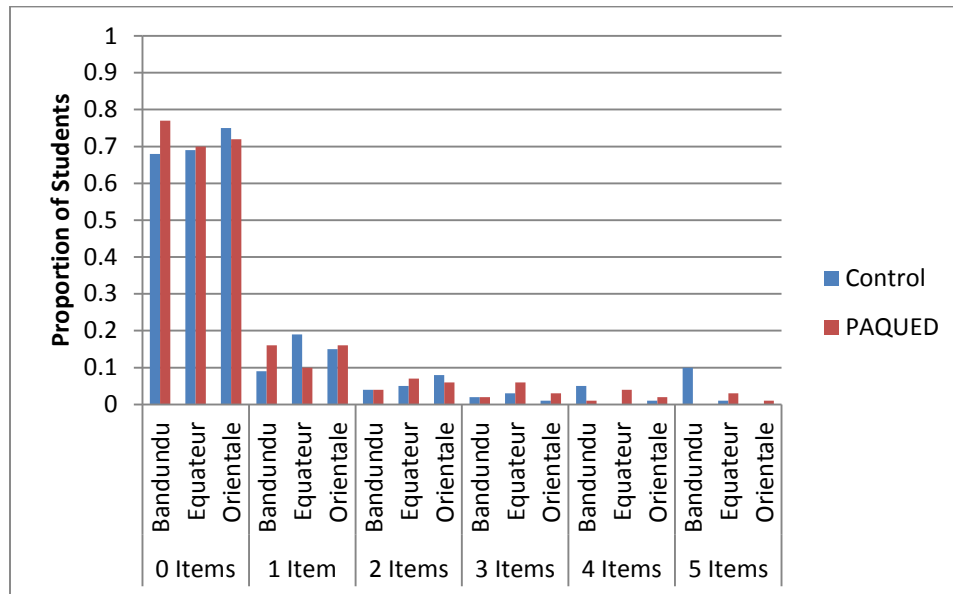


Figure D8. Grade 4 Scores on *Listening Comprehension* at Endline by Province

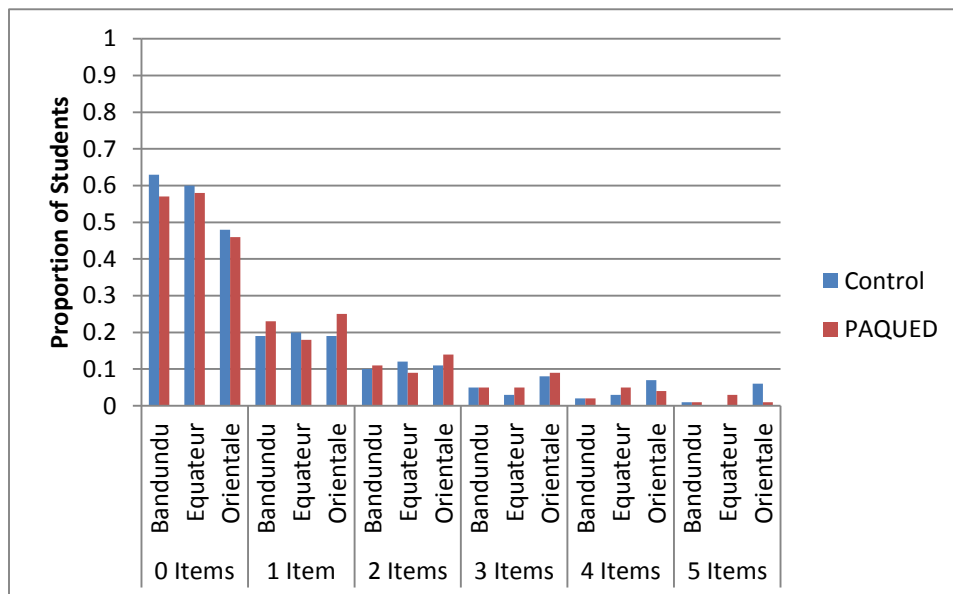
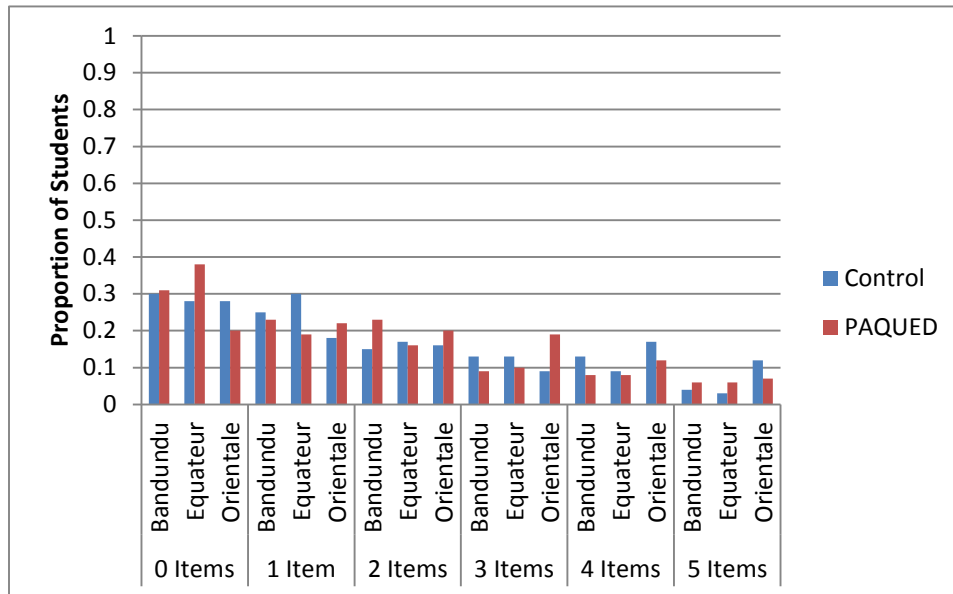


Figure D9. Grade 6 Scores on *Listening Comprehension* at Endline by Province



As illustrated above, a large percentage of students scored zero on this subtask at endline in Grade 2, with no notable difference between Control and PAQUED groups. Within Grade 6, scores were much more evenly distributed across the score points—a very promising trend compared to Grade 2—but still lower than needed to demonstrate proficiency with the French language.

Grapheme Recognition

In the *Grapheme Recognition* subtask, students were presented with a 100-item chart containing the letters of the alphabet as well as common two-letter graphemes in random order. They were required to produce the sounds or names for as many graphemes as possible within one minute. This task was discontinued prior to the end of the minute for students who were unable to produce any of the sounds for the 10 items in the first row. Scores reported for this subtask include percentages of students able to correctly generate numbers of graphemes and the number of grapheme sounds or names that students could correctly generate within one minute.

Table D24 shows the percent of zero scores and percent of items attempted for the *Grapheme Recognition* subtask.

Table D24. Grapheme Recognition Zero Scores and Percent Attempted by Grade and Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2010	2014	2010	2014	2010	2014
2	Bandundu	Control	26%	40%	21%	21%	4.34	5.98
		PAQUED	35%	54%	20%	13%	3.57	3.49
	Equateur	Control	40%	64%	18%	13%	3.77	4.04
		PAQUED	30%	60%	27%	18%	5.39	5.49
	Orientale	Control	51%	46%	17%	18%	3.53	5.34
		PAQUED	34%	47%	23%	14%	4.15	3.49
4	Bandundu	Control	8%	10%	49%	46%	15.53	14.44
		PAQUED	8%	15%	50%	42%	14.18	14.01
	Equateur	Control	6%	19%	50%	50%	16.21	19.98
		PAQUED	7%	26%	56%	48%	20.05	17.36
	Orientale	Control	4%	11%	51%	52%	13.56	19.52
		PAQUED	6%	8%	59%	47%	17.67	16.62
6	Bandundu	Control	1%	1%	72%	70%	32.03	32.87
		PAQUED	2%	4%	72%	67%	29.54	30.16
	Equateur	Control	1%	1%	74%	81%	33.08	42.42
		PAQUED	2%	8%	74%	70%	36.11	31.66
	Orientale	Control	3%	0%	80%	72%	37.10	36.58
		PAQUED	1%	1%	85%	75%	39.23	37.78

As indicated earlier, for this subtask students were given a grid of 100 letters/graphemes for which to generate sounds within 60 seconds. This means that even generating sounds at a rate of one per second would result in mean scores of 60. At endline in Grade 6, across groups, student means ranged from 30.16 to a high of 42.42, indicating that the highest performing students overall generated grapheme sounds at a rate of one sound every one to one-and-a-half seconds. This shows great improvement over performance of Grade 2 students, who were generating grapheme sounds at a rate of 3.49 cgpm to 5.96 cgpm (roughly one sound every 17 seconds and one sound every 10 seconds, respectively), and demonstrates a growing proficiency with this skill.

Table D25 reflects a comparison of differences from 2010 to 2014 across the Control and PAQUED groups. Only one statistically significant difference emerged for any of the D-in-D results using the $p < 0.006$ threshold, indicating that in general both groups grew at comparable rates over time. (The sole difference was that the performance of Grade 6 students in Control schools in Equateur increased from baseline in 2010 through endline in 2014, while performance of Grade 6 students in PAQUED schools decreased. The resultant difference-in-differences was statistically significant.)

Table D25. Grapheme Recognition Difference-in-Differences Analyses by Grade, Province, and Group

Grade 2 Province	Group	Baseline 2010 Mean	SE	Endline 2014 Mean	SE	D-in-D	ES
Bandundu	Control	4.34	0.63	5.98	1.96		
	PAQUED	3.57	0.31	3.49	0.57	-1.73	-0.19
Equateur	Control	3.77	0.49	4.04	1.92		
	PAQUED	5.39	0.79	5.49	0.78	-0.17	-0.01
Orientale	Control	3.53	1.14	5.34	2.15		
	PAQUED	4.15	0.89	3.49	0.45	-2.47	-0.28
Grade 4 Province	Group	Baseline 2010 Mean	SE	Endline 2014 Mean	SE	D-in-D	ES
Bandundu	Control	15.53	2.38	14.44	2.22		
	PAQUED	14.18	0.90	14.01	1.61	0.92	0.04
Equateur	Control	16.21	1.61	19.98	2.29		
	PAQUED	20.05	1.71	17.36	1.21	-6.46	-0.30
Orientale	Control	13.56	3.70	19.52	2.87		
	PAQUED	17.67	2.02	16.62	2.13	-7.01	-0.36
Grade 6 Province	Group	Baseline 2010 Mean	SE	Endline 2014 Mean	SE	D-in-D	ES
Bandundu	Control	32.03	2.22	32.87	3.31		
	PAQUED	29.54	1.97	30.16	3.76	-0.23	-0.01
Equateur	Control	33.08	1.98	42.42 *	3.51		
	PAQUED	36.11	1.28	31.66	1.18	-13.79	-0.60
Orientale	Control	37.10	1.69	36.58	3.42		

Grade 6 Province	Group	Baseline 2010 Mean	SE	Endline 2014 Mean	SE	D-in-D	ES
	PAQUED	39.23	1.67	37.78	2.00	-0.93	-0.04

* significant at $p < 0.006$

Table D26 provides a comparison of mean scores on this subtask by grade, province, group, and sex.

Table D26. Comparison of Grapheme Recognition Mean Scores by Grade Province, Group, and Sex

Grade	Province	Group	Sex	Baseline 2010 Mean	SE	Endline 2014 Mean	SE
2	Bandundu	Control	Male	4.98	0.64	4.54	0.57
			Female	3.28	0.69	7.27	3.35
		PAQUED	Male	3.95	0.70	3.82	0.43
			Female	3.21	0.61	3.14	0.89
	Equateur	Control	Male	4.67	0.86	4.10	1.91
			Female	2.56	0.62	3.97	1.99
		PAQUED	Male	5.77	0.66	6.34	0.93
			Female	4.93	1.19	4.50	1.07
	Orientale	Control	Male	3.65	1.53	7.42	2.73
			Female	3.42	1.22	3.13	1.51
		PAQUED	Male	4.73	0.46	4.22	0.76
			Female	3.52	1.36	2.69	0.37
4	Bandundu	Control	Male	18.45	2.60	15.64	2.31
			Female	12.47	1.83	13.37	3.47
		PAQUED	Male	16.82	1.44	15.29	1.56
			Female	12.11	1.25	12.39	2.10
	Equateur	Control	Male	20.14	2.04	23.71 *	2.26
			Female	10.72	0.86	15.85	2.95
		PAQUED	Male	23.21	2.32	19.53	1.49

Grade	Province	Group	Sex	Baseline 2010		Endline 2014		
				Mean	SE	Mean	SE	
6	Orientale	Control	Female	16.67	1.59	15.11	1.41	
			Male	15.07	3.85	22.03	2.74	
		PAQUED	Female	12.07	4.08	15.90	3.67	
			Male	19.19	2.23	19.62	2.50	
		Bandundu	Control	Male	35.88	3.04	32.61	3.94
				Female	27.16	1.70	33.13	5.57
	Equateur	PAQUED	Male	31.61	1.59	29.28	1.65	
			Female	27.89	3.14	31.06	6.20	
	Orientale	Control	Male	33.67	1.47	42.88	3.06	
			Female	31.65	3.28	41.93	4.54	
		PAQUED	Male	38.27	1.59	35.35	1.96	
			Female	33.06	1.40	26.60	1.33	
Control		Male	38.84	2.04	37.84	3.27		
		Female	35.38	2.09	34.84	4.40		
PAQUED	Male	40.08	1.90	40.45	1.85			
Female	37.54	1.60	34.21	2.74				

Only one statistically significant difference between the sexes emerged using the $p < 0.001$ threshold; Grade 4 boys in Orientale Control schools outperformed girls at endline in 2014. (No such difference was present at baseline in 2010.)

Exploring distributions of scores across the range of possible scores (0–100 items) again shows the difficulty that students had with this subtask. *Figures D10* through *D12* illustrate these distributions by grade.

Figure D10. Grade 2 Scores on *Grapheme Recognition* at Endline by Province

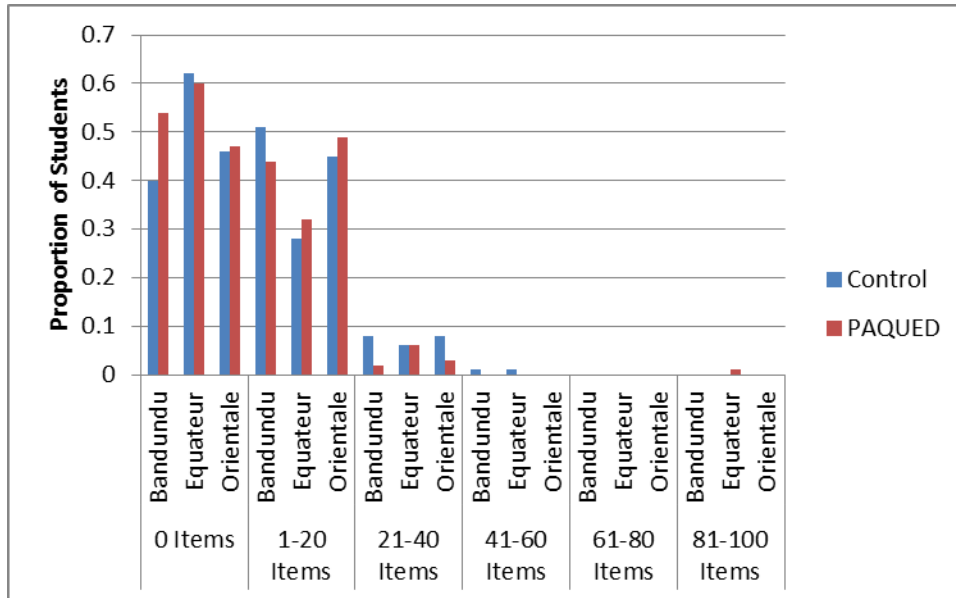


Figure D11. Grade 4 Scores on *Grapheme Recognition* at Endline by Province

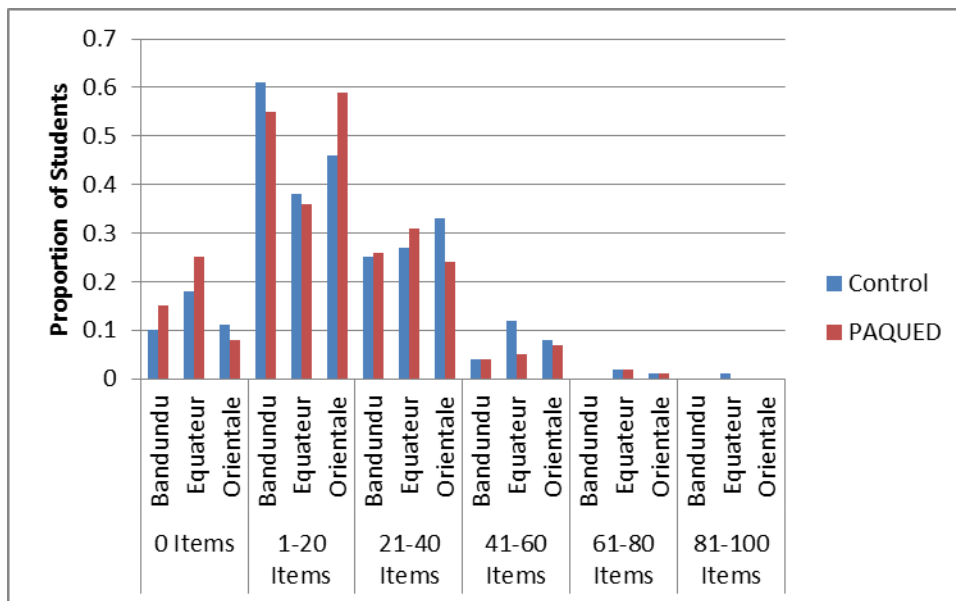
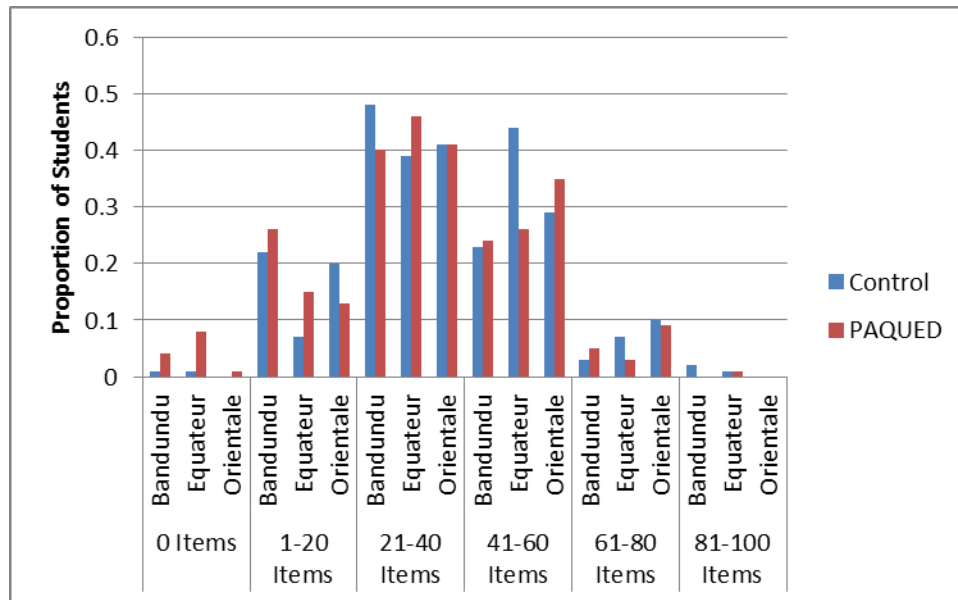


Figure D12. Grade 6 Scores on *Grapheme Recognition at Endline by Province*



As illustrated in *Figures D10* through *D12*, large percentages of students scored between zero and 20 graphemes per minute on this subtask at endline in Grade 2, with few students scoring higher than 40. Within Grade 4, fewer zero scores were observed, with the majority of students scoring between 1 and 60 graphemes per minute, and in Grade 6 a small proportion of PAQUED and Control students scored zero with the majority scoring between 21 and 60 graphemes per minute. Although this distribution shows a promising trend, ideally this skill should be mastered by baseline in Grade 2.

Familiar Word Reading

The *Familiar Word Reading* subtask was only administered to students in Grades 4 and 6. In this task, students were shown a chart of 50 familiar words (e.g., *tu* and *ami*) and were required to read as many words as they could within one minute. This subtask was discontinued before the end of one minute for students who were unable to read any of the first five words. Scores reported for this subtask include percentages of students able to correctly identify words and the number of words that students could correctly identify within one minute (i.e., cwpm).

Table D27 shows the percent of zero scores and percent of items attempted for the *Familiar Word Reading* subtask.

Table D27. Familiar Word Reading Zero Scores and Percent Attempted by Grade and Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2010	2014	2010	2014	2010	2014
4	Bandundu	Control	37%	80%	34%	9%	8.33	2.05
		PAQUED	39%	74%	29%	10%	4.85	2.55
	Equateur	Control	16%	66%	29%	18%	6.86	4.76
		PAQUED	12%	72%	44%	18%	9.06	4.55
	Orientale	Control	41%	50%	29%	25%	6.75	6.59
		PAQUED	12%	66%	53%	17%	12.61	3.94
6	Bandundu	Control	13%	32%	54%	44%	16.38	14.15
		PAQUED	8%	34%	54%	42%	14.42	14.68
	Equateur	Control	0%	12%	58%	63%	18.48	23.73
		PAQUED	3%	34%	60%	46%	20.04	13.77
	Orientale	Control	1%	29%	72%	51%	23.27	19.67
		PAQUED	1%	19%	79%	57%	27.26	20.14

As shown in *Table D27*, reading familiar words was a challenging task for these students as most students attempted fewer than two-thirds of words at endline in Grade 6. Similar to other subtasks, there are notable differences in performance between students in Grade 4 and those in Grade 6, which is promising. However, with means ranging from 13.77 to 23.73 at endline in Grade 6, there is room for improvement on a skill that should ideally be mastered in earlier grades.

Table D28 reflects a comparison of differences from 2010 to 2014 across the Control and PAQUED groups. No statistically significant differences emerged for any of the D-in-D results using the $p < 0.008$ threshold, indicating that both groups grew at comparable rates over time.

Table D28. Familiar Word Reading Difference-in-Differences Analyses by Grade, Province, and Group

Grade 4 Province	Group	Baseline 2010		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
Bandundu	Control	8.33	3.49	2.05	0.99		
	PAQUED	4.85	0.81	2.55	0.55	3.98	0.33
Equateur	Control	6.86	1.56	4.76	1.42		
	PAQUED	9.06	0.77	4.55	0.60	-2.41	-0.19
Orientale	Control	6.75	3.70	6.59	1.99		
	PAQUED	12.61	2.90	3.94	0.80	-8.52	-0.67
Grade 6 Province	Group	Baseline 2010 Mean	SE	Endline 2014 Mean	SE	D-in-D	ES
Bandundu	Control	16.38	2.77	14.15	3.02		
	PAQUED	14.42	1.11	14.68	3.81	2.50	0.11
Equateur	Control	18.48	1.24	23.73	5.15		
	PAQUED	20.04	0.80	13.77	0.99	-11.52	-0.67
Orientale	Control	23.27	2.40	19.67	4.55		
	PAQUED	27.26	3.50	20.14	1.36	-3.53	-0.18

Table D29 provides a comparison of mean scores on this subtask by grade, province, group, and sex.

Table D29. Comparison of Familiar Word Reading Mean Scores by Grade, Province, Group, and Sex

Grade	Province	Group	Sex	Baseline 2010		Endline 2014	
				Mean	SE	Mean	SE
4	Bandundu	Control	Male	11.56	3.39	1.78	0.72
			Female	5.26	2.79	2.29	1.66
		PAQUED	Male	5.32	1.08	3.21	0.68
			Female	4.30	0.82	1.71	0.46
	Equateur	Control	Male	9.23	1.75	5.55	1.82
			Female	3.29	1.05	3.91	1.30

Grade	Province	Group	Sex	Baseline 2010		Endline 2014		
				Mean	SE	Mean	SE	
6	Orientale	PAQUED	Male	10.56	0.80	5.79	0.78	
			Female	7.41	0.99	3.26	0.64	
		Control	Male	8.17	3.64	7.55	1.96	
			Female	5.54	4.28	5.21	2.23	
		PAQUED	Male	13.46	3.08	5.51	1.13	
			Female	11.16	2.69	2.14	0.69	
		Bandundu	Control	Male	18.57	3.22	16.40	2.90
				Female	13.28	2.79	11.92	5.53
	PAQUED		Male	16.48	0.97	13.28	1.69	
			Female	12.79	1.93	16.13	6.14	
	Equateur	Control	Male	19.08	0.81	24.90	4.72	
			Female	16.94	2.76	22.47	6.22	
		PAQUED	Male	21.84	1.52	16.40	1.10	
			Female	17.29	1.78	10.18	1.99	
	Orientale	Control	Male	24.96	2.17	20.61	4.25	
			Female	21.62	2.48	18.39	5.39	
PAQUED		Male	28.55	3.79	23.71 *	1.61		
		Female	25.06	3.36	15.37	1.43		

* significant at $p < 0.002$

Only one statistically significant difference between sexes emerged using the $p < 0.002$ threshold; Grade 6 boys in Orientale PAQUED schools outperformed girls at endline in 2014. (No such difference was present at baseline in 2010.)

Exploring distributions of scores across the range of possible scores (0–50 items) again shows the difficulty that students had with this subtask. **Figures D13** and **D14** illustrate these distributions.

Figure D13. Grade 4 Scores on *Familiar Word Reading* at Endline by Province

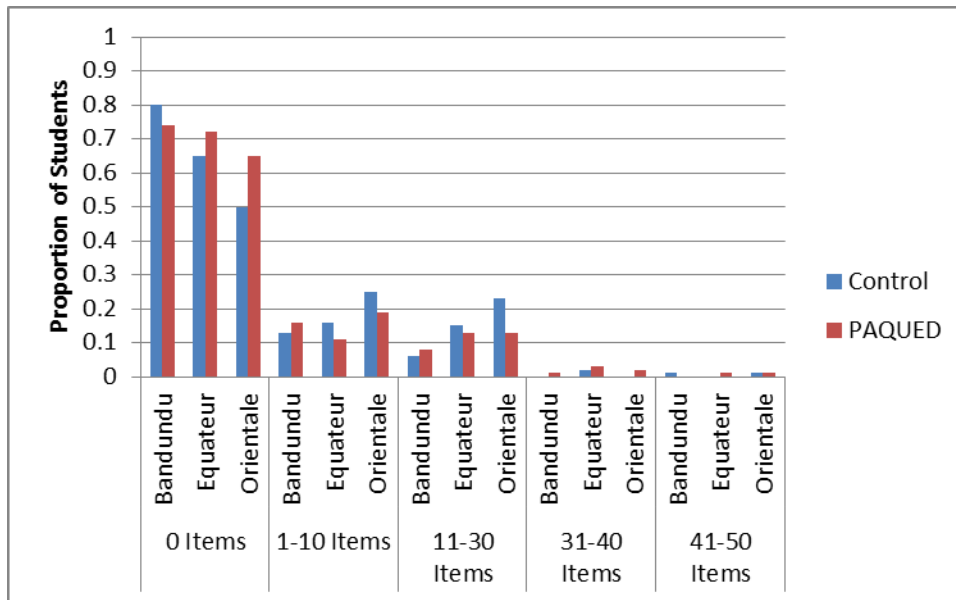
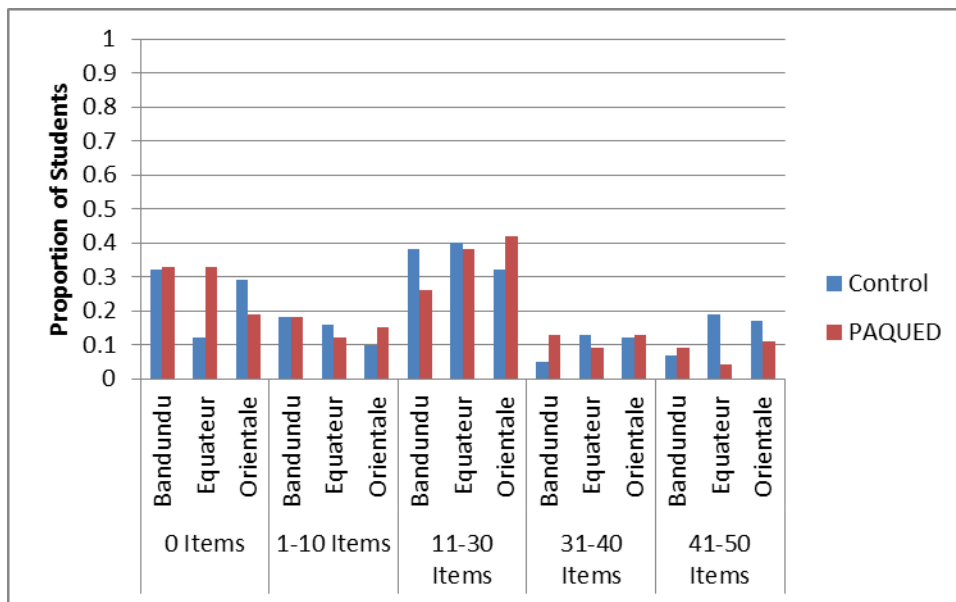


Figure D14. Grade 6 Scores on *Familiar Word Reading* at Endline by Province



As illustrated in *Figures D13* and *D14*, a large percentage (50% for the Orientale Control group and over 50% for all other groups) of students scored zero words per minute on this subtask in Grade 4, although scores did range between one and 30 words per minute. By Grade 6, the distribution was more dispersed, with nearly 20% of Control students in Equateur and Orientale reading between 41 and 50 words per minute.

Invented Word Reading

The *Invented Word Reading* subtask was also administered to only students in Grades 4 and 6. In this task, students were given a chart of 50 invented words (e.g., *tal* and *vor*) and were required to read as many words as they could within one minute. This subtask was discontinued before the end of the minute for students who were unable to read any of the first five words. Scores reported for this subtask include percentages of students able to correctly identify words and the number of words that students could correctly identify within one minute (i.e., cwpm).

Table D30 shows the percent of zero scores and percent of items attempted for the *Invented Word Reading* subtask.

Table D30. Invented Word Reading Zero Scores and Percent Attempted by Grade and Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2010	2014	2010	2014	2010	2014
4	Bandundu	Control	74%	83%	15%	9%	3.25	2.20
		PAQUED	74%	77%	12%	9%	2.11	2.02
	Equateur	Control	70%	74%	12%	15%	2.92	3.81
		PAQUED	48%	72%	26%	16%	5.23	3.94
	Orientale	Control	73%	54%	16%	22%	3.79	5.42
		PAQUED	50%	67%	31%	16%	6.72	3.54
6	Bandundu	Control	34%	30%	39%	40%	10.99	11.91
		PAQUED	36%	39%	40%	35%	10.11	10.68
	Equateur	Control	17%	19%	44%	56%	12.64	18.70
		PAQUED	19%	41%	45%	38%	13.74	10.46
	Orientale	Control	17%	28%	56%	45%	16.74	15.24
		PAQUED	11%	22%	70%	51%	21.29	15.66

As with the Familiar Reading Subtask, student scores on the *Invented Word Reading* subtask were quite low. Across the various groups, zero scores in all the provinces ranged from 54% to 83% at endline in Grade 4, and 19% to 41% in Grade 6. Mean scores at endline in Bandundu were particularly low for both Control and PAQUED groups, with higher mean scores within the Oriental province.

Table D31 reflects a comparison of differences from 2010 to 2014 across the Control and PAQUED groups. No statistically significant differences emerged for any of the D-in-D

results using the $p < 0.008$ threshold, indicating that both groups grew at comparable rates over time.

Table D31. Invented Word Reading Difference-in-Differences Analyses by Grade, Province, and Group

Grade 4 Province	Group	Baseline 2010		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
Bandundu	Control	3.25	1.42	2.20	1.24		
	PAQUED	2.11	0.34	2.02	0.38	0.96	0.10
Equateur	Control	2.92	0.82	3.81	1.31		
	PAQUED	5.23	0.52	3.94	0.58	-2.17	-0.21
Orientale	Control	3.79	2.20	5.42	1.67		
	PAQUED	6.72	2.29	3.54	0.66	-4.81	-0.44
Grade 6 Province	Group	Baseline 2010 Mean	SE	Endline 2014 Mean	SE	D-in-D	ES
Bandundu	Control	10.99	1.74	11.91	2.63		
	PAQUED	10.11	1.09	10.68	2.80	-0.35	-0.02
Equateur	Control	12.64	1.23	18.70	3.63		
	PAQUED	13.74	0.60	10.46	0.77	-9.33	-0.61
Orientale	Control	16.74	1.91	15.24	3.21		
	PAQUED	21.29	2.30	15.66	1.19	-4.15	-0.25

Table D32 provides a comparison of mean scores on this subtask by grade, province, group, and sex.

Table D32. Comparison of Invented Word Reading Mean Scores by Grade, Province, Group, and Sex

Grade	Province	Group	Sex	Baseline 2010		Endline 2014	
				Mean	SE	Mean	SE
4	Bandundu	Control	Male	4.39	1.83	1.61	0.65
			Female	2.04	0.98	2.73	2.12
		PAQUED	Male	2.43	0.60	2.53	0.43
			Female	1.80	0.41	1.39	0.41
	Equateur	Control	Male	4.27	0.95	4.44	1.79
			Female	1.04	0.38	3.12	1.09
		PAQUED	Male	6.56	0.89	4.64	0.70
			Female	3.84	0.44	3.22	0.58
	Orientale	Control	Male	4.07	1.93	6.05	1.57
			Female	3.41	2.83	4.52	1.97
		PAQUED	Male	7.95	2.73	4.95	0.96
			Female	4.93	1.74	1.94	0.42
6	Bandundu	Control	Male	12.74	1.83	13.32	2.39
			Female	8.79	1.95	10.50	4.73
		PAQUED	Male	11.33	0.75	9.83	1.31
			Female	9.12	1.83	11.54	4.40
	Equateur	Control	Male	13.34	0.82	19.35	2.85
			Female	10.95	2.16	18.00	5.10
		PAQUED	Male	15.48	1.07	13.34	1.26
			Female	11.28	1.24	6.54	1.23
	Orientale	Control	Male	17.02	2.30	16.07	3.06
			Female	16.51	2.12	14.11	3.78
		PAQUED	Male	22.23	2.49	18.42 *	1.35
			Female	19.74	2.14	11.96	1.29

* significant at $p < 0.002$

Only one statistically-significant difference emerged using the $p < 0.002$ threshold; Grade 6 boys in Orientale PAQUED schools outperformed girls at endline in 2014. (No such difference was present at baseline in 2010.) Exploring distributions of scores across the range of possible scores (0–50 items) again shows the difficulty that students had with this subtask. *Figures D15* and *D16* illustrate these distributions.

Figure D15. Grade 4 Scores on *Invented Word Reading* at Endline by Province

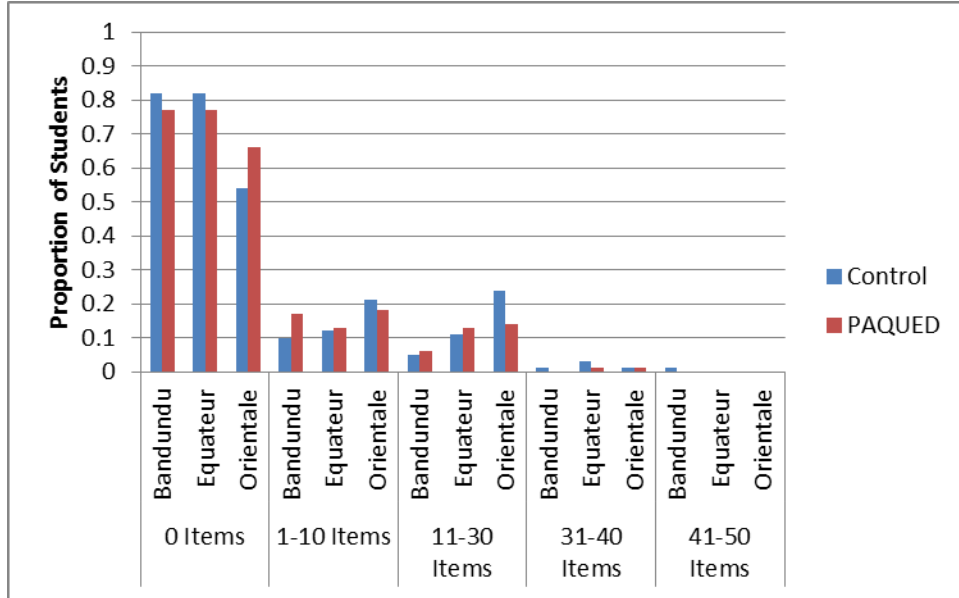
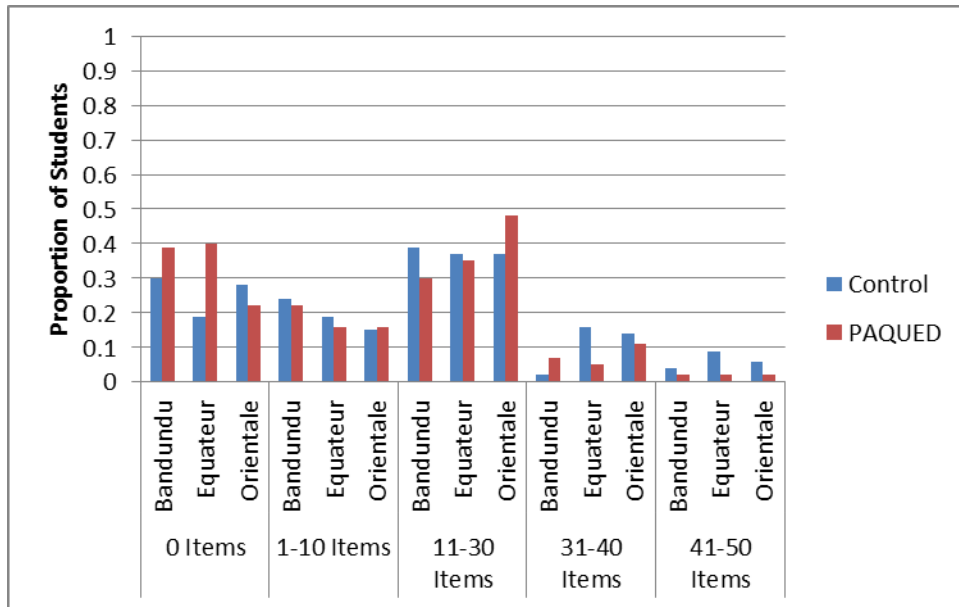


Figure D16. Grade 6 Scores on *Invented Word Reading* at Endline by Province



As exhibited in *Figures D15* and *D16*, like the Familiar Word Reading subtask, a large percentage of students scored zero non-words per minute on this subtask in Grade 4, with very few students able to read more than 30 non-words per minute. Even though in Grade 6 the distribution of scores is more dispersed, there were still relatively few students able to read more than 30 non-words in the 60 seconds. However, more Grade 6 students read between one and 30 non-words in 60 seconds than in Grade 4.

Oral Reading Fluency

The *Oral Reading Fluency* subtask was administered only to students in Grades 4 and 6. In this subtask, students were given a passage containing 54 words and were required to read as much of the passage as they could within one minute. Scores reported for this subtask include percentages of students able to read words aloud and the number of words that students could correctly read aloud within one minute.

Table D33 shows the percent of zero scores and percent of items attempted for the *Oral Reading Fluency* subtask.

Table D33. Oral Reading Fluency Zero Scores and Percent Attempted by Grade and Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2010	2014	2010	2014	2010	2014
4	Bandundu	Control	72%	71%	17%	15%	5.40	4.32
		PAQUED	65%	60%	19%	20%	4.41	5.52
	Equateur	Control	68%	54%	18%	30%	5.59	10.46
		PAQUED	59%	59%	28%	28%	8.77	9.51
	Orientale	Control	71%	38%	22%	37%	6.72	12.09
		PAQUED	53%	49%	38%	30%	11.20	8.41
6	Bandundu	Control	44%	27%	43%	56%	16.55	24.43
		PAQUED	36%	26%	47%	52%	15.50	24.72
	Equateur	Control	22%	8%	59%	74%	24.64	40.75
		PAQUED	25%	30%	57%	55%	25.40	24.42
	Orientale	Control	17%	11%	70%	64%	30.78	33.16
		PAQUED	10%	11%	82%	70%	39.42	34.63

As seen in *Table D33*, mean scores at endline on this subtask range from 24.42 (Equateur Control) to 40.75 (Equateur PAQUED) in Grade 6. Higher scores overall on this subtask,

compared to the *Familiar Word Reading* subtask, are not unusual because reading words in connected text is typically easier than reading words in isolation. That said, even mean scores of 41 words per minute indicates that students, on average, were reading one word approximately every one-and-a-half seconds. This rate of reading speed suggests that attention is being given to individual word reading that should, certainly by Grade 6, be freed up to focus on comprehending of the text being read.

Table D34 reflects a comparison of differences from 2010 to 2014 across the Control and PAQUED groups. No statistically significant differences emerged for any of the D-in-D results using the $p < 0.008$ threshold, indicating that both groups grew at comparable rates over time.

Table D34. Oral Reading Fluency Difference-in-Differences Analyses by Grade, Province, and Group

Grade 4 Province	Group	Baseline 2010		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
Bandundu	Control	5.40	1.75	4.32	2.31		
	PAQUED	4.41	0.99	5.52	0.88	2.19	0.12
Equateur	Control	5.59	1.19	10.46	2.47		
	PAQUED	8.77	0.78	9.51	1.09	-4.13	-0.20
Orientale	Control	6.72	3.73	12.09	3.77		
	PAQUED	11.20	3.46	8.41	1.60	-8.15	-0.43
Grade 6 Province	Group	Baseline 2010 Mean	SE	Endline 2014 Mean	SE	D-in-D	ES
Bandundu	Control	16.55	3.31	24.43	5.22		
	PAQUED	15.50	1.95	24.72	5.98	1.35	0.04
Equateur	Control	24.64	3.40	40.75	6.92		
	PAQUED	25.40	1.74	24.42	1.79	-17.10	-0.59
Orientale	Control	30.78	3.28	33.16	6.93		
	PAQUED	39.42	4.54	34.63	2.03	-7.17	-0.23

Table D35 provides a comparison of mean scores on this subtask by grade, province, group, and sex.

Table D35. Comparison of Oral Reading Fluency Mean Scores by Grade, Province, Group, and Sex

Grade	Province	Group	Sex	Baseline 2010		Endline 2014		
				Mean	SE	Mean	SE	
4	Bandundu	Control	Male	7.61	2.43	3.05	1.16	
			Female	3.02	1.19	5.45	3.93	
		PAQUED	Male	4.75	1.46	6.53	0.97	
			Female	4.05	0.74	4.25	1.01	
		Equateur	Control	Male	7.64	1.83	11.55	2.39
				Female	2.74	1.27	9.25	3.27
	PAQUED		Male	10.74	1.25	11.68	1.55	
			Female	6.81	1.35	7.26	1.70	
	Orientale	Control	Male	7.01	3.13	13.47	3.73	
			Female	6.30	5.21	10.09	4.31	
		PAQUED	Male	12.47	3.72	10.73	1.98	
			Female	9.49	3.21	5.76	1.46	
6	Bandundu	Control	Male	19.10	3.54	25.03	4.39	
			Female	13.19	3.87	23.83	9.28	
		PAQUED	Male	16.40	1.51	22.62	2.70	
			Female	14.77	2.86	26.86	9.38	
		Equateur	Control	Male	25.64	2.51	42.76	6.02
				Female	22.21	5.42	38.61	9.37
	PAQUED		Male	27.25	2.32	29.82	2.32	
			Female	22.78	3.24	17.03	2.41	
	Orientale	Control	Male	31.23	3.48	33.84	5.90	
			Female	30.45	3.85	32.23	8.78	
		PAQUED	Male	41.19	4.92	39.63 *	2.65	
			Female	36.48	4.43	27.95	1.96	

Only one statistically significant difference emerged using the $p < 0.002$ threshold; Grade 6 boys in Orientale PAQUED schools outperformed girls at endline in 2014. (No such difference was apparent at baseline in 2010.)

Exploring distributions of scores across the range of scores (0–55 items) again shows the difficulty that students had with this subtask. *Figures D17* and *D18* illustrate these distributions.

Figure D17. Grade 4 Scores on *Oral Reading Fluency* at Endline by Province

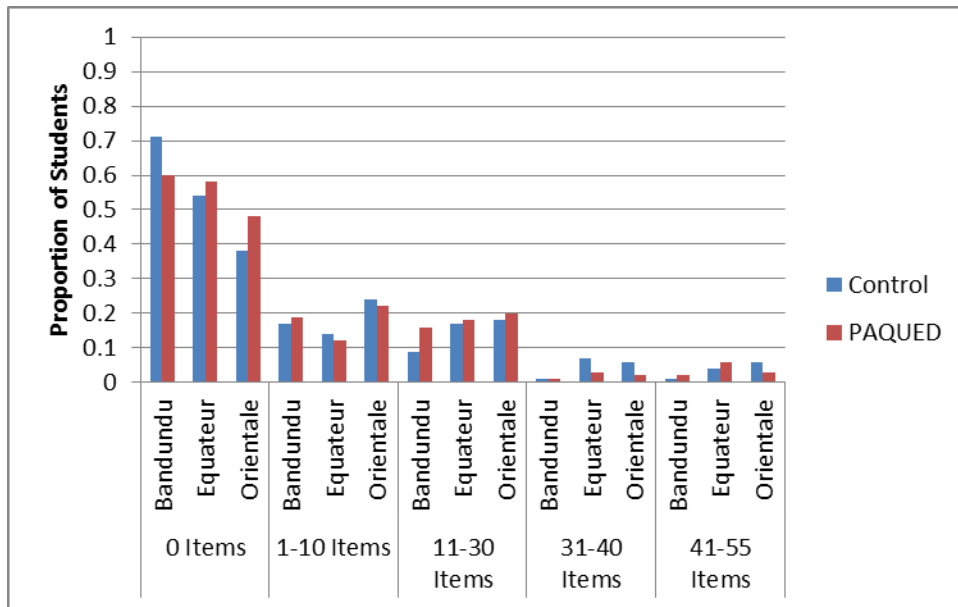
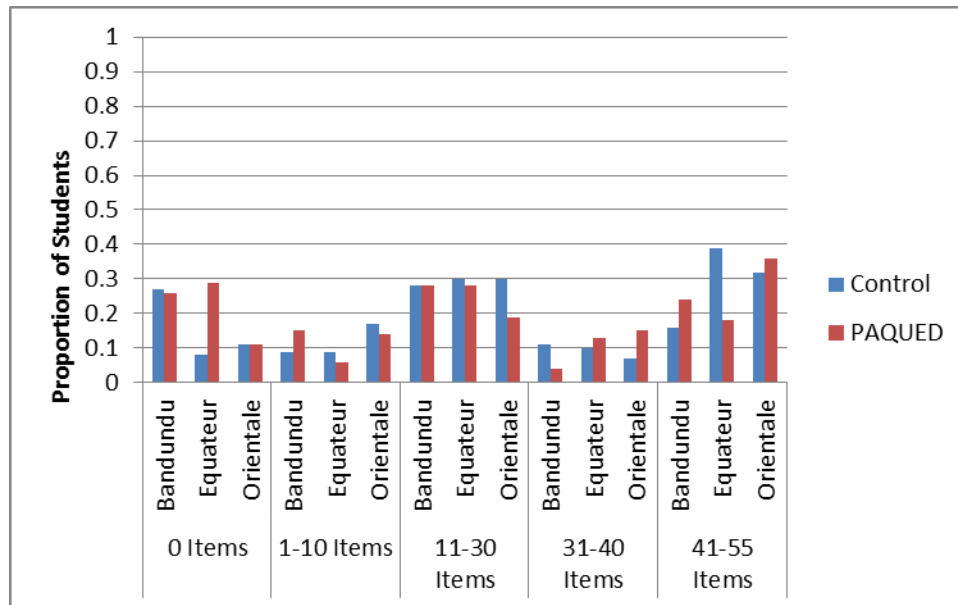


Figure D18. Grade 6 Scores on *Oral Reading Fluency* at Endline by Province



Again, while a large proportion of students in Grade 4 scored zero on this subtask, in Grade 6 the distribution is quite dispersed between zero scores, ranging from 41 to 55 words per minute. It is notable that even in Grade 6, nearly 30% of students in Bandundu (Control and PAQUED) and Equateur (PAQUED) still could not read one single word correctly. However, in general, it is encouraging to see more students reading at a rate that approaches what is needed for comprehension.

Reading Comprehension

After reading the passage in the *Oral Reading Fluency* subtask, students were asked five questions that assessed their basic comprehension of that passage. Scores reported for the *Reading Comprehension* subtask include percentages of students able to answer comprehension questions, based upon the number of students who attempted to answer each question.⁵⁸

Table D36 shows the percent of zero scores and percent of items attempted for the *Reading Comprehension* subtask.

Table D36. Reading Comprehension Zero Scores and Percent Attempted by Grade and Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2010	2014	2010	2014	2010	2014
4	Bandundu	Control	87%	96%	10%	3%	0.23	0.13
		PAQUED	86%	95%	10%	2%	0.18	0.08

⁵⁸ Students were only asked questions that corresponded with the lines of the passage that they were able to read within one minute.

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2010	2014	2010	2014	2010	2014
6	Equateur	Control	77%	91%	17%	3%	0.40	0.15
		PAQUED	67%	89%	22%	6%	0.63	0.22
	Orientale	Control	86%	79%	9%	11%	0.22	0.48
		PAQUED	73%	94%	17%	3%	0.44	0.09
	Bandundu	Control	67%	59%	16%	21%	0.48	0.85
		PAQUED	70%	65%	16%	18%	0.48	0.77
	Equateur	Control	46%	57%	28%	19%	1.04	0.87
		PAQUED	46%	72%	27%	13%	1.08	0.53
	Orientale	Control	65%	59%	13%	20%	0.51	0.93
		PAQUED	52%	49%	18%	23%	0.69	1.00

As shown in *Table D36*, most students across groups were unable to correctly attempt or respond to any comprehension questions. In all the provinces, more than 40% of students had zero scores—even at endline. Percentages of attempted items and mean scores were correspondingly low, ranging from 13% to 23% at endline in Grade 6. As mentioned before, students were only administered comprehension questions that corresponded with the text they were able to read. As seen in the *Oral Reading Fluency* subtask subsection just above, students were, on average, able to read up to 41 words. Therefore, students were administered four questions on average. However, rates of comprehension on those questions remained relatively low.

Table D37 reflects a comparison of differences from 2010 to 2014 across the Control and PAQUED groups. No statistically significant differences emerged for any of the D-in-D results using the $p < 0.008$ threshold, indicating that both groups grew at comparable rates over time.

Table D37. Reading Comprehension Difference-in-Differences Analyses by Grade, Province, and Group

Grade 4 Province	Group	Baseline 2010		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
Bandundu	Control	0.23	0.09	0.13	0.11		
	PAQUED	0.18	0.05	0.08	0.03	0.01	0.01
Equateur	Control	0.40	0.12	0.15	0.08		

Grade 4 Province	Group	Baseline 2010		Endline 2014		D-in-D	ES
		Mean	SE	Mean	SE		
	PAQUED	0.63	0.10	0.22	0.08	-0.16	-0.13
Orientale	Control	0.22	0.13	0.48	0.24		
	PAQUED	0.44	0.11	0.09	0.03	-0.60	-0.74
Grade 6 Province	Group	Baseline 2010 Mean	SE	Endline 2014 Mean	SE	D-in-D	ES
Bandundu	Control	0.48	0.08	0.85	0.31		
	PAQUED	0.48	0.16	0.77	0.30	-0.08	-0.05
Equateur	Control	1.04	0.12	0.87	0.30		
	PAQUED	1.08	0.16	0.53	0.13	-0.38	-0.24
Orientale	Control	0.51	0.07	0.93	0.39		
	PAQUED	0.69	0.10	1.00	0.14	-0.11	-0.08

Table D38 provides a comparison of mean scores on this subtask by grade, province, group, and sex.

Table D38. Comparison of Reading Comprehension Mean Scores by Grade, Province, Group, and Sex

Grade	Province	Group	Sex	Baseline 2010		Endline 2014		
				Mean	SE	Mean	SE	
4	Bandundu	Control	Male	0.37	0.15	0.02	0.01	
			Female	0.09	0.05	0.23	0.19	
		PAQUED	Male	0.18	0.05	0.09	0.03	
			Female	0.18	0.08	0.07	0.03	
		Equateur	Control	Male	0.53	0.15	0.14	0.09
				Female	0.23	0.09	0.16	0.08
	PAQUED	Male	0.86	0.14	0.28	0.10		
		Female	0.39	0.06	0.17	0.07		
	Orientale	Control	Male	0.23	0.11	0.47	0.28	
			Female	0.17	0.10	0.49	0.23	

Grade	Province	Group	Sex	Baseline 2010		Endline 2014	
				Mean	SE	Mean	SE
6		PAQUED	Male	0.44	0.12	0.11	0.05
			Female	0.45	0.13	0.06	0.03
	Bandundu	Control	Male	0.51	0.12	0.64	0.15
			Female	0.44	0.09	1.06	0.54
		PAQUED	Male	0.60	0.22	0.51	0.10
			Female	0.39	0.14	1.03	0.46
	Equateur	Control	Male	1.05	0.12	0.87	0.37
			Female	1.02	0.19	0.87	0.27
		PAQUED	Male	1.16	0.19	0.68	0.17
			Female	0.97	0.13	0.32	0.08
	Orientale	Control	Male	0.65	0.06	1.02	0.37
			Female	0.36	0.07	0.81	0.46
		PAQUED	Male	0.66	0.08	1.09	0.17
			Female	0.74	0.17	0.90	0.14

No statistically significant differences emerged between sexes in any of the provinces using the $p < 0.002$ threshold.

Exploring distributions of scores across the range of scores (0–5 items) again shows the difficulty that students had with this subtask. *Figures D19* and *D20* illustrate these distributions.

Figure D19. Grade 4 Scores on Reading Comprehension at Endline by Province

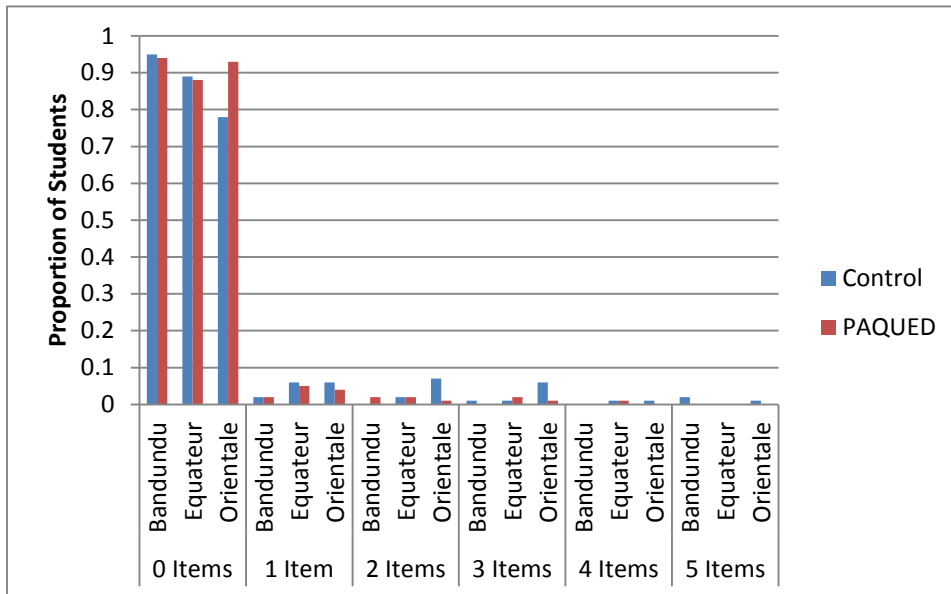
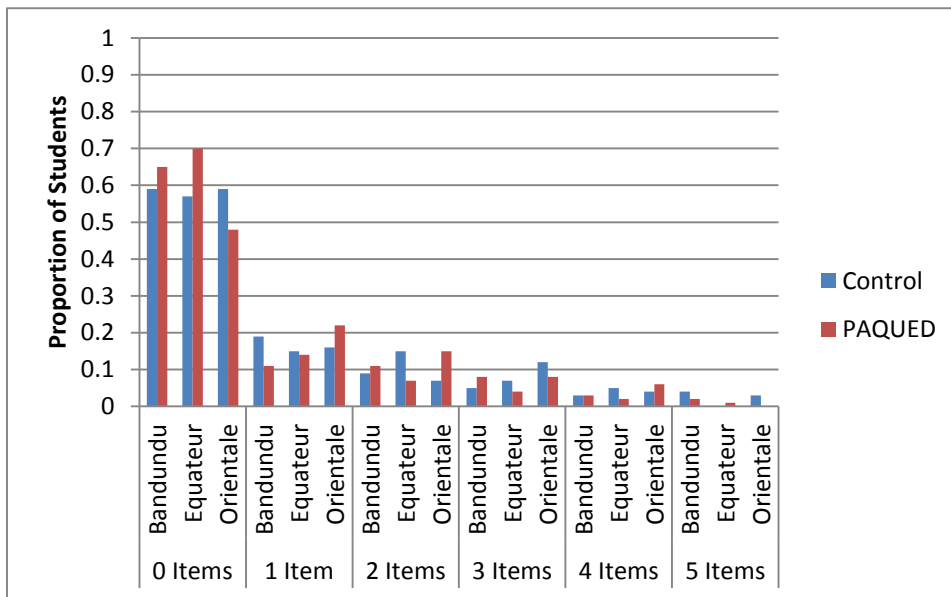


Figure D20. Grade 6 Scores on Reading Comprehension at Endline by Province



Given the low overall scores on the *Oral Reading Fluency* subtask, students were administered a limited number of comprehension questions to answer. Students scoring zero on the Reading Comprehension subtask included students who did not read far enough into the passage to receive the first question and those who received questions, but did not answer them correctly. As seen in **Figures D19** and **D20**, the large majority of students scored zero on this subtask in Grade 4. Even in Grade 6, more than 50% of students scored zero in all groups except for Orientale PAQUED students.

Another informative way to look at the relationship between oral reading fluency and reading comprehension is to explore which reading fluency levels correspond with what levels of reading comprehension, as displayed in *Figures D21* and *D22*.

Figure D21. Grade 4 Overall Correspondence between Oral Reading Fluency and Reading Comprehension

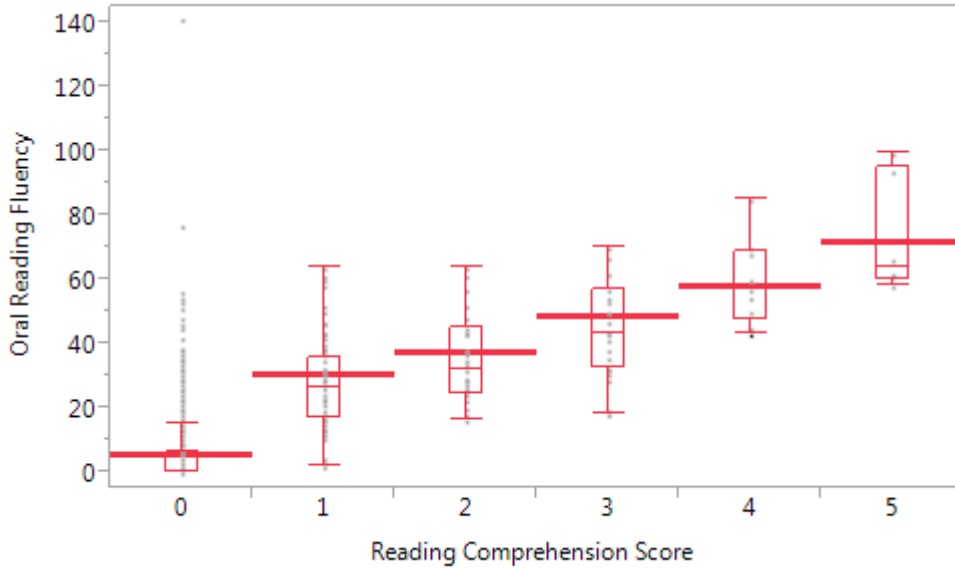
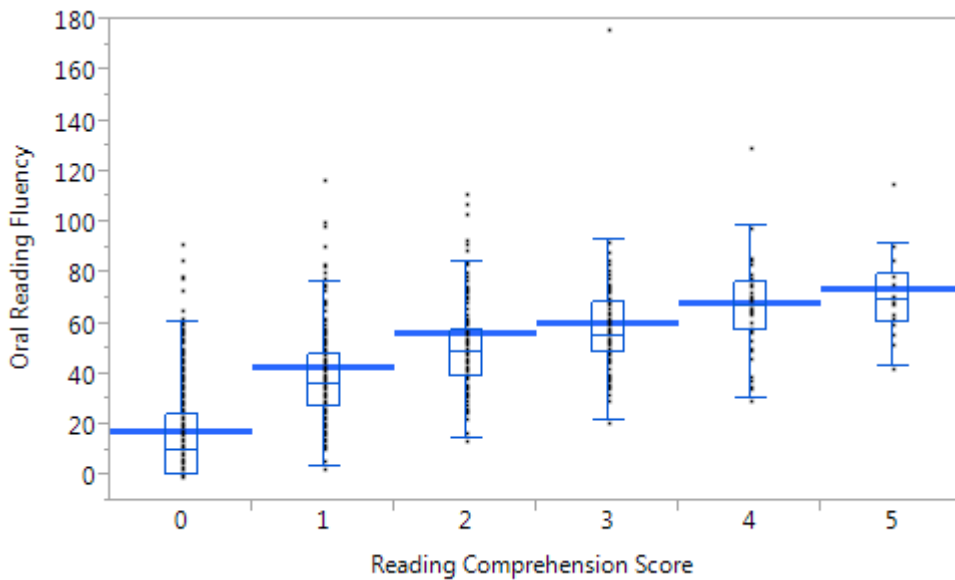


Figure D22. Grade 6 Overall Correspondence between Oral Reading Fluency and Reading Comprehension



Figures D21 and **D22** clearly illustrate the relationship between reading fluency levels and reading comprehension. While many outliers exist, in general Grade 4 students reading between 30 and 40 words per minute tend to accurately respond to one comprehension question, whereas Oral Reading Fluency scores of 60 to 70 correspond with reading comprehension scores of four and five. In Grade 6, Oral Reading Fluency scores of 40 to 50 correspond with a reading comprehension score of one, while scores between 70 and 80 words per minute correspond with reading comprehension scores of four and five.

Dictation

For the *Dictation* subtask, examiners read a short sentence to students (“*Mon ami s’appelle Bola.*”), and students attempted to write the sentence. The three target words that were scored were “*mon*”, “*ami*”, and “*appellee*”. Scores reported for this subtask include the number of words spelled correctly.

Table D39. Dictation Zero Scores and Percent Attempted by Grade and Province

Grade	Province	Group	% Zero Scores		% Attempted		Mean Score	
			2010	2014	2010	2014	2010	2014
4	Bandundu	Control	5%	61%	35%	21%	2.03	0.59
		PAQUED	4%	54%	33%	25%	1.98	0.71
	Equateur	Control	7%	44%	32%	34%	1.91	0.97
		PAQUED	4%	44%	34%	34%	2.03	0.97
	Orientale	Control	10%	48%	26%	28%	1.51	0.81
		PAQUED	5%	55%	37%	23%	2.07	0.68
6	Bandundu	Control	1%	22%	50%	46%	2.97	1.39
		PAQUED	0%	27%	48%	49%	2.85	1.46
	Equateur	Control	3%	15%	46%	65%	2.76	1.93
		PAQUED	0%	27%	47%	44%	2.83	1.32
	Orientale	Control	1%	21%	50%	47%	2.90	1.42
		PAQUED	1%	20%	57%	54%	3.23	1.60

Table D39 shows the percent of zero scores and percent of items attempted for the *Dictation* subtask. Interestingly, zero scores on this subtask at endline (2014) were higher than at baseline (2010). The highest percentages of zero scores in Grade 6 at endline were observed in Bandundu, with 22% of Control students and 27% of PAQUED students

scoring zero. This reflects a pervasive difficulty with the skill of dictation even in Grade 6. Interestingly, mean scores in both grades decrease from baseline (2010) to endline (2014). This is a notable finding that warrants attention in future testing.

Table D40 reflects a comparison of differences from 2010 to 2014 across the Control and PAQUED groups. No statistically significant differences emerged for any of the D-in-D results using the $p < 0.008$ threshold, indicating that both groups grew at comparable rates over time.

Table D40. Dictation Difference-in-Differences Analyses by Grade, Province, and Group

Grade 4		Baseline 2010		Endline 2014		D-in-D	ES
Province	Group	Mean	SE	Mean	SE		
Bandundu	Control	2.03	0.19	0.59	0.13		
	PAQUED	1.98	0.09	0.71	0.07	0.18	0.10
Equateur	Control	1.91	0.32	0.97	0.14		
	PAQUED	2.03	0.15	0.97	0.11	-0.12	-0.07
Orientale	Control	1.51	0.25	0.81	0.16		
	PAQUED	2.07	0.15	0.68	0.08	-0.69	-0.50
Grade 6		Baseline 2010		Endline 2014		D-in-D	ES
Province	Group	Mean	SE	Mean	SE		
Bandundu	Control	2.97	0.09	1.39	0.19		
	PAQUED	2.85	0.16	1.46	0.18	0.19	0.11
Equateur	Control	2.76	0.27	1.93	0.21		
	PAQUED	2.83	0.20	1.32	0.11	-0.68	-0.42
Orientale	Control	2.90	0.18	1.42	0.21		
	PAQUED	3.23	0.20	1.60	0.07	-0.15	-0.10

Table D41 provides a comparison of mean scores on this subtask by grade, province, group, and sex.

Table D41. Comparison of Dictation Mean Scores by Grade, Province, Group, and Sex

Grade	Province	Group	Sex	Baseline 2010		Endline 2014	
				Mean	SE	Mean	SE
4	Bandundu	Control	Male	2.24	0.22	0.66	0.20
			Female	1.83	0.14	0.53	0.21

Grade	Province	Group	Sex	Baseline 2010		Endline 2014		
				Mean	SE	Mean	SE	
6	Equateur	PAQUED	Male	2.14	0.17	0.88	0.10	
			Female	1.89	0.17	0.50	0.09	
		Control	Male	2.17	0.28	1.10	0.14	
			Female	1.56	0.27	0.82	0.20	
		PAQUED	Male	2.19	0.23	1.07	0.14	
			Female	1.87	0.09	0.84	0.12	
	Orientale	Control	Male	1.53	0.24	0.96	0.19	
			Female	1.48	0.27	0.59	0.21	
		PAQUED	Male	2.22	0.16	0.75	0.08	
			Female	1.84	0.15	0.61	0.12	
		Bandundu	Control	Male	3.11	0.14	1.50	0.18
				Female	2.79	0.08	1.26	0.34
	PAQUED		Male	2.98	0.12	1.49	0.08	
			Female	2.75	0.22	1.44	0.37	
	Equateur	Control	Male	2.82	0.18	1.79	0.24	
			Female	2.61	0.46	2.07	0.22	
		PAQUED	Male	2.94	0.27	1.43	0.13	
			Female	2.66	0.14	1.15	0.10	
Orientale		Control	Male	2.75	0.19	1.42	0.18	
			Female	2.97	0.24	1.42	0.28	
	PAQUED	Male	3.28	0.19	1.81	0.10		
		Female	3.17	0.22	1.30	0.09		

No statistically significant difference emerged between sexes in any of the provinces using the $p < 0.002$ threshold.

Exploring distributions of scores across the range of scores (0–3 items) again shows the difficulty that students had with this subtask. *Figures D23* and *D24* illustrate these distributions.

Figure D23. Grade 4 Scores on *Dictation* at Endline by Province

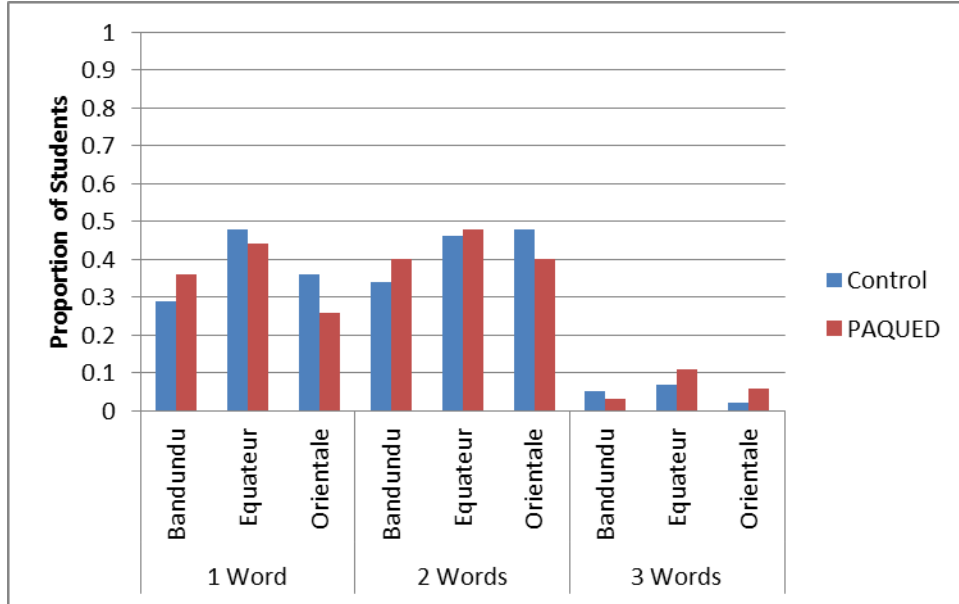
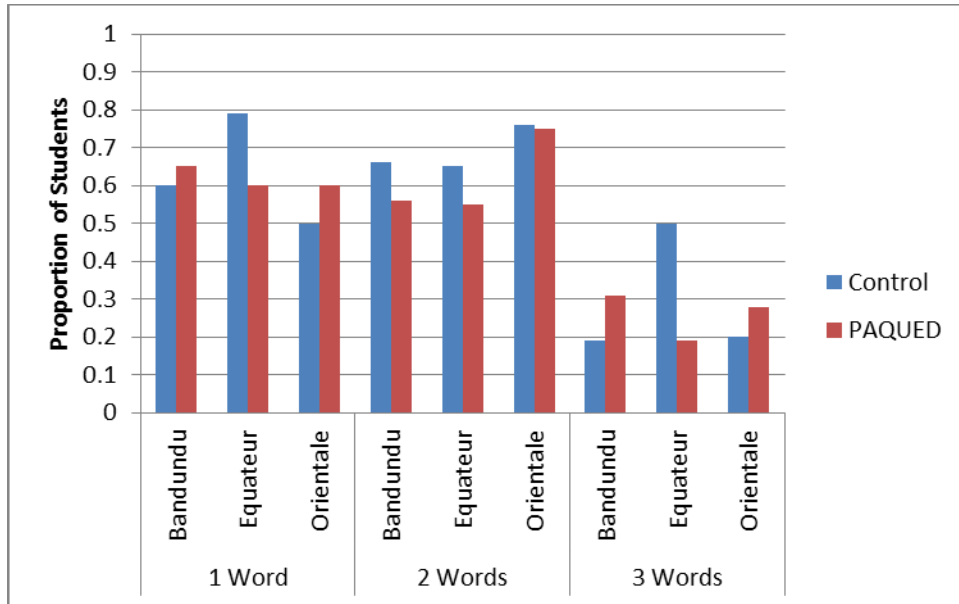


Figure D24. Grade 6 Scores on *Dictation* at Endline by Province



Within the *Dictation* subtask, students were scored on fully correct encoding of three words in a longer sentence. As illustrated in *Figures D23* and *D24*, Grade 4 student

scores were relatively evenly distributed across one and two words correct. Fewer than 20% of students overall were able to correctly write all three words. However, in Grade 6, student performance was stronger for all groups.

Student Characteristics Analyses

A series of chi-square tests were run to determine the extent to which relevant student characteristics are correlated with student performance. Because the greatest student literacy gains appear to be on the *Grapheme Recognition* subtask, that skill as measured at endline was used as an indicator of student competency. In addition, given relatively low levels of performance even on this subtask, only high-performing students (i.e., those scoring in the top quintile on this subtask) were included in the following analyses. **Table D42** shows weighted percentages, chi-square statistics, and p-values for each of the student characteristics identified earlier in this report as being of theoretical interest for this purpose.

Table D42. Chi-Squared Analyses of Grade 2 Student Characteristics with Student High Performance on the Grapheme Recognition Subtask

Student Characteristic	Weighted Percentages		
Student has reading book in class			
	Not high-Performing Students	High-performing Students	Total
No	93%	87%	92%
Yes	7%	14%	9%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 47.9186$			
Design-based $F(1.00, 25.00) = 4.1631$ $p = 0.052$			
Someone in the student's home is able to read			
No	23%	18%	22%
Yes	77%	82%	78%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 13.7059$			
Design-based $F(1.00, 25.00) = 0.8649$ $p = 0.361$			

Student Characteristic	Weighted Percentages		
Student has at least one book at home			
No	88%	86%	88%
Yes	12%	14%	12%
Total	100%	100%	100%

Pearson: Uncorrected $\chi^2(1) = 3.4217$

Design-based $F(1.00, 25.00) = 0.2210$ $p = 0.642$

Student attended kindergarten			
	Not high-Performing Students	High-performing Students	Total
No	84%	76%	82%
Yes	16%	24%	18%
Total	100%	100%	100%

Pearson: Uncorrected $\chi^2(1) = 36.0098$

Design-based $F(1.00, 25.00) = 2.8658$ $p = 0.103$

If teacher assigns homework, student has someone at home to help with it			
No	45%	59%	49%
Yes	55%	41%	51%
Total	100%	100%	100%

Pearson: Uncorrected $\chi^2(1) = 64.9685$

Design-based $F(1.00, 25.00) = 7.0944$ $p = 0.013$

Student speaks French at home			
No	95%	91%	94%
Yes	5%	10%	6%
Total	100%	100%	100%

Pearson: Uncorrected $\chi^2(1) = 32.2163$

Design-based $F(1.00, 25.00) = 2.4737$ $p = 0.128$

Table D42 shows that there were no statistically significant correlations between these student characteristics at high performance on the *Grapheme Recognition* subtask using the $p < 0.002$ threshold.

Logistic regression analyses were also run on these student characteristics to further explore their relationship with student performance, when province, gender, and student SES were held constant. These analyses revealed no statistically significant relationships.

Table D43 shows weighted percentages, chi-square statistics, and p-values for each of the Grade 4 student characteristics identified earlier in this chapter as being of theoretical interest for this purpose.

Table D43. Chi-Squared Analyses of Grade 4 Student Characteristics with Student High Performance on the Grapheme Recognition Subtask

Student Characteristic	Weighted Percentages		
Student has reading book in class			
	Not high-Performing Students	High-performing Students	Total
No	84%	87%	85%
Yes	16%	13%	15%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 4.2111$			
Design-based $F(1.00, 25.00) = 0.4345$ $p = 0.516$			
Someone in the student's home is able to read			
No	19%	21%	20%
Yes	81%	79%	80%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 1.4686$			
Design-based $F(1.00, 25.00) = 0.1651$ $p = 0.688$			
Student has at least one book at home			
No	80%	83%	81%
Yes	20%	17%	19%

Student Characteristic	Weighted Percentages		
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 4.6050$			
Design-based $F(1.00, 25.00) = 0.3469$ $p = 0.561$			
Student attended kindergarten			
No	84%	76%	82%
Yes	16%	25%	18%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 45.3706$			
Design-based $F(1.00, 25.00) = 3.3816$ $p = 0.078$			
If teacher assigns homework, student has someone at home to help with it			
No	57%	55%	56%
Yes	44%	45%	44%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 1.2380$			
Design-based $F(1.00, 25.00) = 0.0581$ $p = 0.811$			
Student speaks French at home			
No	96%	82%	93%
Yes	4%	18%	7%
Total	100%	100%	100%
Pearson: Uncorrected $\chi^2(1) = 283.1895$			
Design-based $F(1.00, 25.00) = 34.1842$ $p = 0.000$			

Table D43 shows that the student speaking French at home was the one characteristic for Grade 4 students that had a significant correlation (at $p < 0.002$) with high performance on the *Grapheme Recognition* subtask.

Logistic regression analyses were also run on these student characteristics to further explore their relationship with student performance, when province, gender, and student

SES were held constant. These analyses, presented in *Table D44*, showed the following statistically significant relationships.

Table D44. Grade 4 Student Characteristics Logistic Regressions

Characteristic	p-value
Student speaks French in the home	0.000*

* significant at $p < 0.002$

Table D45 shows weighted percentages, chi-square statistics, and p-values for each of the Grade 6 student characteristics identified earlier in this chapter as being of theoretical interest for this purpose.

Table D45. Chi-Squared Analyses of Grade 6 Student Characteristics with Student High Performance on the Grapheme Sound Knowledge Subtask

Student Characteristic	Weighted Percentages		
	Not high-Performing Students	High-performing Students	Total
Student has reading book in class			
No	85%	75%	82%
Yes	15%	25%	18%
Total	100%	100%	100%

Pearson: Uncorrected $\chi^2(1) = 73.6352$

Design-based $F(1.00, 25.00) = 5.8873$ $p = 0.023$

Someone in the student's home is able to read			
No	16%	10%	14%
Yes	85%	90%	86%
Total	100%	100%	100%

Pearson: Uncorrected $\chi^2(1) = 25.3480$

Design-based $F(1.00, 25.00) = 4.5932$ $p = 0.042$

Student Characteristic	Weighted Percentages		
------------------------	----------------------	--	--

Student has at least one book at home			
---------------------------------------	--	--	--

	Not high-Performing Students	High-performing Students	Total
No	77%	74%	76%
Yes	24%	26%	24%
Total	100%	100%	100%

Pearson: Uncorrected $\chi^2(1) = 4.1864$

Design-based $F(1.00, 25.00) = 0.2058$ $p = 0.654$

Student attended kindergarten			
-------------------------------	--	--	--

No	84%	74%	81%
Yes	16%	26%	19%
Total	100%	100%	100%

Pearson: Uncorrected $\chi^2(1) = 70.5736$

Design-based $F(1.00, 25.00) = 3.8185$ $p = 0.062$

If teacher assigns homework, student has someone at home to help with it			
--	--	--	--

No	62%	60%	61%
Yes	39%	40%	39%
Total	100%	100%	100%

Pearson: Uncorrected $\chi^2(1) = 0.4927$

Design-based $F(1.00, 25.00) = 0.0284$ $p = 0.867$

Student speaks French at home			
-------------------------------	--	--	--

No	93%	89%	91%
Yes	7%	12%	9%
Total	100%	100%	100%

Pearson: Uncorrected $\chi^2(1) = 24.5393$

Design-based $F(1.00, 25.00) = 2.4188$ $p = 0.132$

As *Table D45* demonstrates, for Grade 6 students there were no characteristics that correlated with high performance on the *Grapheme Recognition* subtask using the $p < 0.002$ threshold.

Logistic regression analyses were also run on these student characteristics to further explore their relationship with student performance, when province, gender, and student SES were held constant. These analyses showed no statistically significant relationships.

Teacher Characteristics Analyses

Two principal component factor indices were established to explore the relationship between teacher characteristics and student achievement on the *Grapheme Recognition* subtask, which generated the following composite factors:

1. Teacher participation—which includes frequency of reported visits by PAQUED personnel, teacher participation in exchange forums at the cluster level, and teacher participation in exchange forums at the school level; and
2. Teacher access to materials—which includes resources the teacher received from PAQUED, number of radios received from PAQUED, and the number of PAQUED kits the teacher reported using.

Regression analyses on these two composites showed no significant substantial impact on student performance.

Logistic regression analyses were also run on these teacher characteristics to further explore their relationship with student performance, when province, gender, and student SES were held constant. These analyses showed no statistically significant relationships.

Head Teacher Characteristics Analyses

It was hypothesized that the sex of the head teacher might impact the extent to which teachers participated in exchange forums and followed the IAI interactive lessons, but this was not the case. No significant relationship between the sex of the head teacher and teacher participation in exchange or IAI interactive lessons emerged.

3. Summary and Conclusions

In general, no consistent, statistically significant trends emerged from the analyses presented in this section on student performance on EGRA measures in the PAQUED schools, either when disaggregating data by region or by sex. However, promising trends can be seen. Overall, as the grades being assessed increases (Grade 2 to Grade 4 to Grade 6), student performance increases; however, further investigation is warranted in cases where performance decreased over time (from 2010 to 2014), whether due to testing error or other factors in the classroom.

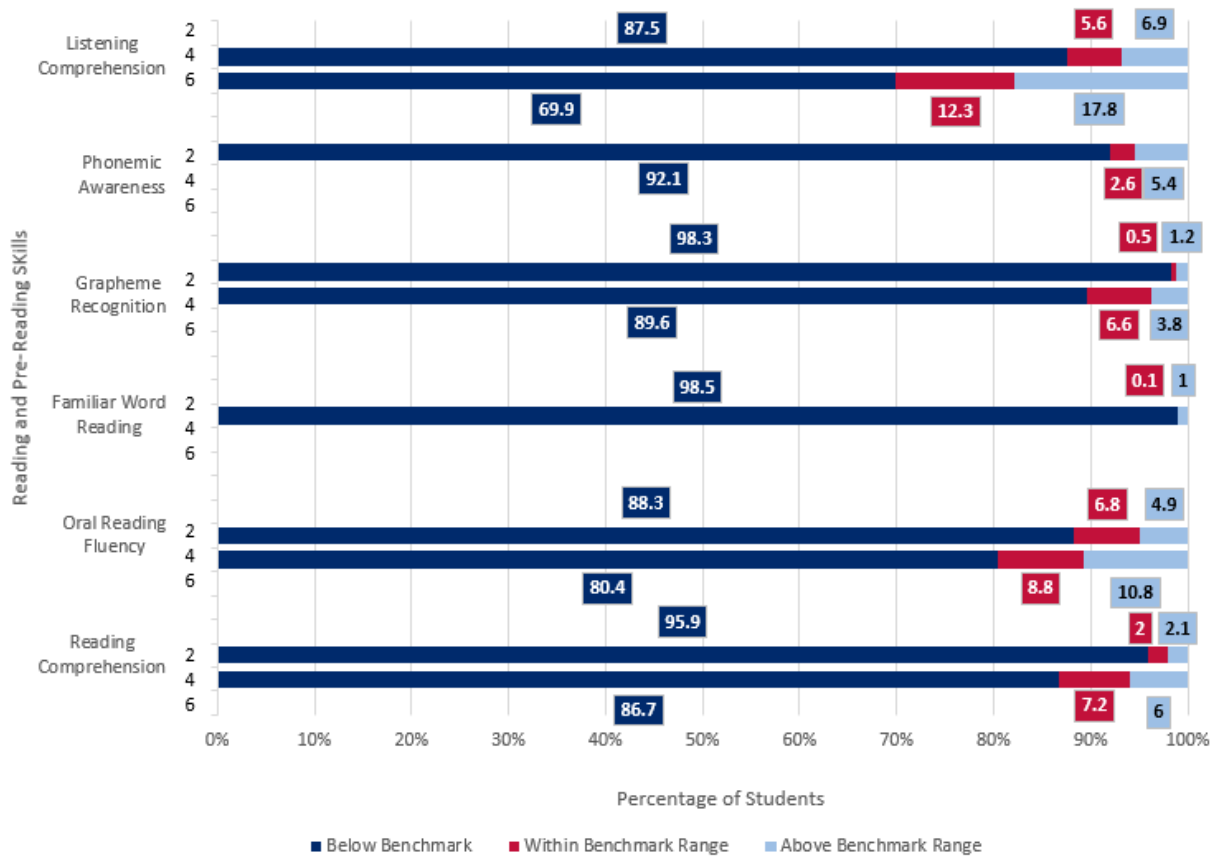
Overall student performance overall at endline is below what is needed to meet national benchmarks, across all subtasks. Even for the oral vocabulary skills of vocabulary, phonemic awareness, and listening comprehension, students failed to demonstrate French oral skills required to effectively read in French. Student mean scores on grapheme and word recognition, as well as connected text reading, were also lower than necessary for reading with fluency and comprehension. *Table D46* and *Figure D25* illustrate this.

Table D46. Student Performance in PAQUED Schools Relative to National Benchmarks⁵⁹

Grade	Reading or Pre-reading Skill	Below benchmark		Within benchmark range		Above benchmark range	
		<i>n</i>	(%)	<i>N</i>	(%)	<i>n</i>	(%)
2	Listening Comprehension						
	Phonemic Awareness	1647	(92.1%)	50	(2.6%)	98	(5.4%)
	Graphemes	1769	(98.3%)	9	(0.5%)	17	(1.2%)
	Familiar Words						
	Oral Reading Fluency						
	Comprehension						
4	Listening Comprehension	1541	(87.5%)	96	(5.6%)	108	(6.9%)
	Phonemic Awareness						
	Graphemes	1602	(89.6%)	88	(6.6%)	55	(3.8%)
	Familiar Words	1715	(98.5%)	5	(0.1%)	19	(1%)
	Oral Reading Fluency	1597	(88.3%)	73	(6.8%)	75	(4.9%)
	Comprehension	1688	(95.9%)	23	(2%)	34	(2.1%)
6	Listening Comprehension	1277	(69.9%)	201	(12.3%)	237	(17.8%)
	Phonemic Awareness						
	Graphemes						
	Familiar Words						
	Oral Reading Fluency	1507	(80.4%)	111	(8.8%)	97	(10.8%)
	Comprehension	1551	(86.7%)	92	(7.2%)	72	(6%)

⁵⁹ The *n* for each grade presented in this table is as follows: Grade 2 = 1795; Grade 4 = 1745; Grade 6 = 1715. Each *n* presented is unweighted, and each percentage presented is weighted.

Figure D25. Performance of Students in PAQUED Schools Relative to DRC Benchmarks



The sole student characteristic which had a significant correlation with high performance on the *Grapheme Recognition* subtask was speaking French at home, but that was only for Grade 4 students.

Chapter E: Results and Analysis of Student Mathematics Performance in Reading Program, Accessible PAQUED, and Accessible Control Schools

Chapter E presents EGMA results for the Accessible PAQUED, Accessible Control, and Reading Program populations. (*Chapter F* presents EGMA results for the PAQUED and Control populations.) EGMA results for the Reading Program schools were presented in *Chapter B* because there was a substantive difference between the reading-focused intervention they received and the intervention that was provided to the PAQUED and Accessible PAQUED schools. However, in the case of mathematics there were no differences between what the Reading Program schools and the PAQUED and Accessible PAQUED schools received. Therefore, they are included in the discussion of Accessible PAQUED and Accessible Control school performance. It is important to note, however, that significance testing was only conducted for differences in performance between (1) Accessible PAQUED and Accessible Control schools and (2) PAQUED and Control schools. While Reading Program schools are included in the same tables and graphs as the Accessible study's schools, the way in which the schools were selected precludes their being appropriately compared to the Accessible Control schools.

This chapter is divided into two sections: (1) presentation of EGMA results in Reading Program, Accessible PAQUED, and Accessible Control schools; (2) presentation of EGMA results disaggregated by sex in Reading Program, Accessible PAQUED, and Accessible Control schools.

Overall, very few significant differences were observed in mathematics performance across the various treatment groups. Furthermore, it was equally common to see control schools outperform treatment schools. The discussion of EGMA results has therefore been substantially abbreviated relative to the discussions provided of EGMA results.

1. EGMA Results in Reading Program, Accessible PAQUED, and Accessible Control Schools

Detailed EGMA results are presented for each of the provinces—Bandundu, Equateur and Orientale. Since the trends in each region are very similar, they are summarized here.

In general, and with the exception of the *Division* subtask in Grade 4, the students in the Reading Program and Accessible PAQUED groups performed slightly better than the students in the Accessible Control group. However, the differences were seldom statistically significant. For all intents and purposes the performance by the students in

the Reading Program, Accessible PAQUED, and Accessible Control groups was quite similar.

The EGMA subtasks can be arranged into two groups: (1) subtasks that assess more procedural knowledge (e.g., *Number Identification* and *Quantity Discrimination*) and subtasks that assess more conceptual knowledge (e.g., *Missing Number*, *Word Problems*, *Addition*, *Subtraction*, *Multiplication* and *Division*). There is a reasonably clear pattern across the groups and grades: students performed better on the more procedural tasks than they did on the more conceptual tasks, and the percentage of students with zero scores was much lower for the more procedural tasks than it was on the more conceptual tasks. This trend was also evident in the 2012 EGMA study that serves as the baseline for the Accessible PAQUED and Accessible Control schools.⁶⁰

While the Grade 2 and Grade 4 EGMA subtasks had some items in common, the majority of the items are grade-appropriate in terms of the curricular expectation of the students for each grade. The only exception to this was the *Missing Number* subtask where all but one of the items were the same for the two grades. In light of this, the generally similar performance across all the subtasks (with the exception of *Missing Number*) by the Grade 2 and Grade 4 students should be interpreted in terms of students being asked grade-level appropriate questions. Rephrased, the performance by Grade 2 and Grade 4 students was comparable if we take into account that the subtask items are grade appropriate. That is, the Grade 4 students performed similarly to the Grade 2 students on items that are grade appropriate—they struggled with the more conceptual items and did better on the more procedural items.

Looking at differences in gains over time across treatment and Control groups provides important insight into the impact of the treatment intervention. Given the way in which Reading Program schools were selected, there was no true Control group against which to compare Reading Program gains, so they are excluded from the comparison and associated discussion. However, it is possible to look at the differences in gains over time (D-in-D) between the Accessible PAQUED and Accessible Control groups to determine the impact of the IAI interventions. The relevant tables in for each region summarize the D-in-D scores from 2012–2014 for these two groups for Grades 2 and 4.

There appeared to be a general trend of some improvement in performance between 2012 and 2014 on many of the EGMA subtasks. The D-in-D analysis, however, shows that the improvement was not statistically significant in favor of the Accessible PAQUED students, suggesting that the apparent improvements (where they exist) cannot be attributed to the intervention activities.

The overriding impression across the regions is that there is no trend of a statistically significant difference in the performance of students in Accessible PAQUED and

⁶⁰ For more details please see Brombacher, A., Davies, C., Ralaingita, W., Slade, T., & Costello, M. (2013). *PAQUED: DRC – EGMA midterm report*. Washington DC: USAID. Available from <https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=542>.

Accessible Control groups in 2014. Furthermore the 2012 and 2014 data for students in the Accessible PAQUED and Accessible Control groups provides no evidence of a statistically significant treatment effect over time.

EGMA Bandundu

Table E1 summarizes the percentages of students in the two grades and three different groups (Reading Program, Accessible PAQUED, and Accessible Control) with zero scores (students unable to respond correctly to a single item on a subtask) and their mean scores at endline (2014) for all EGMA subtasks. With a single exception—Grade 2 zero score % on the *Word Problems* subtask—the differences as they are in the table are not statistically significant at the $p < 0.001$ and $p < 0.002$ levels.⁶¹

Table E1. EGMA Bandundu: 2014 Percent Zero Scores and Mean Percentage Scores by Grade

Subtask (2014)	Grade 2		Grade 4		
	% Zero score	Mean % score	% Zero score	Mean % score	
Number Identification	Reading Program	2%	58%	0%	70%
	Accessible PAQUED	3%	56%	1%	70%
	Accessible Control	1%	48%	0%	57%
Quantity Comparison	Reading Program	1%	67%	1%	69%
	Accessible PAQUED	2%	66%	0%	68%
	Accessible Control	1%	63%	0%	58%
Missing Number	Reading Program	10%	31%	4%	47%
	Accessible PAQUED	14%	25%	3%	43%
	Accessible Control	16%	24%	5%	39%
Word Problems	Reading Program	18%	47%	16%	47%
	Accessible PAQUED	25%	45%	32%	41%

⁶¹ Type 1 errors in statistics occur when a difference is thought to exist where one does not. (Put another way, a type 1 error is the rejection of the null hypothesis when it is actually true.) Due to the large number of comparisons conducted for these sections (*Chapter D*, sections 1-2) of the report, the Bonferroni correction was used to determine the threshold of significance for the various analyses. These sections contain 48 tests of comparison of means for the subtasks administered to all three grades and 24 tests of comparison of means for the subtasks administered to Grades 4 and 6 alone. Applying the Bonferroni correction to these analyses results in $p < 0.05 / (48) = 0.001$ for the Number ID, Quantity Comparison, Missing Number, World Problems, Addition, and Subtraction subtasks and $p < (0.05) / (24) = 0.002$ for the Multiplication and Division subtasks. Please see Annex 1 for a fuller discussion of the Bonferroni correction and the analytical families used in this report.

Subtask (2014)	Grade 2		Grade 4		
	% Zero score	Mean % score	% Zero score	Mean % score	
Addition	Accessible Control	37% *	34%	20%	42%
	Reading Program	19%	59%	12%	52%
	Accessible PAQUED	24%	53%	7%	59%
Subtraction	Accessible Control	28%	45%	22%	40%
	Reading Program	27%	50%	31%	41%
	Accessible PAQUED	42%	42%	25%	52%
Multiplication	Accessible Control	44%	31%	39%	42%
	Reading Program			47%	29%
	Accessible PAQUED			54%	34%
Division	Accessible Control			47%	24%
	Reading Program			38%	26%
	Accessible PAQUED			23%	21%
	Accessible Control			45%	32%

* < 0.001; † < 0.002

Figure E1 displays, by grade, the mean scores for each group on each subtask. There is a clear trend with students performing better on the subtasks assessing procedural knowledge and skills (e.g., Number Identification and Quantity Discrimination) than on the other subtasks. The difference in the performance between Grade 2 and Grade 4 students on the *Missing Number* subtask is attributable to the items being largely the same on both assessments.⁶²

⁶² In EGMA, the design is for subtask items to vary by grade in order to be grade-appropriate. That is, while there would likely be a little overlap in items across the Grade 2, Grade 4, and Grade 6 versions of a subtask, the majority of items would be different. The *Missing Number* subtask is the exception to this rule; it is a subtask where the majority of the items are the same across Grades. As a result, where student performance across grades for most subtasks remains fairly constant in percentage-correct terms because each grade's instrument contains a majority of items that are targeted for that particular grade, in the case of the *Missing Number* subtask performance appears to be increase across grades in percentage-correct terms because the difficulty of the items is largely the same across grades.

Figure E1. EGMA Bandundu: 2014 Mean Percentage Scores by Subtask, Grade, and Treatment Group

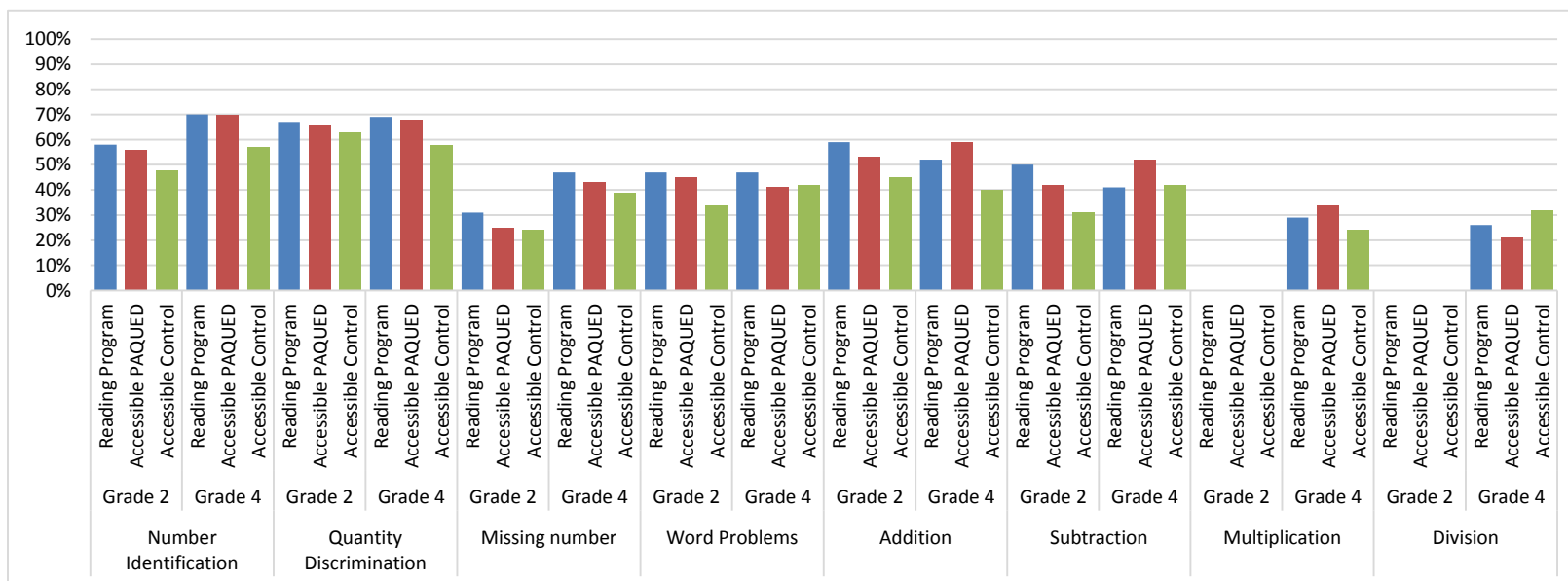


Table E2 summarizes the D-in-D for the Accessible PAQUED and Accessible Control groups between 2012 and 2014 for the two grades. There were no statistically significant differences at the $p < 0.008$ or $p < 0.017$ thresholds. The overarching impression is that there is no strong treatment effect attributable to the intervention.⁶³

Table E2. EGMA Bandundu: 2012 vs. 2014 Difference-in-Differences of Mean Percentage Scores by Grade

Subtask	Grade 2			Grade 4			
	2012	2014	D-in-D ⁶⁴	2012	2014	D-in-D	
Number Identification	Accessible PAQUED	49.40%	56.30%	3.70%	71.20%	70.30%	2.40%
	Accessible Control	45.20%	48.30%		60.10%	56.80%	
Quantity Comparison	Accessible PAQUED	50.20%	66.10%	0.80%	60.50%	67.80%	3.20%
	Accessible Control	47.70%	62.90%		54.30%	58.50%	
Missing Number	Accessible PAQUED	15.70%	24.90%	2.50%	31.30%	42.60%	-1.60%
	Accessible Control	17.10%	23.80%		26.20%	39.20%	
Word Problems	Accessible PAQUED	29.00%	29.80%	6.40%	59.30%	41.00%	-4.40%
	Accessible Control	28.30%	22.70%		56.10%	42.30%	
Addition	Accessible PAQUED	49.00%	52.70%	7.80%	44.60%	58.90%	4.90%
	Accessible Control	49.50%	45.30%		30.90%	40.30%	
Subtraction	Accessible PAQUED	45.00%	42.00%	9.20%	36.50%	51.80%	2.40%
	Accessible Control	43.60%	31.40%		29.20%	42.10%	
Multiplication	Accessible PAQUED				18.00%	33.90%	6.40%
	Accessible Control				14.30%	23.70%	

⁶³ Due to the large number of difference-in-differences comparisons conducted for these sections (**Chapter E**) of the report, the Bonferroni correction was used to determine the threshold of significance for the various analyses. These sections contain 6 tests of difference-in-differences for the subtasks administered to all three grades and 2 tests of difference-in-differences for the subtasks administered to Grades 4 and 6 alone. Applying the Bonferroni correction to these analyses results in $p < 0.05 / (6) = 0.008$ for the Number ID, Quantity Comparison, Missing Number, World Problems, Addition, and Subtraction subtasks and $p < (0.05) / (3) = 0.017$ for the Multiplication and Division subtasks. Please see Annex 1 for a fuller discussion of the Bonferroni correction and the analytical families used in this report.

⁶⁴ Difference-in-differences values are reported according to the value being measured. In the Reading results, those values changed according to the subtask. In the Mathematics results, it is the Mean Percentage Scores which are being reported, so the D-in-D results are presented as percentages.

Subtask	Grade 2			Grade 4		
	2012	2014	D-in-D ⁶⁴	2012	2014	D-in-D
Division	Accessible PAQUED			22.00%	21.50%	-13.50%
	Accessible Control			18.50%	31.50%	

* < 0.008; † < 0.017

EGMA Equateur

Table E3 summarizes the percentages of students in the two grades and three different groups (Reading Program, Accessible PAQUED, and Accessible Control) with zero scores (students unable to respond correctly to a single item on a subtask) and their mean scores at endline (2014) for all EGMA subtasks. The differences presented in the table are not statistically significant at the $p < 0.001$ and $p < 0.002$ levels.

Table E3. EGMA Equateur: 2014 Percent Zero Scores and Mean Percentage Scores by Grade⁶⁵

Subtask (2014)	Grade 2		Grade 4		
	% Zero score	Mean % score	% Zero score	Mean % score	
Number Identification	Reading Program	3%	56%	1%	65%
	Accessible PAQUED	3%	53%	0%	63%
	Accessible Control	7%	50%	1%	60%
Quantity Comparison	Reading Program	1%	68%	2%	65%
	Accessible PAQUED	6%	62%	1%	67%
	Accessible Control	2%	65%	1%	63%
Missing Number	Reading Program	9%	28%	5%	43%
	Accessible PAQUED	15%	28%	4%	46%
	Accessible Control	11%	35%	8%	44%
Word Problems	Reading Program	16%	48%	12%	58%
	Accessible PAQUED	17%	51%	13%	52%
	Accessible Control	22%	44%	9%	50%
Addition	Reading Program	17%	60%	11%	48%

⁶⁵ Significance testing was only conducted for differences between Accessible PAQUED and Accessible Control schools.

Subtask (2014)	Grade 2		Grade 4		
	% Zero score	Mean % score	% Zero score	Mean % score	
Addition	Accessible PAQUED	20%	47%	12%	54%
	Accessible Control	14%	58%	12%	49%
	Reading Program	26%	50%	29%	50%
Subtraction	Accessible PAQUED	33%	41%	34%	46%
	Accessible Control	20%	48%	27%	50%
	Reading Program			41%	39%
Multiplication	Accessible PAQUED			42%	36%
	Accessible Control			37%	44%
	Reading Program			25%	33%
Division	Accessible PAQUED			28%	30%
	Accessible Control			26%	38%

* < 0.001; † < 0.002

Figure E2 displays, by grade, the mean scores for each group on each subtask. There is a clear trend with students performing better on the subtasks assessing procedural knowledge and skills (e.g., Number Identification and Quantity Discrimination) than on the other subtasks. The difference in the performance between Grade 2 and Grade 4 students on the *Missing Number* subtask is attributable to the items being largely the same on both assessments.

Figure E2. EGMA Equateur: 2014 Mean Percentage Scores by Subtask, Grade, and Treatment Group

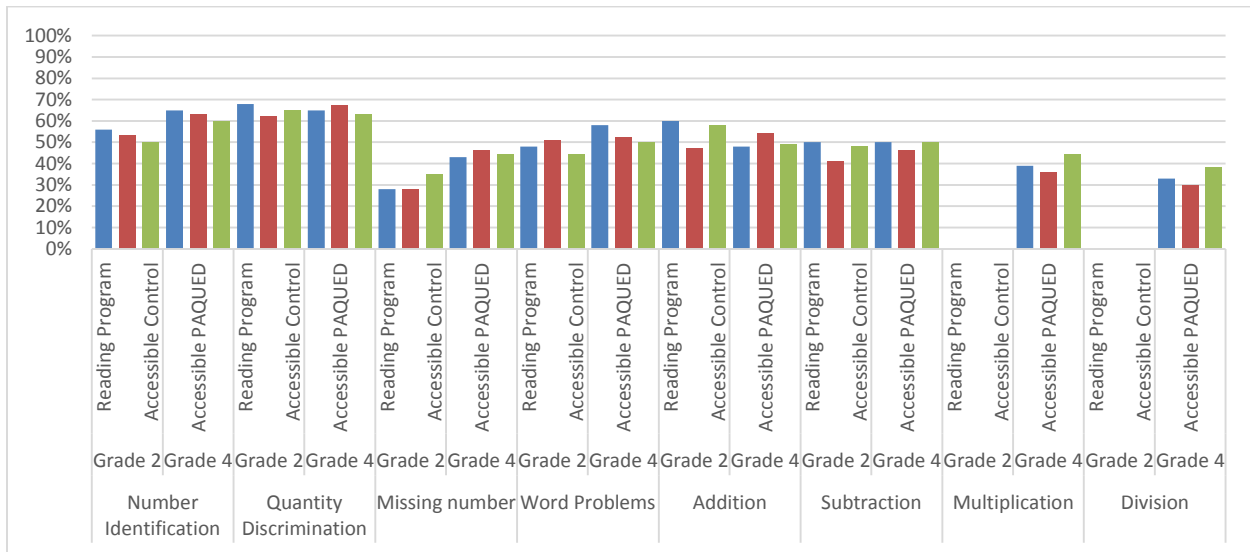


Table E4 summarizes the D-in-D of the Accessible PAQUED and Accessible Control groups between 2012 and 2014 for the two grades.⁶⁶ There were no statistically significant differences at the $p < 0.008$ or $p < 0.017$ thresholds. The overarching impression is that there is no strong treatment effect attributable to the intervention.

Table E4. EGMA Equateur: 2012 vs. 2014 Difference-in-Differences of Mean Percentage Scores by Grade

Subtask		Grade 2			Grade 4		
		2010	2014	D-in-D	2010	2014	D-in-D
Number Identification	Accessible PAQUED	44.60%	53.00%	1.90%	65.70%	62.50%	3.60%
	Accessible Control	43.10%	49.50%		67.00%	60.30%	
Quantity Comparison	Accessible PAQUED	54.50%	62.50%	-0.40%	64.40%	66.80%	-0.30%
	Accessible Control	56.50%	64.90%		60.30%	63.00%	
Missing Number	Accessible PAQUED	18.20%	28.30%	-9.10%	29.40%	45.70%	3.70%
	Accessible Control	15.70%	34.90%		31.60%	44.10%	
Word Problems	Accessible PAQUED	26.10%	33.80%	8.60%	44.50%	51.80%	13.00%

⁶⁶ D-in-D (difference in differences) considers the change over time in performance for two groups of interest—in this case, treatment (Accessible PAQUED) and control (Accessible Control). The change in performance attributable to the intervention (rather than any other environmental factors) is the performance growth in the treatment group net of the performance growth in the control group. D-in-D is reported in the units of the change being measured—in this case, %. A negative D-in-D indicates that performance growth in the control group was greater than performance growth in the treatment group.

Subtask	Grade 2			Grade 4			
	2010	2014	D-in-D	2010	2014	D-in-D	
Addition	Accessible Control	30.30%	29.50%		55.90%	50.20%	
	Accessible PAQUED	42.40%	47.30%	-5.70%	39.50%	54.20%	9.70%
	Accessible Control	47.40%	57.90%		44.10%	49.10%	
Subtraction	Accessible PAQUED	34.80%	40.90%	-6.70%	43.10%	45.90%	4.40%
	Accessible Control	35.10%	47.90%		51.70%	50.10%	
	Accessible PAQUED				22.50%	35.70%	
Multiplication	Accessible Control			26.20%	44.40%	-4.90%	
	Accessible PAQUED			26.30%	30.10%		
Division	Accessible Control			36.20%	38.40%	1.60%	
	Accessible PAQUED						

* < 0.008; † < 0.017

EGMA Orientale

Table E5 summarizes the percentages of students in the two grades and three different groups (Reading Program, Accessible PAQUED, and Accessible Control) with zero scores (students unable to respond correctly to a single item on a subtask) and their mean scores at endline (2014) for all EGMA subtasks. The differences presented in the table are not statistically significant at the $p < 0.001$ and $p < 0.002$ levels.

Table E5. EGMA Orientale: 2014 Percent Zero Scores and Mean Percentage Scores by Grade

Subtask (2014)	Grade 2		Grade 4		
	% Zero score	Mean % score	% Zero score	Mean % score	
Number Identification	Reading Program	4%	48%	0%	68%
	Accessible PAQUED	2%	58%	0%	70%
	Accessible Control	1%	56%	0%	73%
Quantity Comparison	Reading Program	16%	55%	0%	68%
	Accessible PAQUED	0%	72%	1%	77%
	Accessible Control	3%	69%	2%	73%
Missing Number	Reading Program	17%	20%	5%	44%
	Accessible PAQUED	7%	25%	1%	49%

Subtask (2014)	Grade 2		Grade 4		
	% Zero score	Mean % score	% Zero score	Mean % score	
Word Problems	Accessible Control	12%	23%	3%	47%
	Reading Program	63%	14%	28%	27%
	Accessible PAQUED	40%	23%	28%	29%
Addition	Accessible Control	45%	26%	19%	37%
	Reading Program	27%	50%	5%	63%
	Accessible PAQUED	15%	63%	6%	59%
Subtraction	Accessible Control	16%	60%	4%	68%
	Reading Program	44%	33%	13%	64%
	Accessible PAQUED	33%	54%	17%	58%
Multiplication	Accessible Control	41%	41%	10%	72%
	Reading Program			27%	48%
	Accessible PAQUED			30%	44%
Division	Accessible Control			29%	45%
	Reading Program			19%	53%
	Accessible PAQUED			15%	43%
	Accessible Control			31%	55%

* < 0.001; † < 0.002

Figure E3 displays, by grade, the mean scores for each group on each subtask. There is a clear trend with students performing better on the subtasks assessing procedural knowledge and skills (e.g., Number Identification and Quantity Discrimination) than on the other subtasks. The difference in the performance between Grade 2 and Grade 4 students on the *Missing Number* subtask is attributable to the items being largely the same on both assessments.

Figure E3. EGMA Orientale: 2014 Mean Percentage Scores by Subtask, Grade, and Treatment Group

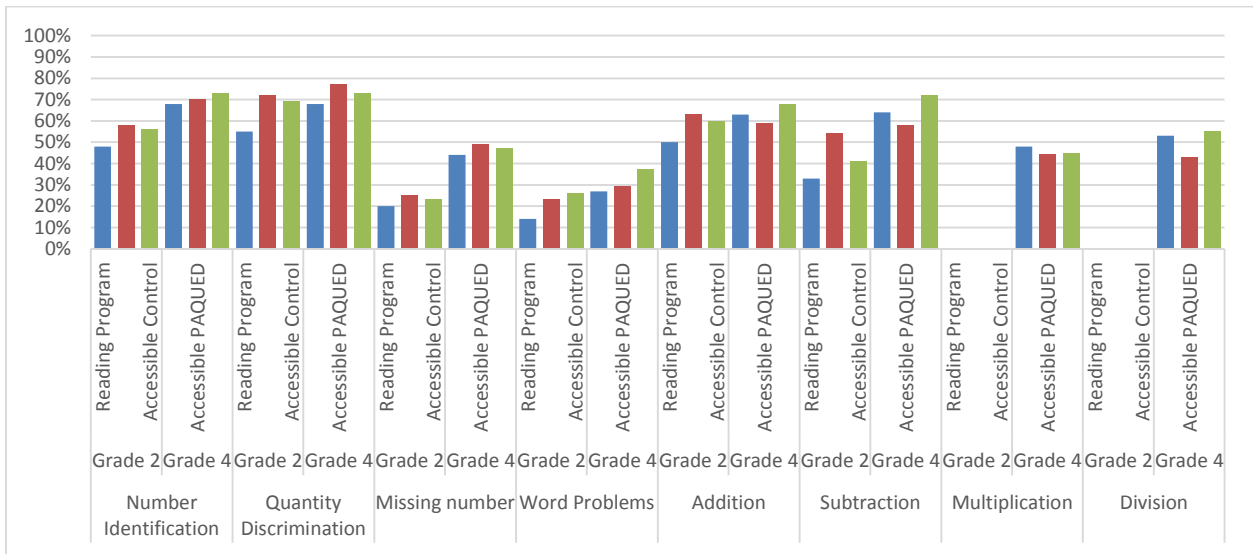


Table E6 summarizes the D-in-D of the Accessible PAQUED and Accessible Control groups between 2012 and 2014 for the two grades. There were no statistically significant differences at the $p < 0.008$ or $p < 0.017$ thresholds. The overriding impression is that there is no strong treatment effect attributable to the intervention. If anything, the treatment group (Accessible PAQUED) appears to have done slightly worse than the Control group.

Table E6. EGMA Orientale: 2012 vs. 2014 Difference-in-Differences of Mean Percentage Scores by Grade

Subtask		Grade 2			Grade 4		
		2010	2014	D-in-D	2010	2014	D-in-D
Number Identification	Accessible PAQUED	57.20%	57.60%	-4.80%	77.50%	70.50%	-2.90%
	Accessible Control	51.30%	56.50%		77.50%	73.30%	
Quantity Comparison	Accessible PAQUED	73.60%	71.70%	-14.7%	75.50%	77.00%	0.10%
	Accessible Control	56.30%	69.10%		71.30%	72.80%	
Missing Number	Accessible PAQUED	20.60%	25.40%	-1.70%	37.70%	49.10%	-2.50%
	Accessible Control	16.10%	22.60%		33.20%	47.10%	
Word Problems	Accessible PAQUED	22.10%	15.60%	-0.10%	48.90%	28.50%	-7.50%
	Accessible Control	23.40%	17.00%		49.60%	36.80%	
Addition	Accessible PAQUED	58.60%	62.50%	-8.80%	53.70%	59.20%	-9.70%
	Accessible Control	47.70%	60.50%		53.30%	68.40%	

Subtask	Grade 2			Grade 4			
	2010	2014	D-in-D	2010	2014	D-in-D	
Subtraction	Accessible PAQUED	35.60%	53.80%	0.00%	50.10%	58.50%	-5.70%
	Accessible Control	22.90%	41.10%		58.40%	72.50%	
Multiplication	Accessible PAQUED				36.60%	43.60%	-2.50%
	Accessible Control				35.40%	45.00%	
Division	Accessible PAQUED				43.70%	43.40%	-12.80%
	Accessible Control				42.50%	55.00%	

* < 0.008; † < 0.017

2. EGMA Results Disaggregated by Sex in Reading Program, Accessible PAQUED, and Accessible Control Schools

While the detailed EGMA results disaggregated by sex are presented by province—Bandundu, Equateur and Orientale—and treatment group (Reading Program and Accessible PAQUED), the trends across the regions and treatment groups are very similar and for that reason are summarized here.

In general, the endline performance of the male and female students was very similar, as were the trends in their performance over time (2012 to 2014). If there is a discernable trend, it is that the male students in Grade 2 appear to consistently perform slightly better than their female counterparts on the more procedural tasks (e.g., Number Identification and Quantity Discrimination). While these differences are apparent in the tables and graphs, they are seldom statistically significant and it is best not to draw too much inference from them.

The reasonably clear pattern with regard to performance on the procedural items versus the more conceptual subtasks described earlier remains evident in the sex-disaggregated data, as does the relative performance of Grade 2 and 4 students (described above).

EGMA Bandundu Accessible PAQUED

Table E7 summarizes the percentages of male and female students in the two grades with zero scores and their mean scores at endline (2014) for all EGMA subtasks. There were no statistically significant differences at the $p < 0.001$ and $p < 0.002$ thresholds.

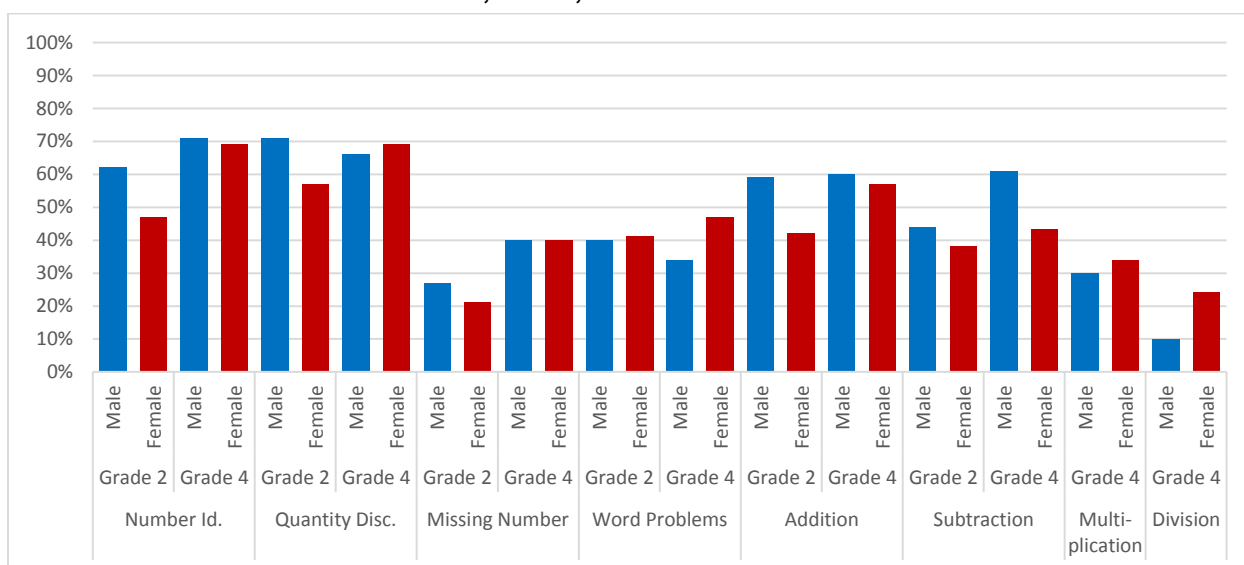
Table E7. EGMA Bandundu Accessible PAQUED Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade

Subtask (2014)		Grade 2		Grade 4	
		% Zero score	Mean % score	% Zero score	Mean % score
Number Identification	Male	2.0%	62.2%	0.7%	71.5%
	Female	3.5%	47.1%	0.5%	69.2%
Quantity Comparison	Male	2.0%	71.7%	0.0%	66.2%
	Female	1.7%	57.4%	0.0%	69.2%
Missing Number	Male	7.6%	27.1%	2.5%	40.1%
	Female	24.2%	21.5%	4.4%	45.0%
Word Problems	Male	24.7%	47.0%	34.2%	34.4%
	Female	26.7%	41.2%	29.7%	47.1%
Addition	Male	21.0%	59.1%	4.4%	60.7%
	Female	29.0%	42.8%	8.5%	57.2%
Subtraction	Male	37.3%	44.4%	22.3%	61.2%
	Female	48.4%	38.2%	26.6%	43.3%
Multiplication	Male			20.1%	33.0%
	Female			26.4%	34.7%
Division	Male			59.9%	18.0%
	Female			48.7%	24.7%

* < 0.001; † < 0.002

Figure E4 displays, by grade, the mean scores for male and female students on each subtask.

Figure E4. EGMA Bandundu Accessible PAQUED Schools: 2014 Mean Percentage Scores by Subtask, Grade, and Sex



EGMA Equateur Accessible PAQUED schools

Table E8 summarizes the percentages of male and female students in the two grades with zero scores and their mean scores at endline (2014) for all EGMA subtasks. There were no statistically significant differences at the $p < 0.001$ and $p < 0.002$ thresholds.

Table E8. EGMA Equateur Accessible PAQUED Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade

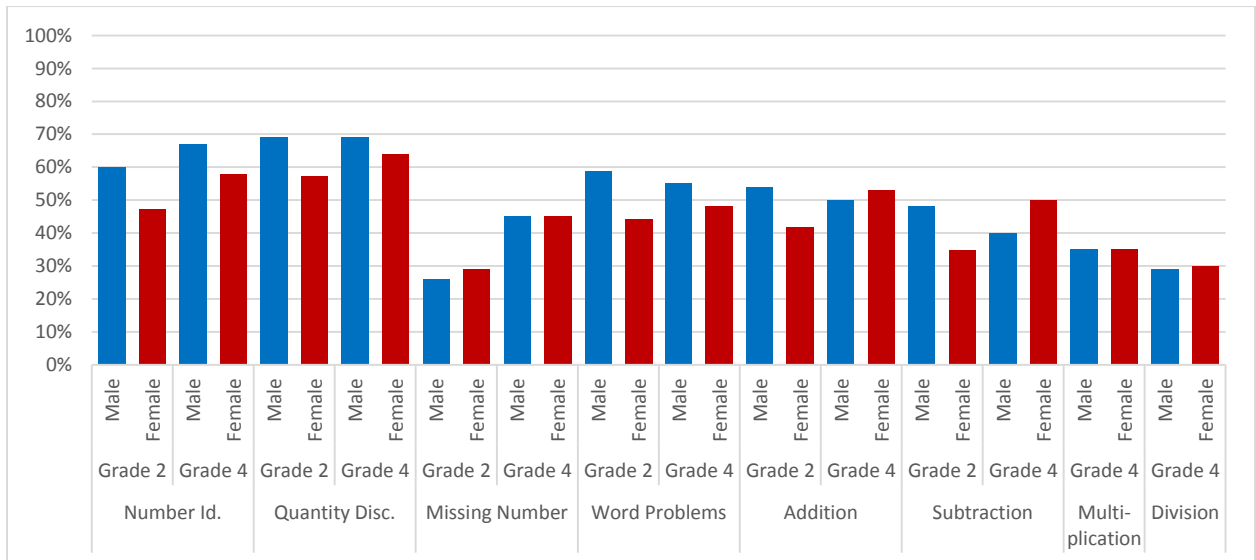
Subtask (2014)		Grade 2		Grade 4	
		% Zero score	Mean % score	% Zero score	Mean % score
Number Identification	Male	1.5%	60.3%	0.0%	67.9%
	Female	3.9%	47.2%	0.0%	57.8%
Quantity Comparison	Male	7.1%	69.1%	1.2%	69.2%
	Female	5.3%	57.2%	0.0%	64.6%
Missing Number	Male	19.2%	26.3%	6.3%	45.8%
	Female	11.9%	29.9%	2.7%	45.6%
Word Problems	Male	10.6%	58.9%	7.0%	55.6%
	Female	22.4%	44.1%	17.7%	48.4%
Addition	Male	18.4%	54.2%	10.6%	55.0%
	Female	22.0%	41.7%	12.8%	53.5%

Subtask (2014)		Grade 2		Grade 4	
		% Zero score	Mean % score	% Zero score	Mean % score
Subtraction	Male	21.2%	48.5%	38.0%	40.6%
	Female	41.6%	34.8%	29.6%	50.5%
Multiplication	Male			31.5%	35.9%
	Female			25.0%	35.6%
Division	Male			50.3%	29.7%
	Female			35.4%	30.5%

* < 0.001; † < 0.002

Figure E5 displays, by grade, the mean scores for male and female students on each subtask.

Figure E5. EGMA Equateur Accessible PAQUED Schools: 2014 Mean Percentage Scores by Subtask, Grade, and Sex



EGMA Orientale Accessible PAQUED schools

Table E9 summarizes the percentages of male and female students in the two grades with zero scores and their mean scores at endline (2014) for all EGMA subtasks. There were no statistically significant differences at the $p < 0.001$ and $p < 0.002$ thresholds.

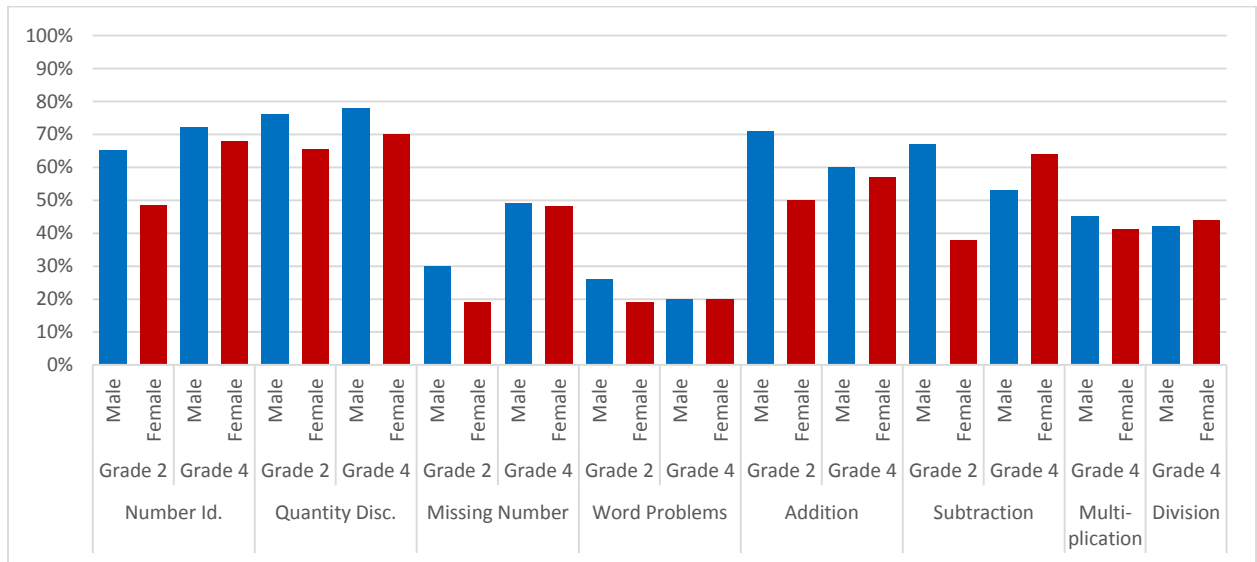
Table E9. EGMA Orientale Accessible PAQUED Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade

Subtask (2014)		Grade 2		Grade 4	
		% Zero score	Mean % score	% Zero score	Mean % score
Number Identification	Male	0.6%	65.2%	0.0%	72.6%
	Female	2.7%	48.6%	0.0%	68.1%
Quantity Comparison	Male	0.1%	76.8%	1.2%	78.7%
	Female	0.0%	65.6%	0.0%	75.0%
Missing Number	Male	5.8%	30.7%	0.5%	49.5%
	Female	9.1%	19.1%	1.5%	48.7%
Word Problems	Male	28.7%	26.4%	22.8%	29.0%
	Female	53.2%	19.9%	33.6%	28.0%
Addition	Male	12.2%	71.3%	2.6%	60.7%
	Female	18.7%	52.0%	10.9%	57.6%
Subtraction	Male	21.0%	67.3%	18.3%	53.3%
	Female	47.0%	37.8%	15.2%	64.4%
Multiplication	Male			15.8%	45.2%
	Female			13.0%	41.9%
Division	Male			32.0%	42.5%
	Female			28.7%	44.4%

* < 0.001; † < 0.002

Figure E6 displays, by grade, the mean scores for male and female students on each subtask.

Figure E6. EGMA Orientale Accessible PAQUED schools: 2014 Mean Percentage Scores by Subtask, Grade, and Sex



EGMA Bandundu Reading Program schools

Table E10 summarizes the percentages of male and female students in the two grades with zero scores and their mean scores at endline (2014) for all EGMA subtasks. There were no statistically significant differences at the $p < 0.001$ and $p < 0.002$ thresholds.

Table E10. EGMA Bandundu Reading Program Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade

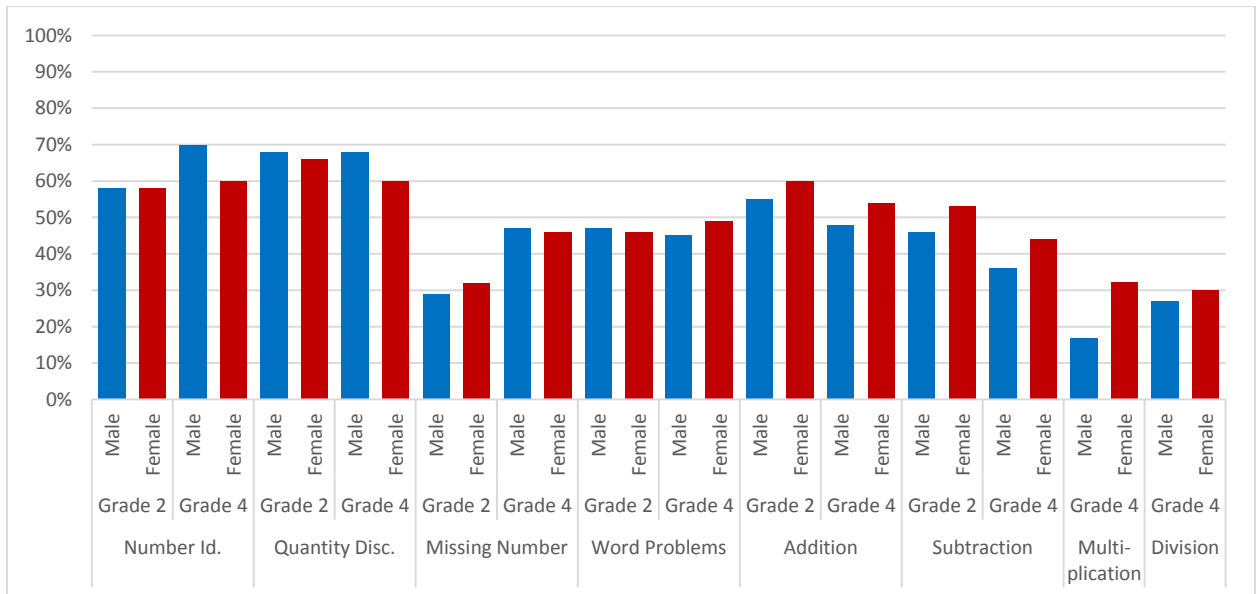
Subtask (2014)		Grade 2		Grade 4	
		% Zero score	Mean % score	% Zero score	Mean % score
Number Identification	Male	2.0%	58.2%	0.0%	70.2%
	Female	1.5%	58.5%	0.0%	69.0%
Quantity Comparison	Male	0.4%	68.1%	2.5%	68.8%
	Female	1.2%	66.8%	0.0%	69.0%
Missing Number	Male	14.2%	29.6%	5.6%	47.2%
	Female	7.5%	32.3%	3.0%	46.7%
Word Problems	Male	17.3%	47.3%	19.0%	45.2%
	Female	18.7%	46.1%	13.6%	49.3%
Addition	Male	20.1%	55.2%	11.3%	48.1%
	Female	17.5%	62.0%	12.5%	54.2%

Subtask (2014)		Grade 2		Grade 4	
		% Zero score	Mean % score	% Zero score	Mean % score
Subtraction	Male	31.2%	46.8%	27.6%	36.2%
	Female	23.2%	53.5%	34.3%	44.5%
Multiplication	Male			52.0%	17.9%
	Female			42.1%	32.1%
Division	Male			37.3%	27.3%
	Female			39.3%	30.6%

* < 0.001; † < 0.002

Figure E7 displays, by grade, the mean scores for male and female students on each subtask.

Figure E7. EGMA Bandundu Reading Program Schools: 2014 Mean Percentage Scores by Subtask, Grade, and Sex



EGMA Equateur Reading Program schools

Table E11 summarizes the percentages of male and female students in the two grades with zero scores and their mean scores at endline (2014) for all EGMA subtasks. There were no statistically significant differences at the $p < 0.001$ and $p < 0.002$ thresholds.

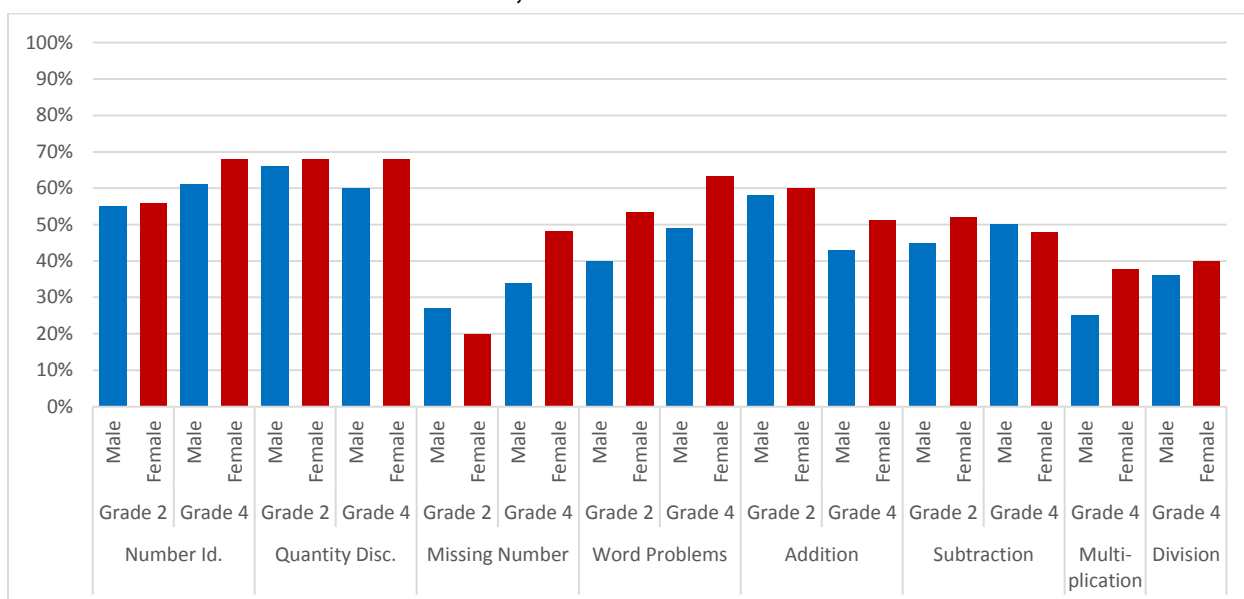
Table E11. EGMA Equateur Reading Program Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade

Subtask (2014)		Grade 2		Grade 4	
		% Zero score	Mean % score	% Zero score	Mean % score
Number Identification	Male	4.1%	55.3%	0.0%	61.2%
	Female	2.8%	56.7%	0.9%	67.9%
Quantity Comparison	Male	0.0%	66.1%	4.1%	60.6%
	Female	0.8%	68.6%	0.5%	68.0%
Missing Number	Male	10.6%	27.8%	6.5%	34.9%
	Female	7.7%	28.0%	4.7%	48.3%
Word Problems	Male	15.8%	40.0%	17.9%	49.3%
	Female	15.6%	53.5%	8.0%	63.2%
Addition	Male	15.4%	58.1%	12.7%	43.4%
	Female	18.6%	61.0%	10.7%	51.2%
Subtraction	Male	28.5%	45.8%	23.1%	51.0%
	Female	23.8%	52.4%	32.4%	48.6%
Multiplication	Male			45.0%	25.7%
	Female			38.4%	37.9% **
Division	Male			23.5%	36.8%
	Female			26.6%	40.2%

* < 0.001; † < 0.002

Figure E8 displays, by grade, the mean scores for male and female students on each subtask.

Figure E8. EGMA Equateur Reading Program Schools: 2014 Mean Percentage Scores by Subtask, Grade, and Sex



EGMA Orientale Reading Program schools

Table E12 summarizes the percentages of male and female students in the two grades with zero scores and their mean scores at endline (2014) for all EGMA subtasks. There were no statistically significant differences at the $p < 0.001$ and $p < 0.002$ thresholds.

Table E12. EGMA Orientale Reading Program Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade

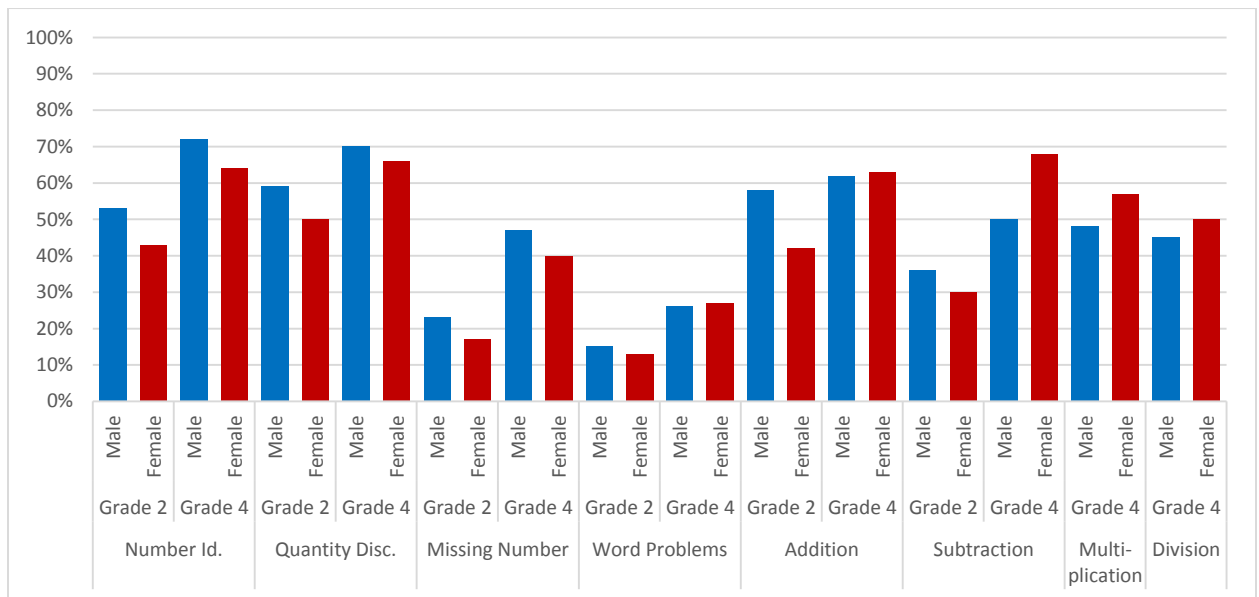
Subtask (2014)		Grade 2		Grade 4	
		% Zero score	Mean % score	% Zero score	Mean % score
Number Identification	Male	3.7%	53.8%	0.0%	72.2%
	Female	3.8%	43.2%	0.0%	64.4%
Quantity Comparison	Male	13.4%	59.6%	0.0%	71.0%
	Female	18.5%	50.5%	0.8%	66.1%
Missing Number	Male	9.5%	23.1%	2.2%	47.9%
	Female	23.3%	17.9%	7.0%	40.1%
Word Problems	Male	57.7%	15.7%	26.4%	26.3%
	Female	68.5%	13.1%	29.8%	27.1%
Addition	Male	20.1%	58.5%	1.9%	62.6%

Subtask (2014)		Grade 2		Grade 4	
		% Zero score	Mean % score	% Zero score	Mean % score
Subtraction	Female	32.7%	42.1%	8.2%	63.3%
	Male	42.7%	36.9%	14.6%	59.0%
Multiplication	Female	46.0%	30.1%	12.3%	68.3%
	Male			33.0%	48.2%
Division	Female			22.2%	57.5%
	Male			23.2%	45.1%
	Female			16.1%	50.5%

* < 0.001; † < 0.002

Figure E9 displays, by grade, the mean scores for male and female students on each subtask.

Figure E9. EGMA Orientale Reading Program Schools: 2014 Mean Percentage Scores by Subtask, Grade, and Sex



Chapter F: Results and Analysis of Student Mathematics Performance in PAQUED and Control Schools

Chapter F presents EGMA results for the PAQUED and Control populations. (*Chapter E* presents EGMA results for the Accessible PAQUED, Accessible Control, and Reading Program populations. EGMA results for the Reading Program schools were presented in *Chapter B* because there was a substantive difference between the reading-focused intervention they received and the intervention that was provided to the PAQUED and Accessible PAQUED schools. However, in the case of mathematics there were no differences between what the Reading Program schools and the PAQUED and Accessible PAQUED schools received. Therefore, they are included in the discussion of Accessible PAQUED and Accessible Control school performance.)

This chapter is divided into two sections: (1) presentation of EGMA results in PAQUED and Control schools; (2) presentation of EGMA results disaggregated by sex in PAQUED and Control schools.

1. EGMA Results in PAQUED and Control Schools

While the detailed EGMA results are presented by province—Bandundu, Equateur, and Orientale—the trends in each province are very similar and for that reason are summarized here.

In general, the performance of the PAQUED and Control students were very similar, as measured by mean scores and trends in performance over time. Where there are differences, they are seldom statistically significant and there is no trend showing consistent differences in favor of either the PAQUED or Control group.

As was noted earlier during the discussion of results in Accessible PAQUED and Accessible Control schools, students in the PAQUED and Control study tend to perform better on procedural tasks than on conceptual tasks. This trend persists across treatment groups and grades, and was also evident in the 2010 baseline study.⁶⁷

Also as noted during the discussion of results in Accessible PAQUED and Accessible Control schools, the majority of EGMA subtask items are grade-appropriate in terms of the curricular expectation of the students for each grade. In light of this, the generally similar performance across all the subtasks (with the exception of Missing Number), by the Grade 2, Grade 4, and Grade 6 students should be interpreted in terms of students being asked grade-level appropriate questions.

⁶⁷ For more details please see Brombacher et al. (2012). *PAQUED: DRC Baseline Report, Early Grade Mathematics Assessment*. Reported by RTI International, accessible at <https://www.eddataglobal.org/countries/index.cfm?fuseaction=pubDetail&ID=411>.

Looking at differences in gains over time across treatment and control groups provides important insights into the impact of a treatment intervention. The relevant tables for each region summarize the D-in-D scores from 2010–2014 for the PAQUED and Control groups for Grades 2, 4, and 6.

Looking back to the case of the Accessible PAQUED schools, there appeared to be a general trend of some improvement in performance between 2012 and 2014 on many of the EGMA subtasks; this general improvement is not evident for the broader PAQUED sample. Although there appears to be improvement on some tasks, on others, performance appears to have decreased. Regardless, the D-in-D analysis shows that any improvements that may exist are not statistically significant in favor of the PAQUED students.

The overriding impression across the regions is that there is no trend of a statistically significant difference in the performance of the PAQUED and Control groups in 2014. Furthermore, the 2010 and 2014 data for students in the PAQUED and Control groups provides no evidence of a statistically significant treatment effect over time.

EGMA Bandundu

Table F1 summarizes the percentages of students in all three grades and both treatment groups (Control and PAQUED) with zero scores (i.e., students unable to respond correctly to a single item on a subtask) and their mean scores at the 2014 endline for all EGMA subtasks in Bandundu. The differences in the table are not statistically significant at the $p < 0.001$ and $p < 0.002$ thresholds.⁶⁸

Table F1. EGMA Bandundu: 2014 Percent Zero Scores and Mean Percentage Scores by Grade.

Subtask (2014)		Grade 2		Grade 4		Grade 6	
		% Zero score	Mean % score	% Zero score	Mean % score	% Zero score	Mean % score
Number Identification	Control	1.4%	48.3%	0.0%	56.8%	0.0%	64.9%
	PAQUED	2.6%	56.3%	0.6%	70.3%	0.6%	66.0%
Quantity Comparison	Control	0.7%	62.9%	0.4%	58.5%	5.1%	39.7%
	PAQUED	1.9%	66.1%	0.0%	67.8%	4.1%	38.5%
Missing Number	Control	15.6%	23.8%	5.1%	39.2%	4.1%	42.6%

⁶⁸ Due to the large number of comparisons conducted for these sections (Chapter D, sections 3-4) of the report, the Bonferroni correction was used to determine the threshold of significance for the various analyses. These sections contain 36 tests of comparison of means for the subtasks administered to all three grades and 24 tests of comparison of means for the subtasks administered to Grades 4 and 6 alone. Applying the Bonferroni correction to these analyses results in $p < 0.05 / (36) = 0.001$ for the Number ID, Quantity Comparison, Missing Number, World Problems, Addition, and Subtraction subtasks and $p < (0.05) / (24) = 0.002$ for the Multiplication and Division subtasks. Please see Annex 1 for a fuller discussion of the Bonferroni correction and the analytical families used in this report.

Subtask (2014)		Grade 2		Grade 4		Grade 6	
		% Zero score	Mean % score	% Zero score	Mean % score	% Zero score	Mean % score
Word Problems	PAQUED	14.1%	24.9%	3.5%	42.6%	3.4%	38.2%
	Control	37.1%	34.0%	20.1%	42.3%	8.2%	55.7%
Addition	PAQUED	25.5%	44.7%	31.9%	41.0%	12.6%	51.8%
	Control	28.5%	45.3%	22.0%	40.3%	5.3%	60.0%
Subtraction	PAQUED	24.1%	52.7%	6.5%	58.9%	10.8%	56.9%
	Control	44.2%	31.4%	38.9%	42.1%	27.5%	45.5%
Multiplication	PAQUED	41.7%	42.0%	24.5%	51.8%	30.0%	42.1%
	Control			44.6%	23.7%	44.5%	27.6%
Division	PAQUED			23.4%	33.9%	52.3%	19.2%
	Control			46.6%	31.5%	49.3%	28.4%
	PAQUED			54.0%	21.5%	52.2%	21.7%

* < 0.001; † < 0.002

Figure F1 displays, by grade, the mean scores for each group on each subtask. There is a clear trend with students performing better on the subtasks assessing procedural knowledge and skills (e.g., *Number Identification* and *Quantity Discrimination*) than on the other subtasks. As alluded to in the introduction to **Chapter E**, the difference in the performance between Grade 2, Grade 4, and Grade 6 students on the *Missing Number* subtask is attributable to the items being largely the same across the assessments.

Figure F1. EGMA Bandundu: 2014 Mean Percentage Scores by Subtask, Grade, and Treatment Group

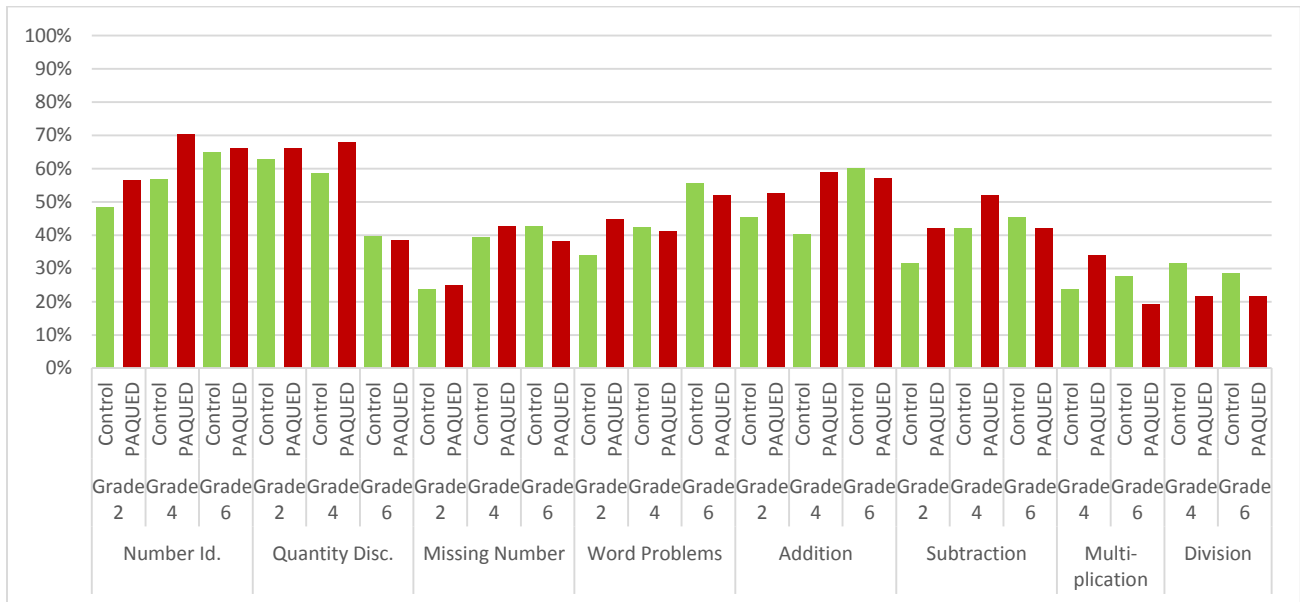


Table F2 summarizes the D-in-D of the Control and PAQUED groups between 2010 and 2014 for the three grades. The differences in the table are not statistically significant at the $p < 0.006$ and $p < 0.008$ thresholds.⁶⁹ The overarching impression is that there is no strong treatment effect attributable to the intervention.

Table F2. EGMA Bandundu: 2010 vs. 2014 Difference-in-Differences of Mean Percentage Scores by Grade

Subtask		Grade 2			Grade 4			Grade 6		
		2010	2014	D-in-D	2010	2014	D-in-D	2010	2014	D-in-D
Number Identification	Control	33.6%	43.4%	0.4%	52.9%	49.3%	3.8%	59.6%	64.9%	7.3%
	PAQUED	30.4%	40.6%		55.3%	55.5%		53.3%	66.0%	
Quantity Comparison	Control	61.2%	45.7%	4.1%	62.7%	54.4%	4.9%	35.5%	39.7%	-0.3%
	PAQUED	57.1%	45.7%		60.3%	57.0%		34.7%	38.5%	
Missing Number	Control	17.9%	16.8%	-0.5%	23.7%	35.5%	-8.7%	22.3%	42.6%	-5.0%

⁶⁹ Due to the large number of difference-in-differences comparisons conducted for these sections (*Chapter F*) of the report, the Bonferroni correction was used to determine the threshold of significance for the various analyses. These sections contain 9 tests of difference-in-differences for the subtasks administered to all three grades and 6 tests of difference-in-differences for the subtasks administered to Grades 4 and 6 alone. Applying the Bonferroni correction to these analyses results in $p < 0.05 / (9) = 0.006$ for the Number ID, Quantity Comparison, Missing Number, World Problems, Addition, and Subtraction subtasks and $p < (0.05) / (6) = 0.008$ for the Multiplication and Division subtasks. Please see Annex 1 for a fuller discussion of the Bonferroni correction and the analytical families used in this report.

Subtask	Grade 2			Grade 4			Grade 6			
	2010	2014	D-in-D	2010	2014	D-in-D	2010	2014	D-in-D	
Word Problems	PAQUED	17.8%	16.2%		30.0%	33.1%		22.9%	38.2%	
	Control	44.8%	30.9%	7.5%	50.3%	41.5%	-6.6%	62.6%	55.7%	-2.0%
	PAQUED	37.6%	31.2%		51.0%	35.6%		60.8%	51.8%	
Addition	Control	40.0%	37.8%	0.4%	28.0%	34.1%	1.0%	47.2%	60.0%	6.3%
	PAQUED	33.9%	32.1%		33.8%	40.8%		37.8%	56.9%	
Subtraction	Control	38.5%	26.1%	5.6%	37.4%	25.2%	6.0%	33.4%	45.5%	-7.2%
	PAQUED	32.3%	25.4%		42.3%	36.1%		37.3%	42.1%	
Multiplication	Control				21.1%	13.5%	-0.1%	16.3%	27.6%	-8.7%
	PAQUED				24.8%	17.1%		16.6%	19.2%	
Division	Control				21.5%	7.5%	-0.4%	21.2%	28.4%	-5.4%
	PAQUED				28.7%	14.2%		19.9%	21.7%	

* < 0.006; † < 0.008

EGMA Equateur

Table F3 summarizes the percentages of students in all three grades and both treatment groups (Control and PAQUED) with zero scores (i.e., students unable to respond correctly to a single item on a subtask) and their mean scores at the 2014 endline for all EGMA subtasks in Equateur. The differences in the table are not statistically significant at the $p < 0.001$ and $p < 0.002$ thresholds.

Table F3. EGMA Equateur: 2014 Percent Zero Scores and Mean Percentage Scores by Grade

Subtask (2014)		Grade 2		Grade 4		Grade 6	
		% Zero score	Mean % score	% Zero score	Mean % score	% Zero score	Mean % score
Number Identification	Control	7.0%	49.5%	0.7%	60.3%	0.0%	67.7%
	PAQUED	2.9%	53.0%	0.0%	62.5%	0.0%	64.9%
Quantity Comparison	Control	1.8%	64.9%	0.8%	63.0%	2.1%	46.1%
	PAQUED	6.0%	62.4%	0.6%	66.8%	3.5%	44.0%
Missing Number	Control	10.8%	34.9%	8.2%	44.1%	5.1%	43.1%

	PAQUED	15.1%	28.3%	4.4%	45.7%	5.3%	40.3%
Word Problems	Control	21.7%	44.3%	8.6%	50.2%	2.5%	59.6%
	PAQUED	17.2%	50.7%	12.7%	51.8%	7.5%	59.1%
Addition	Control	14.1%	57.9%	11.9%	49.1%	4.4%	62.8%
	PAQUED	20.4%	47.2%	11.8%	54.2%	7.0%	57.2%
Subtraction	Control	19.7%	47.9%	26.7%	50.1%	14.6%	53.3%
	PAQUED	32.5%	40.9%	33.6%	45.9%	20.1%	48.8%
Multiplication	Control			25.9%	44.4%	31.9%	32.6%
	PAQUED			28.0%	35.7%	32.1%	31.9%
Division	Control			37.0%	38.4%	24.1%	46.9%
	PAQUED			42.4%	30.1%	30.4%	36.9%

* < 0.001; † < 0.002

Figure F2 displays, by grade, the mean scores for each group on each subtask. There is a clear trend with students performing better on the subtasks assessing procedural knowledge and skills (e.g., *Number Identification* and *Quantity Discrimination*) than on the other subtasks. The difference in the performance between Grade 2, Grade 4, and Grade 6 students on the *Missing Number* subtask is attributable to the items being largely the same across the assessments.

Figure F2. EGMA Equateur: 2014 Mean Percentage Scores by Subtask, Grade, and Treatment Group

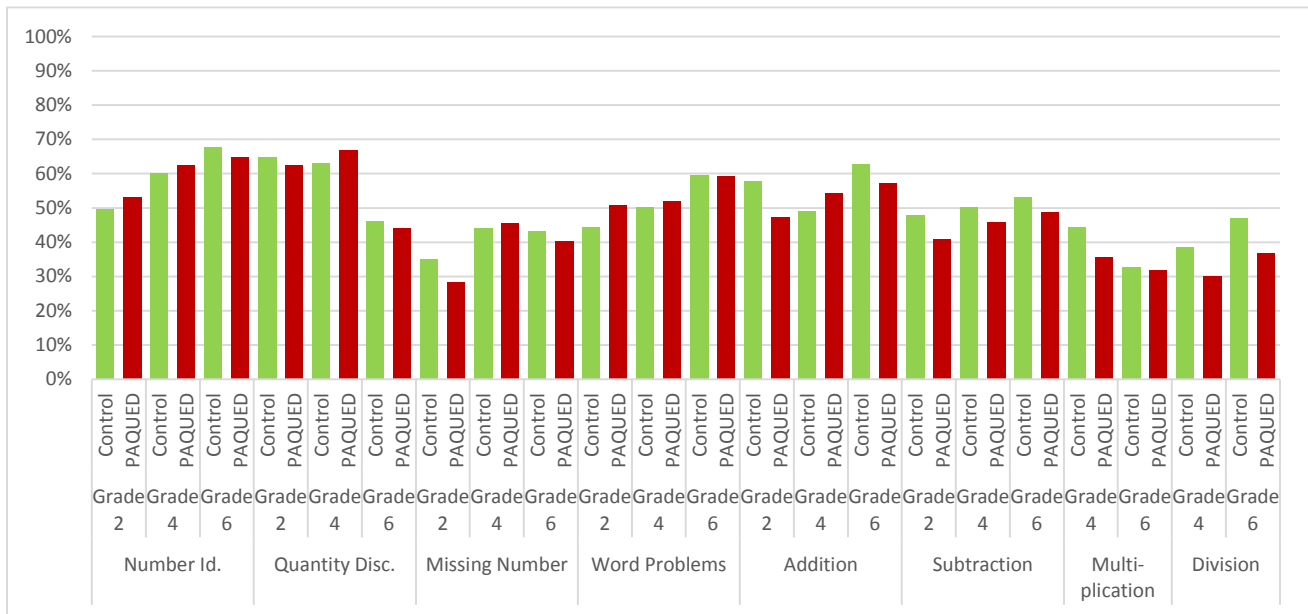


Table F4 summarizes the D-in-D for the Control and PAQUED groups between 2010 and 2014 for all three grades. The differences in the table are not statistically significant at the $p < 0.006$ and $p < 0.008$ thresholds. The overarching impression is that there is no strong treatment effect attributable to the intervention.

Table F4. EGMA Equateur: 2010 vs. 2014 Difference-in-Differences of Mean Percentage Scores by Grade

Subtask		Grade 2			Grade 4			Grade 6		
		2010	2014	D-in-D	2010	2014	D-in-D	2010	2014	D-in-D
Number Identification	Control	29.3%	36.2%	0.9%	56.6%	62.2%	-9.8%	59.5%	67.7%	-3.1%
	PAQUED	38.0%	45.8%		61.3%	57.2%		59.8%	64.9%	
Quantity Comparison	Control	55.0%	49.5%	8.2%	61.2%	63.0%	-4.6%	34.1%	46.1%	-1.6%
	PAQUED	60.0%	62.7%		64.6%	61.9%		33.6%	44.0%	
Missing Number	Control	8.8%	18.2%	3.0%	15.0%	37.1%	-2.5%	13.8%	43.1%	-1.1%
	PAQUED	11.9%	24.3%		18.9%	38.5%		12.1%	40.3%	
Word Problems	Control	29.4%	30.5%	7.2%	38.4%	49.8%	-5.9%	56.4%	59.6%	-0.5%
	PAQUED	34.7%	43.0%		44.6%	50.0%		56.5%	59.1%	

Subtask		Grade 2			Grade 4			Grade 6		
		2010	2014	D-in-D	2010	2014	D-in-D	2010	2014	D-in-D
Addition	Control	32.5%	35.4%	8.2%	31.2%	53.3%	-10.7%	45.5%	62.8%	-0.2%
	PAQUED	37.9%	49.0%		32.9%	44.4%		40.1%	57.2%	
Subtraction	Control	23.8%	24.2%	12.2%	39.1%	48.6%	-13.9%	43.1%	53.3%	-7.6%
	PAQUED	28.8%	41.5%		45.8%	41.5%		46.2%	48.8%	
Multiplication	Control				22.0%	30.3%	4.8%	22.1%	32.6%	-2.1%
	PAQUED				20.9%	34.0%		23.6%	31.9%	
Division	Control				14.9%	32.9%	-3.5%	18.4%	46.9%	-11.0%
	PAQUED				20.4%	35.0%		19.4%	36.9%	

* < 0.006; † < 0.008

EGMA Orientale

Table F5 summarizes the percentages of students in all three grades and both treatment groups (Control and PAQUED) with zero scores (i.e., students unable to respond correctly to a single item on a subtask) and their mean scores at the 2014 endline for all EGMA subtasks in Orientale. The differences in the table are not statistically significant at the $p < 0.001$ and $p < 0.002$ thresholds.

Table F5. EGMA Orientale: 2014 Percent Zero Scores and Mean Percentage Scores by Grade

Subtask (2014)		Grade 2		Grade 4		Grade 6	
		% Zero score	Mean % score	% Zero score	Mean % score	% Zero score	Mean % score
Number Identification	Control	1.4%	56.5%	0.0%	73.3%	0.0%	70.1%
	PAQUED	1.6%	57.6%	0.0%	70.5%	0.4%	72.2%
Quantity Comparison	Control	3.3%	69.1%	2.0%	72.8%	0.0%	51.1%
	PAQUED	0.1%	71.7%	0.6%	77.0%	0.0%	48.8%
Missing Number	Control	11.6%	22.6%	2.9%	47.1%	0.0%	46.2%
	PAQUED	7.3%	25.4%	1.0%	49.1%	1.5%	47.5%
Word Problems	Control	45.3%	25.5%	18.8%	36.8%	14.7%	46.4%
	PAQUED	39.9%	23.4%	27.9%	28.5%	6.6%	49.3%
Addition	Control	16.3%	60.4%	4.3%	68.4%	1.9%	73.6%

Subtask (2014)	Grade 2		Grade 4		Grade 6		
	% Zero score	Mean % score	% Zero score	Mean % score	% Zero score	Mean % score	
Subtraction	PAQUED	15.2%	62.5%	6.5%	59.2%	1.3%	76.5%
	Control	40.7%	41.1%	9.6%	72.5%	8.0%	64.3%
Multiplication	PAQUED	32.9%	53.8%	16.9%	58.5%	7.2%	69.5%
	Control			31.0%	45.0%	20.7%	41.0%
Division	PAQUED			14.5%	43.7%	16.6%	46.4%
	Control			28.8%	55.0%	14.4%	56.1%
	PAQUED			30.5%	43.4%	13.3%	53.4%

* < 0.001; † < 0.002

Figure F3 displays, by grade, the mean scores for each group on each subtask. There is a clear trend with students performing better on the subtasks assessing procedural knowledge and skills (e.g., *Number Identification* and *Quantity Discrimination*) than on the other subtasks. The difference in the performance between Grade 2, Grade 4, and Grade 6 students on the *Missing Number* subtask is attributable to the items being largely the same across the assessments.

Figure F3. EGMA Orientale: 2014 Mean Percentage Scores by Subtask, Grade, and Treatment Group

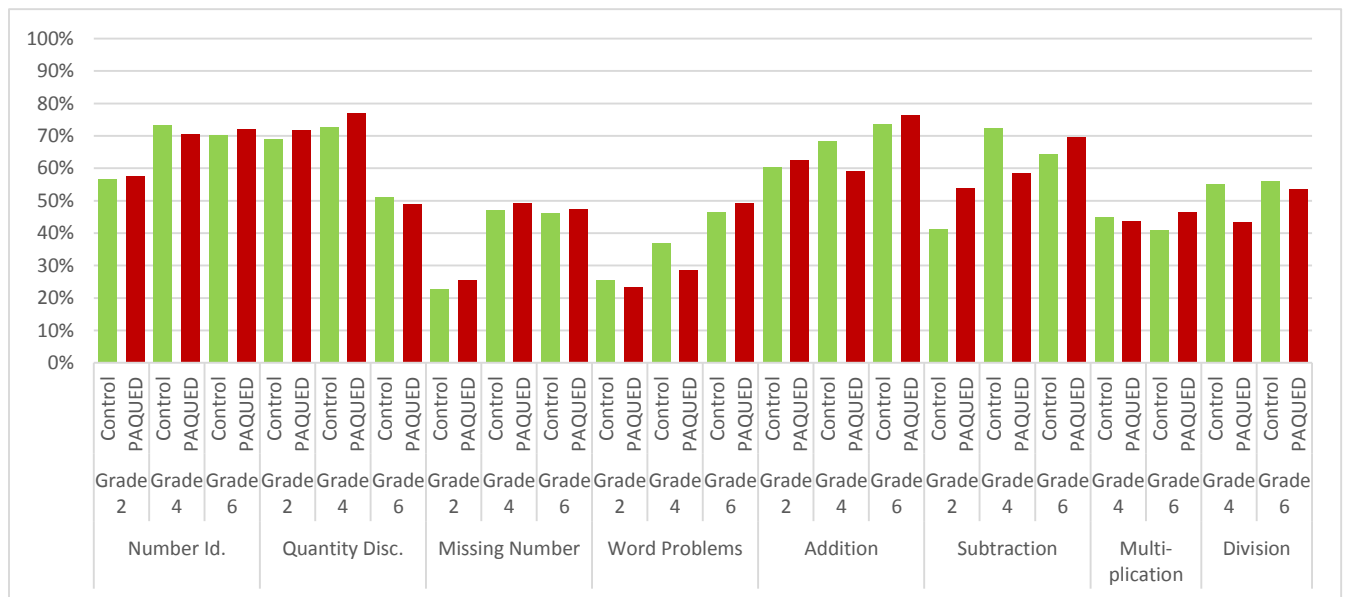


Table F6 summarizes the D-in-D of the Control and PAQUED groups between 2010 and 2014 for the three grades. The differences in the table are not statistically significant at the $p < 0.006$ and $p < 0.008$ thresholds. The overarching impression is that there is no strong treatment effect attributable to the intervention. In fact, the PAQUED group appears to have done slightly worse than the Control group.

Table F6. EGMA Orientale: 2010 vs. 2014 Difference-in-Differences of Mean Percentage Scores by Grade

Subtask		Grade 2			Grade 4			Grade 6		
		2010	2014	D-in-D	2010	2014	D-in-D	2010	2014	D-in-D
Number Identification	Control	37.0%	47.6%	-18.2%	62.1%	66.3%	-8.6%	69.2%	70.1%	-0.4%
	PAQUED	48.7%	41.2%		67.8%	63.4%		71.7%	72.2%	
Quantity Comparison	Control	52.9%	64.6%	-8.0%	66.7%	70.9%	-3.6%	43.2%	51.1%	-6.4%
	PAQUED	64.0%	67.7%		71.4%	72.0%		47.4%	48.8%	
Missing Number	Control	15.9%	20.5%	-3.9%	26.3%	44.3%	-8.8%	25.8%	46.2%	-11.4%
	PAQUED	19.1%	19.8%		38.4%	47.6%		38.6%	47.5%	
Word Problems	Control	29.8%	24.9%	-8.4%	39.3%	33.8%	-5.8%	68.4%	46.4%	0.2%
	PAQUED	40.4%	27.2%		47.5%	36.1%		71.1%	49.3%	
Addition	Control	33.7%	46.8%	-16.8%	34.9%	56.1%	-18.8%	57.6%	73.6%	-1.4%
	PAQUED	50.5%	46.7%		47.7%	50.2%		61.9%	76.5%	
Subtraction	Control	26.8%	32.3%	-15.3%	41.7%	53.5%	-17.3%	45.9%	64.3%	-12.5%
	PAQUED	44.2%	34.4%		54.7%	49.1%		63.6%	69.5%	
Multiplication	Control				17.3%	41.8%	-23.0%	24.7%	41.0%	-12.6%
	PAQUED				29.3%	30.8%		42.7%	46.4%	
Division	Control				17.2%	39.1%	-25.3%	25.0%	56.1%	-12.8%
	PAQUED				31.4%	28.0%		35.1%	53.4%	

* < 0.006 ; † < 0.008

2. EGMA Results Disaggregated by Sex in PAQUED and Control Schools

While the detailed EGMA results disaggregated by sex are presented for each province—Bandundu, Equateur, and Orientale—the trends in each region are very similar and, therefore, are summarized here.

In general, the endline (2014) performance of the male and female students, and the trends in their performance over time (2012 to 2014), were fairly similar. If there is a discernable trend, it is that the male students in the earlier grades (Grade 2 and Grade 4) appear to consistently perform slightly better than their female counterparts on the more procedural subtasks (e.g., *Number Identification* and *Quantity Discrimination*). While these differences are apparent in the tables and graphs, they are seldom statistically significant and it is best not to draw too much inference from them.

The reasonably clear pattern with regard to performance on the procedural items versus the more conceptual subtasks described earlier remains evident in the data disaggregated by sex, as does the relative performance of Grade 2, 4, and 6 students, also described above.

EGMA Bandundu PAQUED Schools

Table F7 summarizes the percentages of male and female students in the three grades with zero scores and their mean scores at endline (2014) for all EGMA subtasks. The differences in the table are not statistically significant at the $p < 0.001$ and $p < 0.002$ thresholds.

Table F7. EGMA Bandundu PAQUED Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade

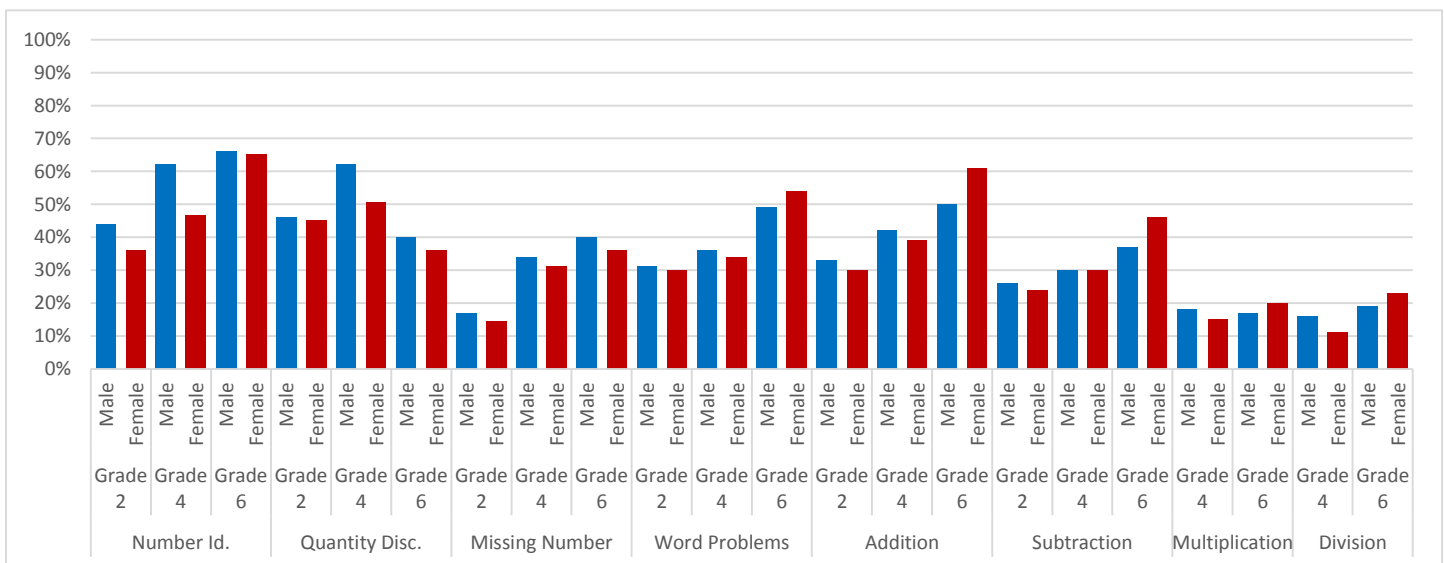
Subtask		Grade 2		Grade 4		Grade 6	
		% Zero score	Mean % score	% Zero score	Mean % score	% Zero score	Mean % score
Number Identification	Male	2.9%	44.2%	0.0%	62.6%	1.2%	66.7%
	Female	7.5%	36.8%	1.7%	46.6%	0.0%	65.2%
Quantity Comparison	Male	8.6%	46.1%	1.1%	62.1%	3.6%	40.6%
	Female	4.6%	45.3%	1.0%	50.5%	4.7%	36.4%
Missing Number	Male	18.0%	17.9%	5.1%	34.2%	2.0%	40.2%
	Female	26.1%	14.5%	7.6%	31.7%	4.8%	36.2%
Word Problems	Male	35.9%	31.9%	14.4%	36.3%	15.6%	49.6%

Subtask		Grade 2		Grade 4		Grade 6	
		% Zero score	Mean % score	% Zero score	Mean % score	% Zero score	Mean % score
Addition	Female	41.8%	30.5%	26.5%	34.8%	9.6%	54.1%
	Male	36.9%	33.5%	12.3%	42.1%	8.6%	52.0%
Subtraction	Female	50.7%	30.6%	25.4%	39.2%	13.1%	61.9%
	Male	47.7%	26.1%	41.3%	37.0%	31.4%	37.4%
Multiplication	Male			60.9%	18.3%	58.5%	17.8%
	Female			63.6%	15.5%	46.1%	20.6%
Division	Male			69.2%	16.3%	56.0%	19.6%
	Female			75.9%	11.5%	48.4%	23.9%

* < 0.001; † < 0.002

Figure F4 displays, by grade, the mean scores for male and female students on each subtask.

Figure F4. EGMA Bandundu Treatment: 2014 Mean Percentage Scores by Subtask, Grade, and Sex



EGMA Equateur PAQUED Schools

Table F8 summarizes the percentages of male and female students in the three grades with zero scores and their mean scores at endline (2014) for all EGMA subtasks. The differences in the table are not statistically significant at the $p < 0.001$ and $p < 0.002$ thresholds.

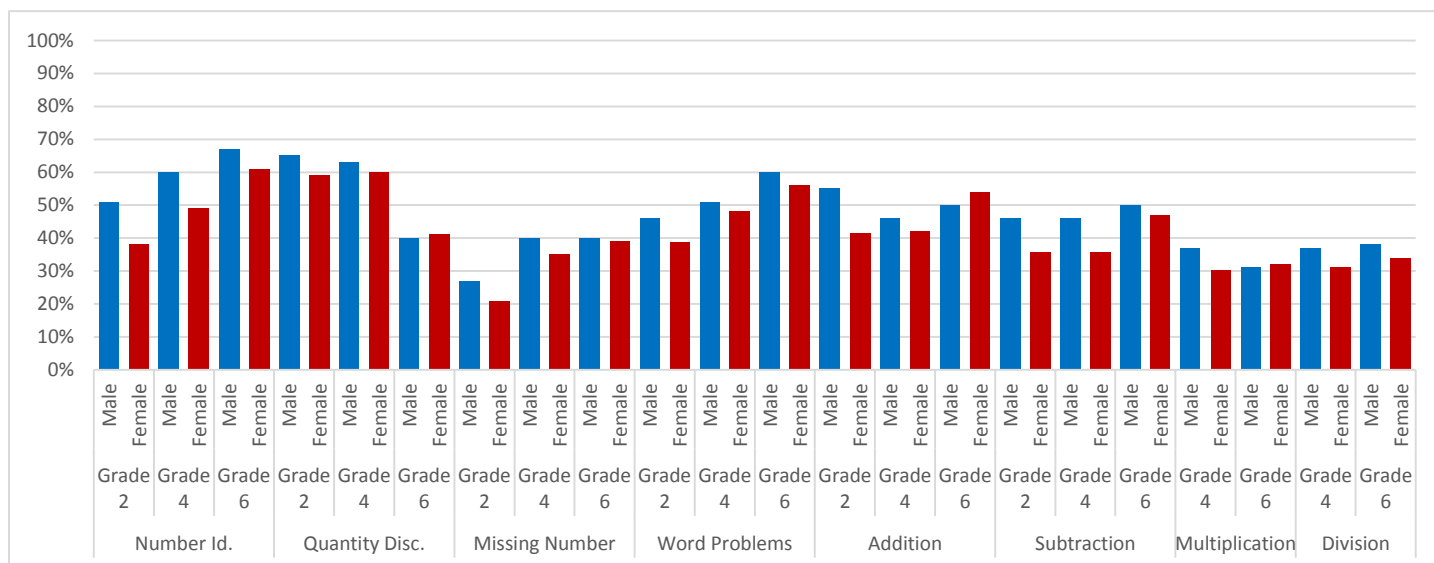
Table F8. EGMA Equateur PAQUED Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade

Subtask		Grade 2		Grade 4		Grade 6	
		% Zero score	Mean % score	% Zero score	Mean % score	% Zero score	Mean % score
Number Identification	Male	3.6%	51.7%	2.1%	65.0%	0.0%	67.8%
	Female	5.9%	38.7%	4.4%	49.1%	0.0%	60.9%
Quantity Comparison	Male	3.3%	65.3%	1.6%	63.1%	3.4%	46.0%
	Female	2.8%	59.5%	3.7%	60.6%	3.7%	41.2%
Missing Number	Male	11.9%	27.1%	6.4%	41.0%	4.6%	40.8%
	Female	20.5%	20.9%	9.6%	35.8%	6.2%	39.6%
Word Problems	Male	21.4%	46.7%	11.5%	51.7%	8.2%	60.7%
	Female	31.9%	38.7%	14.6%	48.2%	6.7%	56.9%
Addition	Male	17.0%	55.3%	14.7%	46.5%	4.9%	59.0%
	Female	26.2%	41.5%	18.7%	42.2%	9.8%	54.7%
Subtraction	Male	21.6%	46.4%	28.0%	46.8%	17.6%	50.0%
	Female	37.3%	35.6%	37.2%	35.8%	23.6%	47.2%
Multiplication	Male			30.7%	37.8%	29.8%	31.7%
	Female			39.0%	30.1%	35.2%	32.1%
Division	Male			35.8%	37.9%	26.1%	38.4%
	Female			43.3%	31.9%	36.3%	34.9%

* < 0.001 ; † < 0.002

Figure F5 displays, by grade, the mean scores for male and female students on each subtask.

Figure F5. EGMA Equateur Treatment: 2014 Mean Percentage Scores by Subtask, Grade, and Sex



EGMA Orientale PAQUED Schools

Table F9 summarizes the percentages of male and female students in the three grades with zero scores and their mean scores at endline (2014) for all EGMA subtasks. The differences in the table are not statistically significant at the $p < 0.001$ and $p < 0.002$ thresholds.

Table F9. EGMA Orientale PAQUED Schools: 2014 Percent Zero Scores and Mean Percentage Scores by Sex and Grade

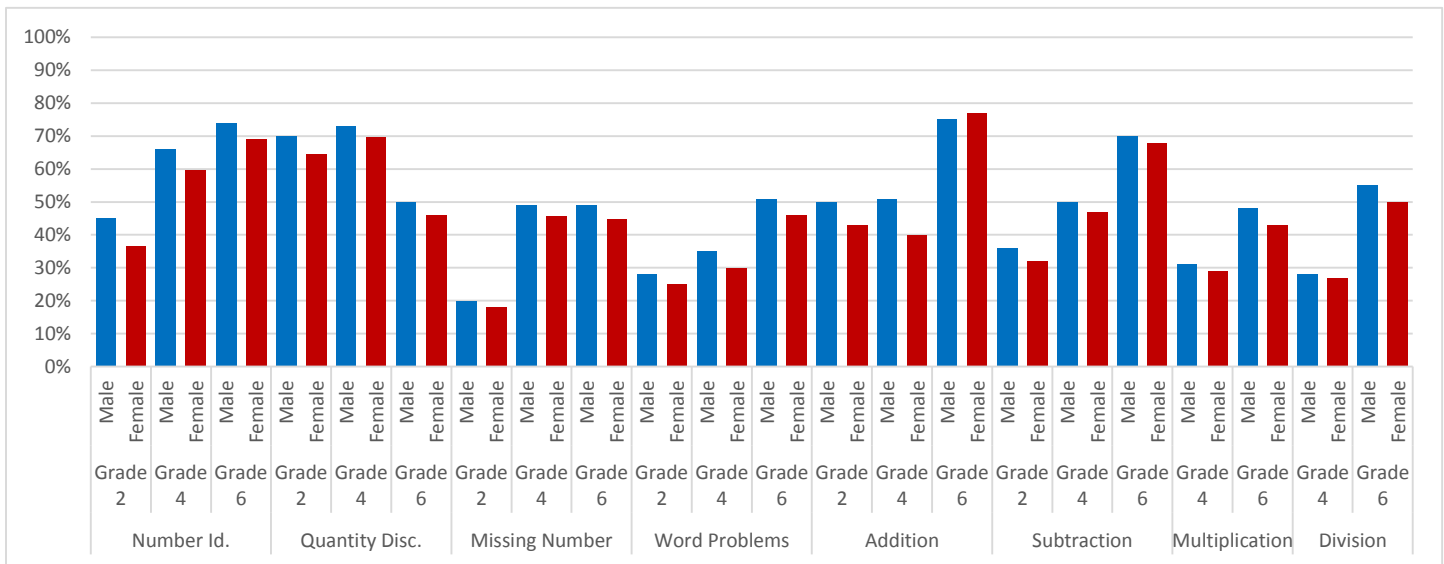
Subtask		Grade 2		Grade 4		Grade 6	
		% Zero score	Mean % score	% Zero score	Mean % score	% Zero score	Mean % score
Number Identification	Male	4.3%	45.2%	0.0%	66.7%	0.7%	74.6%
	Female	4.3%	36.6%	0.0%	59.7%	0.0%	69.1%
Quantity Comparison	Male	0.0%	70.6%	0.7%	73.9%	0.0%	50.7%
	Female	1.4%	64.6%	0.1%	69.8%	0.0%	46.2%
Missing Number	Male	9.2%	20.7%	3.2%	49.2%	2.0%	49.6%
	Female	11.5%	18.8%	3.1%	45.8%	0.8%	44.8%
Word Problems	Male	32.9%	28.4%	19.2%	35.4%	4.8%	51.8%
	Female	46.7%	25.8%	23.4%	37.0%	9.0%	46.0%
Addition	Male	24.3%	50.0%	9.9%	51.2%	1.2%	75.7%

Subtask		Grade 2		Grade 4		Grade 6	
		% Zero score	Mean % score	% Zero score	Mean % score	% Zero score	Mean % score
Subtraction	Female	28.9%	43.1%	10.9%	49.0%	1.5%	77.5%
	Male	42.2%	36.1%	24.7%	50.7%	6.9%	70.3%
Multiplication	Female	46.8%	32.5%	28.3%	47.3%	7.7%	68.5%
	Male			35.6%	31.5%	12.3%	48.4%
Division	Female			33.4%	29.9%	22.4%	43.8%
	Male			43.0%	28.7%	12.1%	55.1%
	Female			40.1%	27.2%	15.0%	51.0%

* < 0.001; † < 0.002

Figure F6 displays, by grade, the mean scores for male and female students on each subtask.

Figure F6. EGMA Orientale Treatment: 2014 Mean Percentage Scores by Subtask, Grade, and Sex



Chapter G: Conclusions and Recommendations

This report presented analyses from a complex group of interrelated studies. The data collected in 2014 and presented here were gathered from 263 schools; 8,323 students; 629 teachers; 265 head teachers; and, crucially, 3 distinct, non-comparable samples. In more than half of the schools studied—the PAQUED and Control schools—Grade 6 students were included and the results were compared to a baseline established in 2010. In slightly less than one third of the schools studied—the Accessible PAQUED and Accessible Control schools—comparisons could be made with a baseline established in 2012, and the study design permitted a high degree of confidence in the estimates generated on the basis of the data. In the remaining sixth of the schools studied—the Reading Program schools—there were no appropriate schools to be used as counterfactuals, and the study design precluded hypothesis testing (checking for significance) over time (from 2012 to 2014).⁷⁰

Table G1 provides a visual representation of the studies presented in this report.

Table G1. Study Populations and Analyses Permitted

Study	Treatment Groups	2010	2012	2014	Evaluation
PAQUED (Grades 2, 4, 6)	PAQUED	Baseline		Endline	Relative growth in performance in PAQUED vs. Control over the period from 2010-2014
	Control				
Accessible PAQUED (Grades 2, 4)	Accessible PAQUED		Baseline	Endline	Relative growth in performance in Accessible PAQUED vs. Accessible Control over the period from 2010-2014
	Accessible Control				
Reading Program (Grades 2, 4)	Reading Program			Snapshot	Performance in 2014

Despite the complexity of the situation being examined, the findings can be summed up fairly simply: these studies do not permit us to make a positive statement regarding the characteristics of an intervention—whether in terms of its components or its intensity—required in order to realize meaningful improvements in student performance in reading or mathematics. However, these studies do allow us to state with confidence that the support provided to Accessible PAQUED and PAQUED schools was not sufficient to generate learning gains that are either significant or substantive.

⁷⁰ As discussed in *Chapter A*, the purposive selection of the Reading Program schools precludes the generalization of their results to the Accessible PAQUED population. The population from which Accessible Control schools were sampled was created by selecting schools which were comparable to the Accessible PAQUED in terms of student population, location, and other school characteristics. As the Reading Program schools cannot be compared to the Accessible PAQUED schools, they cannot therefore be compared to the Accessible Control schools either.

Discussion of Reading Outcomes

From a theoretical standpoint, it is entirely reasonable to expect that an intervention which focused on the skills being measured for the impact evaluation for only a single 30-minute lesson each week (see *Chapter A, PAQUED Inputs Targeting Teachers*) would have minimal impact under the best of circumstances. (Due to the requirement to align with the existing French curriculum, this was the level of reading-specific support provided to the largest population studied, the PAQUED schools.) Given the implementation challenges the project encountered in the PAQUED schools, it is thus hardly surprising that only a handful of comparisons revealed statistically-significant differences. Given the large sample sizes assessed, large numbers of analyses conducted, and the absence of consistent directionality in the statistically-significant differences noted, it is likely these are little more than noise.⁷¹

While implementation of the intervention as designed went more smoothly in Accessible PAQUED schools than in the larger population of PAQUED schools, it had no discernible impact on student reading performance. While not conclusive, this is at least suggestive that the package of inputs that constituted the intervention were inadequate to the task of improving student reading, and it was not simply a matter of poor implementation. (These inputs may have been technically appropriate for developing French oral language skills—i.e., speaking and listening—that were not explored as deeply as reading skills deeply by the EGRA as adapted in 2009 for the DRC context.)

On the other hand, the intensified, focused intervention delivered to the Reading Program schools certainly bore the hallmarks of the sort of well-designed reading-improvement schemes that have demonstrated success elsewhere: books in the hands of students, including decodable and leveled texts; read-aloud books and teacher guides for the instructors; comprehensive training for teachers on effective reading strategies and the use of the materials provided; at least semi-regular coaching and mentoring by someone trained in reading pedagogy; and—crucially—adequate, daily instructional time focused on reading.

While scores on most reading subtasks appeared to improve slightly from 2012 to 2014, the study design showed that relatively few gains were statistically significant and not the result of chance. Such statistical confirmation would be particularly valuable because the *Vocabulary, Initial Sound Identification, Grapheme Recognition, Familiar Word Reading, Invented Word Reading, Oral Reading Fluency, Reading Comprehension*, and

⁷¹ The statistically significant differences were the following: Grade 4 male students in Control schools in Equateur outperformed their female classmates on the *Vocabulary* subtask at the 2010 baseline (but were no longer doing so by the 2014 endline); Grade 4 male students in Control schools in Equateur outperformed their female classmates on the *Grapheme Recognition* subtask at the 2014 endline (although they had not been doing so at the 2010 baseline); and Grade 6 male students in PAQUED schools in Orientale outperformed their female classmates on the *Familiar Word, Invented Word, and Oral Reading Fluency* subtasks at the 2014 endline (when they had not been doing so at the 2010 baseline); and Grade 6 male students in Control schools in Bandundu outperformed their female classmates on the *Vocabulary* subtask at baseline in 2010 (but were no longer doing so by endline in 2014).

Dictation subtasks—that is, every subtask save *Listening Comprehension*—all included at least one instance of student performance in Reading Program schools appearing to decline, rather than improve, from 2012 to 2014. It would be important to determine whether these apparent changes (either growth or decline) represent legitimate differences in student learning or simply result from the vagaries of random student sampling. However, even were hypothesis testing possible, the absence of a counterfactual (i.e., control group) would preclude asserting that any observed gains were due to the Reading Program intervention rather than some other unknown force.

When looking across the three studies, a few promising trends can be seen. Whether in PAQUED, Accessible PAQUED, or Reading Program schools, the reading performance of students in the higher grades exceeded that of students in the lower grades assessed. (i.e., Grade 6 scores were generally greater than Grade 4 scores, which in turn were greater than Grade 2 scores.) While this is naturally what one would hope to see in any education system, given the challenges facing students, teachers, and school administrators in the DRC, it is well not to take it for granted. Also, while not statistically significant at the more conservative ($p < 0.0006$) threshold, a trend for girls from certain provinces in the Reading Program schools to outperform boys at endline—even after no difference emerged in those schools in 2012—suggests that the type of intervention used in these schools may either encourage greater participation by girls or improve the efficacy of the learning experience for girls and should be further explored in subsequent implementations.

It is also promising that support for students appears to improve student performance. For students in Reading Program schools, for example, being taught by teachers who participated in the program (as measured by engagement in forums, receiving training, and receiving visits by program staff) enhanced student performance on the *Grapheme Recognition* subtask. Similarly, having access to books in the classroom significantly correlated with student performance in Grades 2 and 4. It is also promising that other forms of support to students—albeit ones over which schools cannot often exert direct influence—such as attending kindergarten, having books at home, speaking French at home, and having someone at home who is able to read, correlate with student performance.

Ultimately, overall student performance in 2014, in all populations and treatment groups, is below what is needed to meet national benchmarks across all subtasks. Even for oral skills such as vocabulary, phonemic awareness, and listening comprehension, students failed to demonstrate French oral skills required to effectively read in French. Student mean scores on grapheme and word recognition, as well as connected text reading, were also lower than necessary for reading French with fluency and comprehension.

Discussion of Mathematics Outcomes

There is little to be said about the effect of PAQUED’s intervention on student performance in mathematics, regardless of the study in question. Overall student

performance was poor. Statistically significant differences in student performance on mathematics were largely absent from *Chapters E* and *F* of this report.

As discussed in more detail in *Chapter E*, EGMA subtasks can be arranged into two groups: those which assess more procedural knowledge and those which assess more conceptual knowledge. In general students performed better on the more procedural tasks than they did on the more conceptual tasks, and the percentage of students with zero scores was much lower for the more procedural tasks than it was on the more conceptual tasks. This is hardly surprising, and is not likely a function of the PAQUED intervention.

What may be a function of the Reading Program intervention—and may warrant further study—is an apparent reversal of the tendency for boys to outperform girls in math. While hardly any of the comparisons of math performance between the sexes met the conservative thresholds of significance set by the Bonferroni correction, a hint of a trend appeared in favor of girls in Reading Program schools outperforming boys. This contrasts with the possible (albeit not statistically significant) trend in the other populations studied for boys to outperform girls. Again, in light of the apparent absence of a consistent relationship between the intensity of overall support provided to a school and student performance on math, it is likely best to avoid reading too much into this observation.

Recommendations

The absence of a clear narrative of improved student reading or math performance driven by the PAQUED interventions—whether the original package of inputs (PAQUED and Accessible PAQUED schools) or the modified package provided to Reading Program schools—precludes issuing definitive recommendations for the improvement of student learning outcomes in the DRC. However, the following points bear consideration.

- **Re-evaluate the impact of the Reading Program after another year of intervention.**

Most elements of the Reading Program exclusively targeted teachers and students in the *degré élémentaire*—that is, Grades 1 and 2. Therefore, Grade 2 students assessed at endline had only been exposed to one school year of the program. It is possible that students who had completed both Grade 1 *and* 2 under the Reading Program might demonstrate substantively greater performance gains compared to students who had only received the intervention while in Grade 2.

- **Use a research design appropriate to the task.**

The research that concluded with the 2014 endline EGRA and EGMA was never designed with the Reading Program schools in mind. As such, the Reading Program schools lacked a proper baseline, lacked a proper counterfactual, and lacked the sample size required to conduct deep analyses of all the questions that might be of theoretical interest. If the DRC education community wants discussions about the future path of reading instruction in the Congo to be

concretely informed by quantitative data on the viability of the Reading Program's approach, this would be an important step.⁷²

The above notwithstanding, there are elements of the Reading Program intervention that appear promising and are in line with both the MEPSP's vision and strategic documents and also with approaches that some USAID implementers have successfully applied elsewhere (including in Kenya, Liberia, Malawi, Egypt, and Jordan). Provision of books to students—including decodable and leveled readers—that are aligned with teacher guides that focus explicitly on the teaching of reading is one such element. Another, is the incorporation of content on reading pedagogy into in-service teacher training modules (such as the school- and cluster-level *forums d'échange*) is another. Finally, providing teachers with regular, ongoing support by coaches who have been trained in reading pedagogy is a third. That the composite regression model of teacher participation demonstrated a positive and significant correlation with student performance on the *Grapheme Recognition* subtask is an encouraging early sign of the Reading Program's potential.

⁷² A team led by Professor Pierre Mukendi of the *Université de Kinshasa* (UNIKIN), a member of the National Reading Commission, completed a qualitative endline study of the PAQUED project. The report his team produced, and the Lessons Learned Workshop that was convened in August 2014 to discuss both its findings and the preliminary analyses of the 2014 endline EGRA/EGMA studies, provide valuable insights that go beyond what the data that are the subject of this report can state. To request a copy of Prof. Mukendi's report, please contact Susan Ross at sross@edc.org.

Annex I: Details on Study Methodology

This Annex provides more detail regarding the methodology of the current study. The majority of the discussion focuses on the sample designs (and how they shifted over time) and associated weighting procedures. The final section of the Annex briefly discusses the analytical methods applied to the data.

Sample Designs for 2010, 2012, 2014

This section presents a discussion of the various sample designs that have been employed by the *Projet d'Amélioration de la Qualité de l'Éducation* (PAQUED) project from 2010 through 2014, in order to clarify the constraints in terms of comparability and generalizability that have informed the structure of the current report.

The logistical challenges in the Democratic Republic of Congo (DRC) are significant, and so PAQUED took two important issues under consideration when determining the eligibility of a subdivision to participate in the intervention: accessibility and safety. For a subdivision to meet all the *Accessibility* criteria it could not be located in a flood zone, it must be no more than three days' journey away from the center of the province, and it must be primarily accessible by land. For a subdivision to meet the *Safety* criterion, visiting most schools in the subdivision could not pose a high security risk to project personnel.

At the time of the 2010 early grade reading assessment (EGRA)/early grade mathematics assessment (EGMA), there were 18 subdivisions in Bandundu, 17 subdivisions in Equateur, and 14 subdivisions in Orientale that met the criteria for inclusion in the PAQUED intervention.

2010 Sample Design

The EGRA and EGMA data collection was carried out in grades 2, 4, and 6, in the three PAQUED provinces of Bandundu, Equateur, and Orientale, encompassing a total of 144 schools. Of these schools, 109 were from PAQUED (approximately 36 per province) and 35 were Control schools (approximately 12 per province), which would allow for two types of comparisons to be made over the life of the project: (1) learning gains in PAQUED project schools versus non-project (Control) schools, and (2) interprovincial comparisons.

The study used a clustered sampling approach to select PAQUED schools as well as similar Control schools. RTI International randomly selected six subdivisions in each province from among subdivisions that the PAQUED partners had identified as being eligible according to the factors detailed above. RTI then randomly selected six PAQUED schools from each of those subdivisions. Finally, PAQUED identified three to four schools with characteristics similar to those of the program schools, and from among these, RTI selected two at random to serve as Control schools.

Sampling at the school level involved randomly selecting 13 children per grade to participate in the assessment. Specifically, one teacher from each grade was selected at random and then 13 children were selected from those present in that teacher’s class on the day of assessment. Each child was assessed using both the EGRA and EGMA instruments. A Ministry of Primary, Secondary, and Professional Education (*Ministère de l’Enseignement Primaire, Secondaire, et Professionnelle*, [MEPSP]) supervisor, who led a team of four assessors, was responsible for randomly selecting children from each grade. Student sampling was not gender specific, so the proportion of females and males in the sample should represent the proportions occurring in the student population. Assessment was done in the morning, the time when children would normally be in classes and would likely be most alert.

The final study sample consisted of 5,461 students from 144 schools in three provinces. We sampled students in nearly equal numbers by province, grade level, and sex.

Table A1.1: 2010 Sample: Geographic Distribution by Treatment Type

Province	PAQUED Schools	Control Schools
Bandundu	36	12
Equateur	36	11
Orientale	37	11
Total	109	35

2012 Sample Design

While the 2012 study also applied criteria related to *Accessibility* and *Safety*, PAQUED applied a different definition of accessibility in 2012 than it had used in 2010. The new definition added the requirement that, to the best of PAQUED’s knowledge, the school be located within 20 km of an urban center. Furthermore, for a subdivision to be eligible for inclusion in the Midterm assessment—as opposed to eligible for inclusion in the PAQUED intervention—it needed to contain at least six schools deemed “accessible” according to the more stringent 2012 definition of the term.

At the time of the 2012 EGRA/EGMA, these changes to eligibility criteria meant the number of subdivisions from which RTI could randomly select schools had dropped to six in Bandundu (from 18 in 2010), eight in Equateur (from 17), and five in Orientale (from 14).

The 2012 EGRA and EGMA data collection was carried out in Grades 2 and 4 in the three PAQUED provinces of Bandundu, Equateur, and Orientale. It included a total of 95

schools, of which 60 were from PAQUED and 35 were control schools (approximately 12 per province). The sample was designed to allow for two types of comparisons to be made over the remainder of the project: (1) Learning gains in Accessible PAQUED project schools versus non-project (Accessible Control) schools, and (2) interprovincial comparisons.

The study used a clustered sampling approach to select schools deemed by PAQUED to be “accessible” schools as well as similar Accessible Control schools. First, RTI randomly selected six subdivisions in each province from among subdivisions that the PAQUED partners had identified as being eligible.⁷³ RTI then randomly selected six program schools in each of those subdivisions. The partners identified three to four schools with characteristics similar to those of the program schools, and from among these, selected two at random to serve as Accessible Control schools.

Sampling at the school level involved randomly selecting 13 children per grade to participate in the assessment. One teacher from each grade was selected at random, and then 13 children were randomly selected from those present in that teacher’s class on the day of assessment. Student sampling was not gender specific, so the proportion of females and males in the sample should represent the proportions occurring in the student population. Each child was assessed using both the EGRA and EGMA instruments.

The final 2012 sample included 2,453 students from 95 schools in three provinces. We sampled students in nearly equal numbers by province, grade level, and sex.

Table A1.2: 2012 Sample Geographic Distribution by Treatment Type

Province	Accessible PAQUED Schools	Accessible Control Schools
Bandundu	20	12
Equateur	20	11
Orientale	20	12
Total	60	35

2014 Sample Design Considerations

Because the 2010 baseline study and the 2012 midterm study had different target population/frame files, RTI recommended that the 2014 endline assessment have two sample components:

⁷³ In Orientale, where there were only five eligible subdivisions, only five were selected.

- **Sample A**, which would follow the 2010 baseline sample design and sampling frame, and
- **Sample B**, which would follow the 2012 midterm sample design and sampling frame.

The data collected under the Sample A study are reported in *Chapters C* and *D* of this report as the PAQUED schools results. The data collected under the Sample B study encompass both the Reading Program Schools (reported in *Chapter B* of this report) and the Accessible PAQUED schools (reported in *Chapters C* and *D*).

2014 Sampling Frame

The first step in drawing the 2014 samples involved modifying the list of treatment schools used in the 2010 sample selection to only include schools which were still participating in the PAQUED program four years later.⁷⁴ RTI drew the first stage of the sample (schools within provinces) from this revised population frame. For the second stage of sampling (children within schools), a total of 13 students per grade were randomly selected to participate in the assessment. One teacher from each grade was selected at random, and then 13 children were randomly selected from those present in that teacher's class on the day of assessment. Student sampling was not gender specific, so the proportion of females and males in the sample should represent the proportions occurring in the student population. Each child was assessed using both the EGRA and EGMA instruments.

Schools for both Sample A and Sample B were selected independently from the 2010 and 2012 sample, meaning that even if a school was sampled in either 2010 or 2012, it still remained on the sampling frame for 2014. Because the 2014 sample was selected randomly, schools assessed in prior years could have been included in the 2014 sample.

2014 Sample Selection

The Sample A study design permits two types of comparisons. First, it enables the evaluators to identify learning gains in PAQUED schools between 2010 and 2014 and to understand how these gains compare to those realized by Control schools over the same span. Second, it permits interprovincial comparisons of PAQUED Intervention schools. The analyses presented in *Chapters C* and *D* of this report includes these comparisons for students in grades 2, 4, and 6.

Likewise, the Sample B study design permits two types of comparisons. First, it enables PAQUED to identify learning gains in Accessible PAQUED schools between 2012 and

⁷⁴ EDC informed RTI that some schools that were originally part of the 2010 treatment population had dropped out of the program. These schools were therefore excluded from the sampling frame. Only schools which had been part of the treatment program since 2010 were considered eligible for inclusion in the two 2014 sample frames.

2014 and to understand how these gains compare to those realized by Accessible Control schools over the same span. Second, it permits interprovincial comparisons of Accessible PAQUED schools. The analyses presented in *Chapters C* and *D* of this report includes these comparisons for students in grades 2 and 4. As Grade 6 students were not evaluated during the 2012 assessment, the Sample B study could not provide insight into changes in their performance over time and they were thus excluded from the 2014 study as well.

To draw Sample B, RTI randomly selected six subdivisions in each province from among the subdivisions that PAQUED had identified as being eligible using the 2012 criteria.⁷⁵ RTI randomly selected four Accessible PAQUED schools in each of the six subdivisions. The schools that would serve as Accessible Control schools in 2014 were drawn from the same Accessible Control population frame that was used in 2012.

Reading Program Schools included in 2014

Following the project realignment in January 2013, 45 schools from across the three provinces began to receive an additional intervention and close support. These schools, known in this report as Reading Program⁷⁶ schools, received a modified intervention that is discussed in more detail in the *Elements of the “Reading Program” Intervention Post-Realignment* section of *Chapter A*. Only 20 of the 44 Reading Program schools were both sampled and assessed during the 2012 EGRA/EGMA, however, which means that the sample size is too small to permit trustworthy analyses of the Reading Program schools’ change in performance over time. The 24 remaining schools were included in Sample B in 2014 (in addition to the 20 first assessed in 2012).

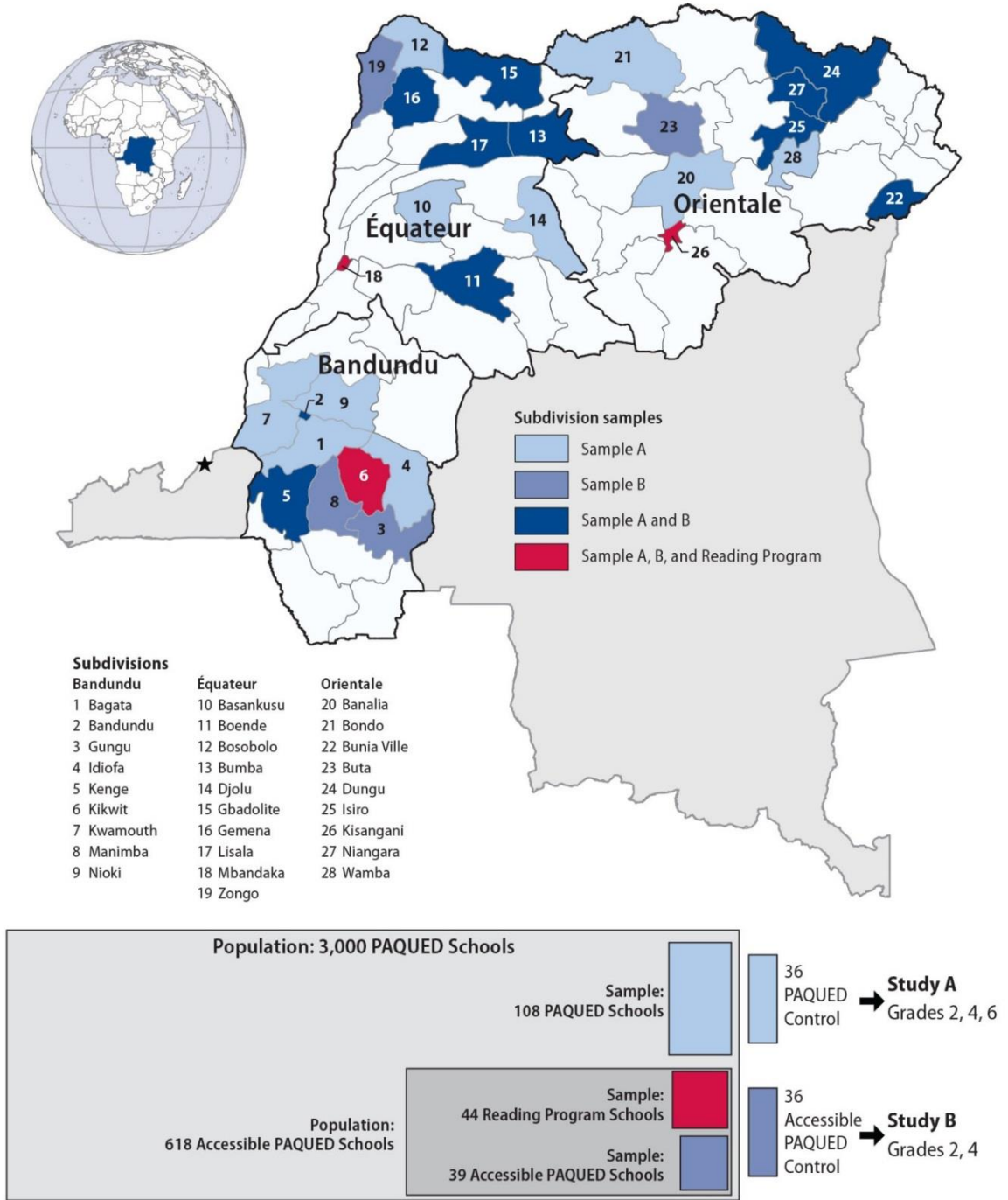
Figure A1.1, reproduced from *Chapter A* of the report, provides two visual representations of the 2014 Endline EGRA/EGMA sample and the populations of schools from which it was drawn. As explained in the main narrative, the top portion of the image includes a map of the educational subdivisions in which schools were assessed, color-coded according to the population(s) whose schools were located in that subdivision. The bottom of the image provides scale representation of the relative sizes of the populations being studied and the samples drawn. (i.e., the ratio of the area of the light grey *Population: 3,000 PAQUED Schools* rectangle to the area of the dark grey *Population: 618 Accessible PAQUED Schools* is 3,000:618, or ~4.85:1.) In the map and visualization, Study A refers to the PAQUED and Control schools; Study B refers to the Accessible PAQUED and Accessible Control schools; and the Reading Program schools are nested

⁷⁵ As a reminder, the 2012 criteria for subdivision eligibility required that the subdivision contain at least six “accessible” schools, where school accessibility was defined as being located within 20 km of an urban center. Once again, because there were only five subdivisions in Orientale that met the 2012 eligibility criteria, only five were selected.

⁷⁶ In PAQUED’s quarterly reporting and discussions with USAID, these were known as “experimental” schools. However, because their selection and the study design did not adhere to experimental or quasi-experimental principles, this report refers to them as “Reading Program” schools in order to avoid misleading the reader.

within the Study B shape because they were a subset of the Accessible PAQUED schools.

Figure A1.1 Subdivisions Containing Schools Assessed at Endline and Relative Size of the Study Populations



The following two tables show the sample sizes that were projected for both schools and students for Samples A and B. The actual number of schools visited and students sampled is reported in the Descriptive Statistics sections of *Chapters B–D*. *Tables A1.3* and *A1.5* show the school-level sample sizes, while *Tables A1.4* and *A1.6* show the student-level sample sizes. Each table also shows the geographic breakdown by treatment type.

Table A1.3: 2014 Sample A Number of Schools Sampled by Geographic Distribution and Treatment Type

Province	PAQUED Schools	Control Schools	Total
Bandundu	36	12	48
Equateur	36	12	48
Orientale	36	12	48
Total	108	36	144

Table A1.4: 2014 Sample A Number of Students Sampled by Geographic Distribution, Treatment Type, and Grade

Province/Grade	PAQUED Students	Control Students	Total
Bandundu	1404	468	1872
Grade 2	468	156	624
Grade 4	468	156	624
Grade 6	468	156	624
Equateur	1404	468	1872
Grade 2	468	156	624
Grade 4	468	156	624
Grade 6	468	156	624
Orientale	1404	468	1872
Grade 2	468	156	624
Grade 4	468	156	624
Grade 6	468	156	624

Total	4212	1404	5616
Grade 2	1404	468	1872
Grade 4	1404	468	1872
Grade 6	1404	468	1872

Table A1.5: 2014 Sample B Number of Schools Sampled by Geographic Distribution and Treatment Type

Province	Accessible PAQUED Schools	Accessible Control Schools	Reading Program Schools	Total
Bandundu	12	12	17	41
Equateur	12	12	16	40
Orientale	15 ⁷⁷	12	11 ⁷⁸	38
Total	39	36	44	119

Table A1.6: 2014 Sample B Number of Students Sampled by Geographic Distribution, Treatment Type, and Grade

Province/Grade	Accessible PAQUED Students	Accessible Control Students	Reading Program Students	Total
Bandundu	312	312	442	1066
Grade 2	156	156	221	533
Grade 4	156	156	221	533
Equateur	312	312	416	1040
Grade 2	156	156	208	520
Grade 4	156	156	208	520
Orientale	390	312	286	988
Grade 2	195	156	143	494
Grade 4	195	156	143	494

⁷⁷ There were only five subdivisions in Orientale that met the eligibility requirement of containing at least 6 “accessible” schools. In order to meet the necessary sample size, three schools were sampled per subdivision in Orientale rather than two schools per subdivision as in the other provinces.

⁷⁸ There were only 11 Reading Program schools in Orientale

Province/Grade	Accessible PAQUED Students	Accessible Control Students	Reading Program Students	Total
Total	1014	936	1144	3094
Grade 2	507	468	572	1547
Grade 4	507	468	572	1547

Probability of Selection and Design Weight – Samples A and B

The methods described in this section for probability of selection and design weight were used to weight both Sample A and Sample B.

The probability of selection for subdivisions within the province is the total number of subdivisions (SBDs) sampled in the province divided by the population number of SBDs in the province

$$pos_{1hi} = \#(\text{SBDs sampled})_h / \#(\text{population SBDs})_h$$

where $\#(\text{SBDs sampled})_h$ is the number of SBDs sampled in the h^{th} province and $\#(\text{population SBDs})_h$ is the number of population SBDs in the h^{th} province. The design weight in a province is the inverse probability of selection in the province. That is, the design weight for the i^{th} SBD in the h^{th} province (d_{hi}) is

$$d_{1hi} = 1 / pos_{1hi}$$

The probability of selection for schools within a subdivision is the total number of schools sampled in the subdivisions divided by the population number of schools in the subdivision

$$pos_{2hi} = \#(\text{schools sampled})_h / \#(\text{population schools})_h$$

where $\#(\text{schools sampled})_h$ is the number of schools sampled in the h^{th} SBD and $\#(\text{population schools})_h$ is the number of population schools in the h^{th} SBD. The design weight in a SBD is the inverse probability of selection in the SBD. That is, the design weight for the i^{th} school in the h^{th} SBD (d_{hi}) is

$$d_{2hi} = 1 / pos_{2hi}$$

The probability of selection for students within school by grade and gender is the total number of students sampled by gender in the grade divided by the population number of students in that grade

$$pos_{jlk} = \#(\text{students sampled})_{jlk} / \#(\text{population students})_{jl}$$

where $\#(\text{students sampled})_{jlk}$ is the number of students in the j^{th} school for the k^{th} gender and $\#(\text{population students})_{jl}$ is the number of students in the j^{th} school for the l^{th} grade. The design weight at this level for a student in a school is the inverse probability of selection in the school by grade and gender

$$d_{jlk} = 1 / pos_{jlk}$$

The final analysis weight used for all the analyses in the 2014 endline assessment report is the product of the first-stage design weight (WT) and the second-stage design weight

$$WT = d_{1hi} * d_{2hi} * d_{jlk}$$

This discussion of weighting only addresses subdivisions and schools. Readers may have noticed that there was an additional layer of random sampling at each school—that of teachers within grades. A lack of reliable data in our population frames on the number of teachers (or class sections) per grade in each school made it impossible to properly account for this in the sample design and weighting processes. As the bias exists across all three data collection periods (2010, 2012, and 2014), it does not affect the resultant analyses.

Weights for Sample A

Sample A: Stage 1

Table A1.7: 2014 Sample A Frame and Sample Counts: Subdivisions (Stage 1)

Province	Number of Eligible SBDs	Number of Sampled SBDs	WT1 = (# eligible SBDs/number sampled SBDs) ⁷⁹	Finite Population Correction (FPC) 1
Bandundu	23	6	3.833	23
Equateur	28	6	4.667	28
Orientale	25	6	4.167	25
Total	76	18	--	--

⁷⁹ Note WT1 and FPC1 are the same for either control or treatment schools since the eligibility of a SBD was defined based on the number of PAQUED schools contained within it.

Sample A: Stage 2

Table A1.8: 2014 Sample A Frame and Sample Counts for PAQUED Schools (Stage 2)

Province	Number of PAQUED Schools within Sampled SBDs	Number of Sampled PAQUED Schools within Sampled SBDs	WT2 = (# eligible PAQUED schools/number sampled PAQUED schools)	FPC1
Bandundu	220	36	6.111	220
Equateur	208	36	5.778	208
Orientale	233	36	6.472	233
Total	661	108	--	--

Table A1.9: 2014 Sample A Frame and Sample Counts for Control Schools (Stage 2)

Province	Number of Control Schools within Sampled SBDs	Number of Sampled Control Schools within Sampled SBDs	WT2 = (# eligible Control schools/number sampled Control schools)	FPC1
Bandundu	883	12	73.583	883
Equateur	453	12	37.75	453
Orientale	496	12	41.33	496
Total	1,832	36	--	--

Weights for Sample B

Sample B: Stage 1

Table A1.10: 2014 Sample B Frame and Sample Counts for Subdivisions (Stage 1)

Province	Number of Eligible SBDs	Number of Sampled SBDs	WT1 = (# eligible SBDs/number sampled SBDs) ⁸⁰	FPC1
Bandundu	6	6	1	6
Equateur	8	6	1.33	8
Orientale	5	5	1	5
Total	19	17	--	--

Sample B: Stage 2

Table A1.11: 2014 Sample B Frame and Sample Counts for Accessible PAQUED Schools (Stage 2)

Province	Number of Accessible PAQUED Schools within Sampled SBDs	Number of Sampled Accessible PAQUED Schools within Sampled SBDs	WT2 = (# eligible Accessible PAQUED schools/number sampled Accessible PAQUED schools)	FPC1
Bandundu	244	12	20.333	244
Equateur	107	12	8.917	107
Orientale	149	15	9.933	149
Total	500	39	--	--

⁸⁰ Note WT1 and FPC1 are the same for either control or treatment schools since the eligibility of a SBD was defined based on the number of Reading Program schools contain within it.

Table A1.12: 2014 Sample B Frame and Sample Counts for Accessible Control Schools (Stage 2)

Province	Number of Accessible Control Schools within Sampled SBDs	Number of Sampled Accessible Control Schools within Sampled SBDs	WT2 = (# eligible Accessible Control schools/number sampled Accessible Control schools)	FPC1
Bandundu	28	12	2.333	28
Equateur	28	12	2.333	28
Orientale	32	12	2.667	32
Total	88	36	--	--

Probability of Selection and Design Weight – Reading Program Schools

Following the 2012 EGRA/EGMA assessments EDC selected 43 Accessible PAQUED schools from across the three provinces to begin receiving the modified, intensified *IAI Plus* treatment. These schools are referred to throughout this report as “Reading Program” schools. Twenty of the 43 Reading Program schools were both sampled and assessed during the 2012 assessment. These schools were not part of a probability sample so any analysis will be explicit to the sample only and cannot be used to make generalizations to the Reading Program population.

Since all schools were selected with certainty after the 2012 assessment, the probability of selection at the school level equals to 1.

$$W_i^1 = 1, \text{ where}$$

W_i^1 = School-level weight for school i .

Sampling at the school level involved randomly selecting 13 children per grade to participate in the assessment. The final level weight for gender k , grade j and school i is defined as:

$$W_{ijk}^2 = \frac{N_{ijk}}{n_{ijk}}, \text{ where}$$

W_{ijk}^2 = Student-level weight for the i^{th} school, gender k and grade j .

N_{ijk} = Number of student in the i^{th} school, by gender k and grade j .

n_{ijk} = Number of sampled students in the i^{th} school, by gender k and grade j .

The final analysis weight is the product of the school and student level weights:

$$W_{ijk}^3 = W_i^1 * W_{ijk}^2$$

Analytical Methods

The main analytical methods used to examine the EGRA and EGMA data for this report included descriptive analyses and simple regression. All analyses utilized the final analysis weight (see Probability of Selection and Design Weight subsection above). Analyses included means comparisons, percentage comparisons, and regression analyses. Statistically significant differences were reported where applicable. Regression analyses were used to examine, and as appropriate provide comparisons for, the effects of key school, student, and teacher characteristics (i.e., SES, language spoken at home, and classroom practices) on performance. For PAQUED and Control schools, the 2014 (endline) data are compared to the 2010 (baseline) data. For Accessible PAQUED and Accessible Control schools, the 2014 (endline) data are compared to the 2012 (midterm) data. For Reading Program schools, the 2014 data are compared to the subset of Reading Program schools assessed in 2012.

Defining alpha values

To guard against inflated type I error rates, Bonferroni adjustments were applied to the alpha value of 0.05 within each chapter of the report. Type I errors occur when we reject the null hypothesis even though it is true. When inflated, we increase the probability of reporting a result as significant when it is not significant.

Because this report contains three non-comparable populations, Bonferroni adjustments were calculated separately for each population.

For the population of Reading Program Schools (**Chapter B** in the report), two families of tests were defined as detailed below.

Family #1 – all tests that examined the existence of a gender effect within treatment group and time.

Family #2 – all tests that examined the impact of outside characteristics on student performance.

Table A1.13: Chapter B Bonferroni Adjustment Families for Reading Program Study, EGRA Results

Subtask	Family #1 Tests (Comparisons of means)	p-value (p < #)	Family # 2 Tests (Outside Characteristics)	p-value (p < #)
Vocabulary				
Initial Sound Identification				
Listening Comprehension	12 per subtask	0.004		
Grapheme Recognition				
Familiar Word				
Invented Word				
Oral Reading	6 per subtask	0.008		
Reading Comprehension				
Dictation				
Student Characteristics			24 per subtask	0.002

For the populations of Accessible PAQUED (and Accessible Control) schools and PAQUED (and Control) schools analyzed in *Chapter C* of the report, three families of tests were defined as detailed below.

Family #1 – All tests that examine changes between treatment groups over time.

Family #2 – all tests that examined the existence of a gender effect within treatment group and time.

Family #3 – all tests that examined the impact of outside characteristics on student performance.

Table A1.14: Chapter C Section 1 - Bonferroni Adjustment Families for Accessible PAQUED Study, EGRA Results

Subtask	Family #1		Family #2		Family #3 Tests	
	Tests ()	p-value (p < #)	Tests ()	p-value (p < #)	(Outside Characteristics)	p-value (p < #)
Vocabulary						
Initial Sound Identification						
Listening Comprehension	6 per subtask	0.008	24 per subtask	0.002		
Grapheme Recognition						
Familiar Word						
Invented Word						
Oral Reading	3 per subtask	0.017	12 per subtask	0.004		
Reading Comprehension						
Dictation						
Student Characteristics					20 per subtask	0.0025

Table A1.15: Chapter C Section 2 - Bonferroni Adjustment Families for PAQUED Study, EGRA Results

Subtask	Family #1		Family #2		Family #3 Tests	
	Tests (D-in-D)	p-value (p < #)	Tests (Comparison of means)	p-value (p < #)	(Outside Characteristics)	p-value (p < #)
Vocabulary						
Initial Sound Identification						
Listening Comprehension	9 per subtask	0.006	36 per subtask	0.001		
Grapheme Recognition						
Familiar Word						
Invented Word						
Oral Reading	6 per subtask	0.008	24 per subtask	0.002		
Reading Comprehension						
Dictation						
Student Characteristics					22 per subtask	0.002

For the populations of Accessible PAQUED (and Accessible Control) schools and PAQUED (and Control) schools analyzed in *Chapter D* of the report, two families of tests were defined as detailed below.

- Family #1 Tests = Difference of Difference Calculations

- Family #2 Tests = 2014 comparisons - treatment vs. control; gender; grade

Table A1.16: Chapter D Sections 1 and 2 - Bonferroni Adjustment Families for Accessible PAQUED Study, EGMA Results

Subtask	Family #1 Tests (D-in-D)	p-value (p < #)	Family # 2 Tests (Comparisons of Means)	p-value (p < #)
Number ID				
Quantity Comparison				
Missing Number	9 per subtask	0.006	36 per subtask	0.001
Word Problems				
Addition				
Subtraction				
Multiplication	6 per subtask	0.008	24 per subtask	0.002
Division				

Table A1.17: Chapter D Sections 3 and 4 - Bonferroni Adjustment Families for Accessible PAQUED Study, EGMA Results

Subtask	Family #1 Tests	p-value (p < #)	Family # 2 Tests	p-value (p < #)
Number ID				
Quantity Comparison				
Missing Number	6 per subtask	0.008	48 per subtask	0.001
Word Problems				
Addition				
Subtraction				
Multiplication	3 per subtask	0.017	24 per subtask	0.002
Division				

***Annex II: The Early Grade Reading Assessment
(EGRA)/Early Grade Mathematics
Assessment (EGMA) Data
Collection Process***

This Annex provides more detail regarding the process of training, selecting, and deploying assessors; the schedule of data collection; and the quality control mechanisms employed to ensure data quality.

Assessor Training and Selection

Training of Assessor Candidates was conducted over 11 days (two full weeks plus an intervening Saturday) from April 17–30, 2014. On three of the 11 days, trainees spent the morning in schools practicing their instrument administration and sampling procedures with children and school staff. Concurrent training workshops were run in the cities of Bandundu Ville (Bandundu), Mbandaka (Equateur), and Kisangani (Orientale). Each training was led by an experienced Tangerine-based early grade reading assessment (EGRA) trainer, with facilitation support from RTI International’s local staff or the local logistics subcontractor’s staff.

A majority of candidates invited to the Assessor Training Workshop had participated in data collection activities in either 2010 or 2012, and many individuals had served as assessors in both preceding years.

The first three days of the training calendar focused on re-introducing the assessment to the trainees. After reviewing the instruments’ subtasks and their administration protocols, trainees were introduced to the 2014 versions of the instruments as rendered on paper. While data collection was ultimately to be conducted using electronic instruments rendered on tablet computers using RTI’s Tangerine® software, it was important that assessors be prepared to revert to paper-based instruments in the event of tablet failure while in the field, therefore, the first practicum day included administration of the instruments on paper. However, for the remainder of the training workshop and the final two practicum days, assessors worked with the tablets.

Trainees practiced repeatedly administering the tests as a group, in pairs, and in trios. On multiple occasions throughout the training workshop, the candidates’ inter-rater reliability (IRR) was evaluated. No candidate achieving less than a 90% score on the IRR assessments was eligible for retention as an assessor. Note that assessors were required to do more than simply perform well on the IRR. They were also scored on their ability to develop a rapport with the student, to faithfully adhere to each subtask’s administration protocols, and to effectively manipulate the tablets.

To conduct the IRR tests, the workshop facilitators annotated a paper version of the student assessment protocol. The annotations included instructions about errors to be made (such as skipping entire rows on the timed grid subtasks, “reading” a true French word in place of the invented word prompt, pointing to the wrong body part on the Vocabulary subtask, etc.), the type of inflection and demeanor to be adopted (read hesitantly; be easily distracted), the student demographic information to be provided, etc. A Congolese member of the facilitation team was given the annotated protocol and instructed to take some time to review it and become familiar with the “performance”

required. This facilitator would then play the role of the student while the workshop facilitator or one of the stronger candidates administered the assessments to the “student.” The assessor candidates congregated around the pair and followed along, completing their assessments in concert with the person administering the tool. The group’s results were then uploaded to the cloud and analyzed for agreement with the mode response for each item. The candidates’ scores were calculated on the basis of the proportion of agreement with the mode response, and remediation activities were planned to address those individual items for which a minimum 80% agreement across all assessors had not yet been obtained.

Trainees who excelled on all of the above measures and also displayed leadership qualities were engaged as Team Supervisors.

The Data Collection Schedule

To allow for maximum exposure of students to the intervention, data was collected as near to the end of the final trimester of the academic year as was practicable. Schools were visited over a period of six weeks beginning May 2, 2014. In most instances, schools were visited in the morning so students could be assessed while at their most alert (exceptions were made for schools operating on the afternoon shift.) Teams of assessors visited schools at an average rate of one school every two days, using the second day to travel from site to site.

Upon arrival at a school, the Team Supervisor approached the Head Teacher to introduce the team, present a letter of authorization from the Ministry of Primary, Secondary, and Professional Education (*Ministère de l’Enseignement Primaire, Secondaire et Professionnelle*, MEPS), and describe the broad contours of the day of data collection. The team would then proceed to sample a teacher at random from each of the grades concerned. Students from the selected teacher’s classroom were then lined up and counted off, with every *n*th student being assessed. (To determine the *n*th student, teams divided the class enrollment by the number of students to be assessed [13]. The quotient then provided the *n*. For instance, in a classroom of 39 students, $39 \div 13 = 3$. Every 3rd student would be selected for assessment.)

Assessors administered the EGRA and early grade mathematics assessment (EGMA) instruments to students. The Supervisor would observe, provide quality assurance, and manage the delivery of sampled students to the next available assessor. As the end of student assessment approached, assessors would administer the Teacher Interview instrument to the teacher whose students were sampled and the Head Teacher Interview instrument to the head teacher of the school.

Quality Control

The midterm assessment was conducted using RTI-developed, open-source electronic data collection software known as Tangerine.⁸¹ The software, which has been customized for the EGRA and EGMA, includes built-in data validity checks, controlled timing, and conditional logic configurations that can help drastically reduce assessor-driven error. Where mobile connectivity allows, the data can be uploaded for nightly review of data quality and quantity. Collecting student results directly into a digital format eliminates the need for hiring data entry clerks to transcribe paper forms, minimizing the introduction of new errors.

During the 2014 endline assessment, most teams of assessors in the provinces of Bandundu and Orientale were able to upload their results on a regular basis (i.e., at least once every two to three days). Connectivity in Equateur was less reliable, and teams were only able to upload data infrequently. Whenever a new batch of results became available, project staff at RTI's home office reviewed the data to confirm that the results were in line with proper EGRA- and EGMA-administration procedures.⁸² Wherever deviations from standard protocol were noted, project staff contacted supervisors and assessors and reminded them of the proper protocols to be followed. The nightly review of data also permitted RTI's statisticians to conduct basic cleaning on an ongoing basis, which accelerated the timeline within which data could be summarized, analyzed, and reported upon.

⁸¹ For more about this software, see the Tangerine website, www.tangerinecentral.org.

⁸² If a child had earned a zero score (first 10 items incorrect) on the letter name knowledge subtest, but the "time remaining" record indicated that more than 30 seconds had elapsed, that would be an indication that the assessor was not observing the three-second rule. Project staff would then contact the assessment team by SMS (text message) or phone and remind the assessor that a child should only be able to hesitate for a full three seconds per item before being directed to move on to the next item.

Annex III: About the Assessment Instruments

This Annex discusses the instruments used for data collection during the 2010, 2012, and 2014 assessments of the *Projet d'Amélioration de la Qualité de l'Éducation* (PAQUED) intervention. The first section describes the original process of adaptation for context in the Democratic Republic of Congo (DRC). The second section presents test statistics, including a discussion of the tests' reliability and principal components analysis. The third section describes the content of the assessments' subtasks, noting why, how, and where the assessments differed from administration to administration.⁸³ The fourth section addresses how the results of the oral reading fluency passage and associated reading comprehension questions were statistically equated across time periods. It also describes quality control measures that were applied during the data collection period.

Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) Adaptation in DRC

The EGRA and EGMA tools have been adapted and administered in more than 70 languages across 50 countries. Initial adaptation of the EGRA and EGMA instrument in the DRC context took place in December 2009. The individuals charged with that adaptation were DRC officials from the Ministry of Primary, Secondary, and Professional Education (*Ministère de l'Enseignement Primaire, Secondaire, et Professionnel*, MEPSP), the *Direction des Programmes Scolaires et Matériel Didactique* (DIPROMAD), the *Inspecteur Général d'Éducation* (IGE), and *Le Service National de Formation* (SERNAFOR) with aid from *Centre de Linguistique Théorique et Appliquée* (CELTA) language experts, who were lead through the EGRA adaptation process by Dr. Souhila Messaoud-Galusi of RTI and the EGMA adaptation process by mathematics education expert Aarnout Brombacher.

For EGRA, participants reviewed the totality of the test—including each individual item (e.g., graphemes and words)—to ensure that language content would be familiar to children being assessed. The short stories used for the reading passage and the *Listening Comprehension* subtask were prepared by ministry officials, with guidance from RTI on early grade reading considerations such as language complexity, story structure and types of questions (including direct and inferential). Local language experts gave guidance on characteristics of the national languages (Lingala, Swahili, Kikongo, and Ciluba) in comparison with French so that interference could be minimized. Changes to subtasks resulting from adaptation included (1) replacing a potential item with a more familiar item of a similar difficulty level for phonemic awareness, and (2) changing a consonant-consonant-vowel (CCV) word to a consonant-vowel (CV) word to get closer to the syllabic structure of the majority of national languages for dictation. For EGMA, the task

⁸³ The full instruments are available upon request. Please contact Timothy Slade at tslade@rti.org.

focused on ensuring that the EGMA subtests and items corresponded to the DRC curricula for the targeted grades.

Once adaptation was complete, the instruments were pre-tested and then piloted. The pre-test was conducted in four schools at the periphery of Kinshasa and chosen by local ministry officials as representing higher and lower performing rural schools. The sample was randomly chosen and included 83 children in Grade 2; 87 in Grade 4; and 86 in Grade 6. The pre-test was evaluated using a Rasch analysis⁸⁴ to show the continuum of difficulty across the items in the test and highlighted areas of the assessment displaying ceiling effects and floor effects. It also pointed to particular items that were potentially measuring the same construct, and therefore redundant or that were not performing well. In second grade, all subtests that required word level reading skills were eliminated because there was a nearly complete floor effect for the more difficult sections (e.g., the Familiar Word Reading, Invented Word Reading, and Oral Reading Fluency subtasks). Additionally, this change was in accordance with the curriculum in the DRC, since MEPS policy is for French to be a primarily oral language in schools until Grade 3.

Test Statistics

Test Reliability

In order to determine whether (and how) the various subtasks in the EGRA/EGMA assessments as implemented in DRC were reliable, and if they tested an underlying skill (i.e., presumably early grade reading and mathematics skills), reliability tests are presented in the tables below. They include simple bivariate correlations between the various subtasks (presented in *Tables A3.1* and *A3.2*) and Cronbach's Alpha Reliability Tests (presented in *Tables A3.3* and *A3.4*). The data used for each of these analyses is drawn from the largest, most representative dataset available: performance of Grades 2, 4, and 6 students in PAQUED schools.

⁸⁴ Rasch Analyses apply a probabilistic model that uses item responses to determine student ability and item difficulty. Put another way, the Rasch Model uses student responses on items to create a metric to measure an unobserved construct. For further reading, please see <http://www.rasch.org/rasch.htm>

Table A3.1: Correlation Matrix for EGRA Tasks

Subtask	Vocabulary	Initial Sound Identification	Grapheme Recognition	Familiar Word Reading	Invented Word Reading	Oral Reading Fluency	Reading Comprehension	Listening Comprehension	Dictation
Vocabulary	1								
Initial Sound Identification	0.3300	1							
Grapheme Recognition	0.5901	.3700	1						
Familiar Word Reading	0.4538	.2593	.7426	1					
Invented Word Reading	0.4306	.2544	.7231	.8914	1				
Oral Reading Fluency	0.4656	.2649	.7463	.8955	.8672	1			
Reading Comprehension	0.4173	.2095	.5248	.6710	.6659	.7032	1		
Listening Comprehension	0.4995	.2618	.4497	.4262	.4261	.4498	.5073	1	
Dictation	0.3285	.2681	.5699	.6068	.5823	.6168	.4421	.3033	1

Table A3.2: Correlation Matrix for EGMA Tasks

Subtask	Counting (Ones)	Counting (Tens)	Counting Objects (Single)	Counting Objects (Grouped)	Number Identification	Quantity Comparison	Missing Numbers	Word Problems	Addition	Subtraction	Multiplication	Division
Counting (Ones)	1											
Counting (Tens)	0.1936	1										
Counting Objects (Single)	0.3027	0.2655	1									
Counting Objects (Grouped)	0.0742	0.3013	0.1757	1								
Number Identification	0.1571	0.3651	0.0286	0.1957	1							
Quantity Comparison	0.1091	0.2001	0.1743	0.165	0.185	1						
Missing Numbers	0.1389	0.319	0.3986	0.1409	0.2637	0.2781	1					
Word Problems	0.0932	0.2388	0.2947	0.1804	0.1933	0.1016	0.4683	1				
Addition	0.1012	0.2428	0.0992	0.1035	0.4077	0.0501	0.2917	0.2535	1			
Subtraction	0.0933	0.2363	0.0675	0.1388	0.3534	0.188	0.2875	0.218	0.5221	1		
Multiplication	0.0527	0.2105	-0.0006	0.1286	0.322	0.1682	0.2185	0.1418	0.4217	0.5339	1	

Subtask	Counting (Ones)	Counting (Tens)	Counting Objects (Single)	Counting Objects (Grouped)	Number Identification	Quantity Comparison	Missing Numbers	Word Problems	Addition	Subtraction	Multiplication	Division
Division	0.0636	0.2102	0.0197	0.0826	0.3222	0.127	0.2271	0.1875	0.4223	0.5113	0.6548	1

The results of Cronbach’s alpha reliability tests conducted for EGRA and EGMA using the Sample A (PAQUED schools) data are presented in *Tables A3.3* and *A3.4* below.

Table A3.3: Reliability Analysis of EGRA Tool

Subtask	Item-Test Correlation	Item-Rest Correlation	Average Inter-Item Correlation	Alpha
Vocabulary	0.7263	0.5887	0.5174	0.8956
Initial Sound Identification	0.5375	0.3848	0.5710	0.9141
Grapheme Recognition	0.8234	0.7300	0.4768	0.8794
Familiar Word Reading	0.8859	0.8457	0.4689	0.8760
Invented Word Reading	0.8708	0.8259	0.4723	0.8775
Oral Reading Fluency	0.8950	0.8575	0.4669	0.8751
Reading Comprehension	0.7674	0.6934	0.4949	0.8869
Listening Comprehension	0.6774	0.5409	0.5278	0.8994
Dictation	0.7087	0.6224	0.5076	0.8919
Overall Test			0.4999	0.9000

Table A3.4: Reliability analysis of EGMA Tool

Subtask	Item-Test Correlation	Item-Rest Correlation	Average Inter-Item Correlation	Alpha
Counting (Ones)	0.3706	0.2142	0.2578	0.7926
Counting (Tens)	0.5613	0.4125	0.2432	0.7795
Counting Objects (Single)	0.4508	0.2489	0.2547	0.7898
Counting Objects (Grouped)	0.4202	0.2698	0.2548	0.7899
Number Identification	0.5884	0.3268	0.2422	0.7786
Quantity Comparison	0.4576	0.2364	0.2656	0.7991
Missing Numbers	0.68	0.5346	0.2316	0.7683

Subtask	Item-Test Correlation	Item-Rest Correlation	Average Inter-Item Correlation	Alpha
Word Problems	0.5768	0.411	0.2492	0.785
Addition	0.6292	0.4962	0.2354	0.772
Subtraction	0.6854	0.5719	0.2303	0.7669
Multiplication	0.6525	0.5295	0.2354	0.772
Division	0.6511	0.5271	0.2357	0.7723
Overall Test			0.2448	0.7955

Modifications to the EGRA and EGMA Instruments Over Time

In order to preserve testing integrity, instrument modifications are required in situations of multiple administrations. However, it is critical to retain comparability of results between the baseline and subsequent midterm or endline assessments. Therefore, most modifications must be minor, such as re-randomization of items within a given subtask.

When updating the DRC EGRA instruments for subsequent rounds of assessment, most changes were limited to re-randomization of items in a given subtask. The exceptions were the *Oral Reading Fluency* and *Reading Comprehension* subtasks each year, and the *Listening Comprehension* task. When updating DRC EGMA instruments for subsequent rounds of assessment, minor modifications to the content of each subtask were made. In general roughly 20% of items were retained unchanged from one year to the next, and the remainder were replaced with different items of similar difficulty. *Table A3.5* indicates the nature of any changes made to the EGRA and EGMA subtests between baseline and midterm.

Table A3.5: Comparison of EGRA and EGMA Instruments

	Year of Assessment			Modifications to Content
	2010	2012	2014	
Early Grade Reading Assessment				
Title of Subtest				
<i>Vocabulaire</i> (Vocabulary)	Minor Content Modifications	Minor Content Modifications		Re-randomization of items within each subsection (parts of the body; objects in the environment; spatial relationships)

	Year of Assessment			Modifications to Content
	2010	2012	2014	
<i>Identification du Son Initial</i> (Initial Sound Identification)	Minor Content Modifications	Minor Content Modifications	Minor Content Modifications	Internal re-randomization of items 1-5 (before the zero-score cutoff) and of items 6-10 (after the zero-score cutoff)
<i>Connaissance des Graphèmes</i> (Grapheme Recognition)	Minor Content Modifications	Minor Content Modifications	Minor Content Modifications	Re-randomization of all 100 items
<i>Lecture des Mots Familiers</i> (Familiar Word Reading)	Minor Content Modifications	Minor Content Modifications	Minor Content Modifications	Re-randomization of all 50 items
<i>Lecture des Mots Inventés</i> (Invented Word Reading)	Minor Content Modifications	Minor Content Modifications	Minor Content Modifications	Re-randomization of all 50 items
<i>Lecture du Texte (Petite Histoire)</i> (Oral Reading Fluency)	Title updated to <i>Lecture de Texte</i> ; Minor Content Modifications	Minor Content Modifications	Minor Content Modifications	Complete replacement of reading passage and comprehension questions – see Equating for more details
<i>Compréhension du Texte Lu</i> (Reading Comprehension)	Title updated to <i>Compréhension</i> ; Minor Content Modifications	Minor Content Modifications	Minor Content Modifications	
<i>Compréhension à l’Audition</i> (Listening Comprehension)	Content Modifications	(no change)	(no change)	Replaced Listening Comprehension passage and questions between 2010 and 2012
<i>Écriture d’une Phrase Complète</i> (Dictation)	(no change)	(no change)	(no change)	None
Early Grade Mathematics Assessment				
<i>Numération</i> (Rote Counting)	(no change)	(no change)	(no change)	None
<i>Compter: Correspondance Biunivoque</i> (Rational Counting)	(no change)	(no change)	(no change)	None
<i>Identification des Nombres</i> (Number Identification)	Minor Content Modifications	Minor Content Modifications	Minor Content Modifications	Approximately 20% of the items within each subtask were retained unchanged from each version of the EGMA instrument to the next. The remaining items were

	Year of Assessment			Modifications to Content
	2010	2012	2014	
<i>Comparaison des Quantités</i> (Quantity Comparison)	Minor Content Modifications	Minor Content Modifications	Minor Content Modifications	modified in a way that preserved similarity to the original items (i.e., numbers of comparable difficulty were used).
<i>Echelle d'Éstimation des Nombres</i> (Number Estimation)	Dropped		N/A	
<i>Chiffre Manquant</i> (Missing Number)	Visual representation of missing item prompt changed from underscore to empty box; minor content modifications		Minor Content Modifications	Approximately 20% of the items within each subtask were retained unchanged from each version of the EGMA instrument to the next. The remaining items were modified in a way that preserved similarity to the original items (i.e., numbers of comparable difficulty were used).
<i>Problèmes</i> (Word Problems)	Minor Content Modifications	Minor Content Modifications	Minor Content Modifications	Approximately 40% of the items within each subtask were retained unchanged from the 2012 version of the EGMA instrument to the 2014 version.
<i>Exercices de Calcule: Section de l'Addition</i> (Addition)	Minor Content Modifications	Minor Content Modifications	Minor Content Modifications	
<i>Exercices de Calcule: Section de Soustraction</i> (Subtraction)	Minor Content Modifications	Minor Content Modifications	Minor Content Modifications	50% of the items within each subtask were retained unchanged from the prior year's EGMA. The remaining items were modified in a way that preserved similarity to the original items (i.e., numbers of comparable difficulty were used).
<i>Exercices de Calcule: Section de Multiplication</i> (Multiplication)	Minor Content Modifications	Minor Content Modifications	Minor Content Modifications	
<i>Exercices de Calcule: Section de Division</i> (Division)	Minor Content Modifications	Minor Content Modifications	Minor Content Modifications	
<i>Attributs des Figures</i> (Shape Recognition)	Dropped		N/A	
<i>Extension du Schéma</i> (Pattern Extension / Geometric Properties)	Dropped		N/A	

Equating

Student reading performance on an EGRA can be falsely inflated if students are given access to and directed to memorize the story that will be used for the *Oral Reading Fluency* subtask. It has become standard practice to eliminate this potential source of bias by replacing the story used at each administration with a new one that is of equivalent difficulty. While developing multiple passages of exactly the same reading level is extremely challenging, a statistical analysis procedure known as *means equating* can be used to ascertain the relative difficulty of a pair of passages.

Five new oral reading passages were drafted prior to the PAQUED Midterm assessment in 2012, and appropriate reading comprehension questions were developed for each. Note that in the following discussion, the matched pair of a story and its related comprehension questions is called a “dyad.” These five new dyads, along with the original dyad from the baseline assessment, were administered to a sample of 212 Grade 4 and Grade 6 children from six schools near Kisangani. Every child was assessed on all six of the dyads. The order in which the dyads were administered was randomized, mitigating any bias that might otherwise have been introduced by assessment fatigue or knowledge gained from repeating similar tasks.

The students’ results were then analyzed using means equating. This procedure results in a conversion table that allows a given score on one passage to be converted to an equivalent score on another passage. In this instance, the passage selected for the endline is being *equated*, and the passage that was used at baseline is serving as the *reference*.

Conceptually, a factor greater than one would indicate that the equated passage was more difficult than the reference passage. Conversely, a factor less than one would indicate that the equated passage was *less* difficult. Multiplying the score a student earned on the *Oral Reading Fluency* subtask at endline by the appropriate equating factor provides a new *equated score*, which allows for a more accurate understanding of the passages’ relative difficulty than raw scores.

The example in **Table A3.6** illustrates what would happen if two different students—one sampled at baseline and one sampled at endline—were to earn the same scores on both oral reading passages.

Table A3.6: Example of Application of the Means Equating Method

Subtest	Student 1: baseline score	Student 2: Original endline score	Equating multiplier	Student 2: Equated endline score
Oral Reading Fluency	30	30	1.18	35.32

While the raw scores suggest that the students' performance was equal, in reality the endline student was assessed using a slightly more difficult passage. Thus, the equated endline score provides a clearer view of the endline student's performance relative to that of the baseline student. All oral reading fluency scores presented in this report are equated scores unless otherwise mentioned.

Annex IV: Reading Program Schools: 2012-2014 Comparisons

In order to more fully explore the change from 2012 (baseline) to 2014 (endline), EDC requested that independent sample t-tests be used to determine whether the growth over time within the 20 Reading Program Schools that were tested at both points in time was statistically significant. It should be noted, however, that across the other samples within this report, statistically significant growth over time did appear for all treatment conditions (that is, control conditions as well as treatment conditions). Difference-in-differences analyses were conducted in order to determine the degree of growth attributable to the intervention rather than to any other contextual causes. Since no control group exists for these 20 Reading Program Schools against which to compare growth, and because growth in student scores over the two years of intervention is expected, it is impossible to attribute any gains observed in these 20 schools to the treatment intervention.⁸⁵

Table A4.1 shows means and standard errors for 2012 and 2014 scores for Grade 2 students for each of the subtasks. Not surprisingly, students in these schools showed statistically significant growth over time on the subtasks of Vocabulary, Listening Comprehension, and Grapheme Recognition. Interestingly, no significant growth was seen in Initial Sound Identification, despite low scores at baseline.

Table A4.1: Reading Program Schools Change Over Time, Grade 2

Subtask	2012		2014		2012-2014
	Mean	SE	Mean	SE	Difference
Vocabulary	8.21	0.40	9.34	0.20	1.13*
Listening Comprehension	0.66	0.09	1.12	0.08	0.47*
Initial Sound Identification	1.27	0.18	1.47	0.15	0.20
Grapheme Recognition	8.16	0.85	16.02	1.09	7.87*

*p<0.017

Table A4.2 shows the same information for assessments administered to students in Grade 4. Interestingly, only one subtask showed statistically significant gains from baseline to endline in Grade 4: Listening Comprehension. Even though students even at endline did not show mastery of the remaining subtasks, no significant gains were observed over time; in fact, in the subtasks of Familiar Word Reading, Invented Word Reading, and Dictation mean scores decreased over time.

⁸⁵ EDC designed and conducted a study in the Reading Program schools that examined changes in teacher knowledge, attitudes, and practice as a result of the modified intervention, and the impact of those changes on student performance. Please see Louge, N. (2014). *2014 final evaluation report teachers' literacy knowledge, instructional practices, and their students' reading performance in PAQUED supported schools in the Democratic Republic of Congo*. Education Development Center.

Table A4.2: Reading Program Schools Change Over Time, Grade 4

Subtask	2012		2014		2012-2014
	Mean	SE	Mean	SE	Difference
Vocabulary	11.08	0.41	12.07	0.20	0.99
Listening Comprehension	1.20	0.11	1.83	0.10	0.63*
Initial Sound Identification	2.73	0.38	1.91	0.19	-0.82
Grapheme Recognition	21.26	1.46	21.91	1.12	0.65
Familiar Word Reading	8.67	1.15	6.74	0.74	-1.93
Invented Word Reading	6.57	0.89	4.81	0.53	-1.76
Oral Reading Fluency	11.09	1.80	13.74	1.47	2.65
Reading Comprehension	0.32	0.07	0.37	0.06	0.06
Dictation	1.01	0.12	0.96	0.08	-0.06

* statistically significant at $p < 0.017$

Tables A4.3 presents the means and standard errors for 2012 and 2014 for Grade 2 students for each of the mathematics subtasks. Statistically significant growth over time was seen for the Missing Number subtask, although a statistically significant decline was seen for Word Problems. These changes are broadly consonant with changes observed in Accessible Control schools (see *Chapter E*).

Table A4.3: Reading Program Schools (n=20) Change Over Time, Mathematics, Grade 2⁸⁶

Subtask	Group	2012		2014		2012-2014
		Mean	SE	Mean	SE	Difference
Number Identification	Treatment	48.05%	2.40%	52.35%	1.85%	4.30%
Quantity Comparison	Treatment	58.82%	2.04%	60.12%	1.98%	1.30%
Missing Number	Treatment	16.22%	0.93%	23.12%	1.10%	6.90% *
Word Problems	Treatment	42.18%	3.11%	29.48%	1.80%	- 12.70% *

⁸⁶ There were no control schools against which performance in Reading Program schools could be compared. Because the number of schools and student records is so small, this table does not disaggregate results by province.

Subtask	Group	2012		2014		2012-2014
		Mean	SE	Mean	SE	Difference
Calculations: Addition	Treatment	46.91%	2.94%	49.80%	2.64%	2.89%
Calculations: Subtraction	Treatment	31.90%	3.75%	36.61%	2.29%	4.71%
Calculations: Multiplication	Treatment	-	-	-	-	-
Calculations: Division	Treatment	-	-	-	-	-

* statistically significant at $p < 0.017$

Table A4.4 presents the same information for Grade 4 students. Statistically significant performance growth was seen for the Missing Number, Addition, and Multiplication subtasks, but statistically significant performance decline was seen for the Number Identification and Word Problems subtasks. Again, these changes in performance are broadly consonant with those observed in Accessible Control schools discussed in *Chapter F* of this report.

Table A4.4: Reading Program Schools (n=20) Change Over Time, Mathematics, Grade 4⁸⁷

Subtask	Group	2012		2014		2012-2014
		Mean	SE	Mean	SE	Difference
Number Identification	Treatment	70.42%	2.46%	63.65%	1.59%	- 6.77% *
Quantity Comparison	Treatment	62.24%	2.77%	65.09%	1.38%	2.85%
Missing Number	Treatment	30.36%	1.47%	42.78%	1.69%	12.42% *
Word Problems	Treatment	58.40%	2.56%	38.67%	1.93%	- 19.73% *
Calculations: Addition	Treatment	43.84%	2.45%	52.99%	2.01%	9.15% *
Calculations: Subtraction	Treatment	45.28%	3.42%	48.84%	2.52%	3.56%
Calculations: Multiplication	Treatment	23.80%	2.62%	35.52%	2.25%	11.72% *

⁸⁷ There were no control schools against which performance in Reading Program schools could be compared. Because the number of schools and student records is so small, this table does not disaggregate results by province.

Subtask	Group	2012		2014		2012-2014
		Mean	SE	Mean	SE	Difference
Calculations: Division	Treatment	31.44%	2.29%	36.55%	2.36%	5.11%

* significant at $p < 0.017$