



READINESS TO LEARN IN MINORITY FRANCOPHONE COMMUNITIES

REPORT OF FINDINGS FROM THE PRESCHOOL PHASE



SRDC Board of Directors

Richard A. Wagner
Partner, Norton Rose Fulbright LLP

Maria David-Evans
IPAC Immediate Past President and Former Deputy Minister,
Government of Alberta

Robert Flynn
Emeritus professor, School of Psychology, University of Ottawa

John Helliwell
Co-director, program on Social Interactions, Identity and
Well-Being, Canadian Institute for Advanced Research

Suzanne Herbert
Former Deputy Minister, Government of Ontario

Guy Lacroix, Ph. D.
Professor of Economics, Université Laval

Renée F. Lyons, Ph. D.
Bridgepoint Chair in Complex Chronic Disease Research and
TD Scientific Director, University of Toronto

Sharon Manson Singer
Former President, Canadian Policy Research Networks

Jim Mitchell
Founding partner of Sussex Circle

SRDC President and CEO

Jean-Pierre Voyer

Authors

Louise Legault, Ph.D.

Glenn Thompson, Ph.D.

Danielle Patry, Ph.D.

Robyn Carson

Lorraine Lefebvre

Paul Lalonde

The Social Research and Demonstration Corporation (SRDC)

is a non-profit research organization, created specifically to develop, field test, and rigorously evaluate new programs. SRDC's two-part mission is to help policy-makers and practitioners identify policies and programs that improve the well-being of all Canadians, with a special concern for the effects on the disadvantaged, and to raise the standards of evidence that are used in assessing these policies.

Since its establishment in December 1991, SRDC has completed over 200 projects and studies for various federal and provincial departments, municipalities, as well as other public and non-profit organizations. SRDC has offices located in Ottawa, Toronto, and Vancouver.

Readiness to Learn Sponsor

The Readiness to Learn project was funded by Employment and Social Development Canada.



The analysis report in this paper is the responsibility of the authors and does not represent the views of the Government of Canada.

Table of Contents

	Page
Table of Contents.....	i
Tables and Figures.....	v
Executive Summary.....	1
1. Presentation of the Readiness to Learn in Minority Francophone Communities Project	3
1.1. Project Background.....	3
1.2. The Readiness to Learn project.....	5
1.3. The Project's Contribution.....	6
1.4. Development Modeling for Children in Minority Settings	6
1.4.1. The Microsystem: The Family Environment (Contextual Variables)	8
1.4.2. The Microsystem: The Family Environment (Family Processes)	10
1.4.3. The Microsystem: The Childcare Environment	11
1.4.4. The Mesosystem: Relationship Between the Family and the Childcare Environment.....	12
1.4.5. The Macrosystem: Community Variables.....	12
1.5. A Definition of School Readiness.....	13
1.6. About this Report.....	14
2. Implementation Activities	15
2.1. Recruitment of the Communities	15
2.2. Recruitment of the Participating Families.....	16
2.3. Human Capital Required for Implementation.....	16
2.3.1. Champions	16
2.3.2. Community Coordinators	17
2.3.3. Educators.....	17
2.3.4. Family Literacy Practitioners	18
2.3.5. Trainers	18
2.3.6. Evaluators.....	19
2.4. Training.....	19
2.4.1. Basic Training for Educators in the Daycare Program.....	19
2.4.2. Follow-up Training for Educators in the Daycare Program	19
2.4.3. Basic Training for the Family Literacy Program	20
2.4.4. Follow-up Training for the Family Literacy Workshops	20
2.5. Description of the Daycare Component.....	21
2.5.1. Delivery of the Daycare Program.....	22
2.6. Description of the Family Literacy Component.....	23
2.6.1. Harmonization of the Two Program Components.....	24
2.6.2. Delivery of the Family Literacy Program	24
2.6.3. Resource Centre — Purchase of Material	25
3. Methodology.....	27
3.1. Target Population.....	27
3.2. Experimental Design.....	27
3.3. Internal Validity	28
3.3.1. Sample Size.....	28
3.3.2. Targeted Sampling Strategy	28
3.3.3. Pre-Intervention Measures	29
3.3.4. Attrition.....	29
3.3.5. Prevention of Comparison Group Contamination	30

3.4. Sample by Community and by Treatment Group	31
3.4.1. Retention Rate	32
3.5. Measures	33
3.5.1. Data Collection Plan	33
3.5.2. Child Measures	35
3.5.3. Parent Measures	40
3.5.4. Parents' Knowledge, Attitudes and Beliefs.....	44
3.5.5. Dosage.....	46
3.6. Study Hypotheses	46
3.6.1. Impact of the Daycare Component on Children's Language Skills	46
3.6.2. Impact of the Daycare Component on Children's Identity and Culture.....	47
3.6.3. Impact of the Family Literacy Component on Parents' Attitudes and Behaviours	47
4. Methodology for the Implementation Review	49
4.1. Analysis Model for the Qualitative Data of the Implementation Review	49
4.2. Data Collection Tools	50
4.2.1. Implementation Review for the Daycare Component	50
4.2.2. Implementation Review for the Family Literacy Workshops	54
4.3. Method of Data Analysis	57
4.3.1. Analysis Strategies	57
5. Implementation Results for the Daycare Program.....	61
5.1. Analysis Strategy for Daycare Observations	62
5.2. Structural Quality and Differentiation of the Observed Daycares	64
5.3. Structural Elements in the Observed Environments: Fidelity and Differentiation.....	65
5.4. Educational Content: Fidelity, Quality and Differentiation of Activity Content	68
5.5. Quality of Educator–Child Interactions	75
5.6. Dosage	76
5.7. Educator Reactions in the Program Daycare Group	76
5.8. Facilitators and Barriers.....	78
5.9. Perceived Effects	81
5.10. Summary.....	83
6. Implementation Review for the Family Literacy Program	87
6.1. Content Fidelity for the Family Literacy Workshops	88
6.2. quality of family workshop delivery	90
6.3. Dosage	91
6.4. Parental Reactions to Workshops	94
6.5. Facilitators and barriers.....	96
6.6. Perceived Effects	98
6.7. Summary.....	99
7. Preliminary Analyses.....	101
7.1. Quality Control Process	101
7.2. Missing Values analysis.....	102
7.2.1. Breakdown of the Missing Values	102
7.2.2. Pattern of Missing Data.....	104
7.2.3. Data Imputation Strategy	108
7.3. Confounding Variables	110
7.3.1. Associations with Treatment Groups and Child Outcomes.....	110
7.3.2. Unstable Effect of Covariates over Time	111
7.4. Representativity of the Sample: Readiness to Learn project vs. SVOLM (Survey of the Vitality of Official- Language Minorities).....	112
7.4.1. Issue	112
7.4.2. Immigrant Status and Linguistic Profile	113
7.4.3. Sociodemographic Characteristics	116
7.5. Summary of the Implications for Impact Analyses.....	119

8. Approach to Analyses.....	121
8.1. Hierarchical Linear Modeling (HLM)	121
8.2. The Differences-in-Differences method.....	122
8.3. Conceptualization of Treatment and Dosage	123
8.3.1. Heterogeneous Exposure to Treatment	123
8.3.2. Evaluating the Effect of Dosage/Exposure	123
8.3.3. Evaluating the Effect of Daycare Program Fidelity/Quality	125
8.4. Bias Control in a Longitudinal Study	127
8.4.1. Changes in Group Composition	127
8.4.2. Non-parallel Developmental Slopes.....	128
8.4.3. Validity of Effects Based on a Variable Other than Group.....	128
8.4.4. Family Literacy Workshops: A Special Case.....	129
8.4.5. Moderation and Mediation Tests.....	131
8.5. Plan for the Next Chapters	131
9. The Tested Program's Impact on Children	133
9.1. Treatment of Variables	134
9.1.1. Outcome Measures.....	134
9.1.2. Substantive Predictors and Covariates	136
9.1.3. Non-Parametric Cross-Validation	140
9.2. In the First Year	140
9.2.1. Technical Details.....	141
9.2.2. Analyses by Group	142
9.2.3. Analyses by Dosage	150
9.2.4. Analyses by Daycare Program Quality and Fidelity	159
9.2.5. Analyses by Linguistic Profile	169
9.3. In the Second Year.....	173
9.3.1. Technical Details.....	173
9.3.2. Analyses by Group.....	175
9.3.3. Analyses by Dosage	184
9.3.4. Analyses by Quality and Fidelity of the Daycare Program in the First Year	188
9.3.5. Analyses by Linguistic Profile	197
9.4. Summary.....	205
10. Impact of the Family Literacy Component	207
10.1. Direct Impact on Parents.....	207
10.1.1. Variables Retained for the Analysis	207
10.1.2. Analyses by Group	208
10.1.3. Analyses Concerning Parents in the Program Daycare Group.....	216
10.1.4. Summary of the Impact Analyses on Parents.....	225
10.2. Indirect Impact on Children	225
10.2.1. In the First Year	225
10.2.2. In the Second Year	226
10.3. Summary.....	227
11. Discussion.....	229
11.1. Project Implementation.....	229
11.1.1. The Daycare Program	230
11.1.2. The Family Literacy Program	231
11.1.3 Recommendations for Project Implementation in the Future.....	233
11.2. The Tested Program's Impact	233
11.2.1. Analysis of Program Impact by Group.....	234
11.2.2. Daycare Program versus Family Literacy Program	235
11.2.3. The Children's Linguistic Profile.....	236
11.2.4. Limitations of the Findings and Future Research.....	239
11.3. Conclusion	240
References	243

Appendix A: Evaluation Tools and Timetable	255
Appendix B: Content of the Three Versions of the ÉPE-AD	257
Appendix C: Summary of Topics Addressed in the Family Literacy Workshops.....	263
Appendix D: Administration Procedure for the ÉPE-AD (Pre-Intervention Measure)	265
Appendix E: Validation of ÉPE-AD Imputation	267
Appendix F: Result of the Confounding Variables Analysis.....	269
Appendix G: Analysis Results for the ÉVIP-R and EOWPVT-F Scales.....	279
Appendix H: Analysis Results for the ÉPE-AD Scales (Second Year)	281
Appendix I: Results of Impact Analyses by Linguistic Profile.....	283
Appendix J: Analysis Results for Frequency and Language of Literacy Activities	287

Tables and Figures

	Page
Figure 1.1: Bronfenbrenner’s Complete Ecological Model (1979)	8
Table 2.1: Delivery Schedule for the Preschool Program	23
Table 3.1: Participant Breakdown by Community since Enrolment	32
Table 3.2: Participant Breakdown by Treatment Group.....	32
Table 3.3: Reasons for Child Withdrawal from the Readiness to Learn Project	33
Table 3.4: Response Rates for Parent Surveys	34
Table 3.5: Response Rates for Family Literacy Workshop Surveys	34
Table 3.6: Response Rate for Child Assessments	35
Table 3.7: Cronbach Alpha for ÉPE-AD Domains by Assessment Period	37
Table 3.8: Cronbach Alphas for ÉPE-AD Vocabulary Subscales by Assessment Period	38
Table 3.9: Pearson Correlation Coefficients for ÉPE-AD Scales and for Expressive and Receptive Vocabulary Scales	40
Table 3.10: Cronbach Alpha for the Literacy Activities Scale and the Language of Literacy Activities Scale for Each Survey Period	43
Table 4.1: Dummy Results for the Structural Fidelity Matrix.....	58
Table 5.1: Observation Periods by Year.....	62
Table 5.2: Number of Classes in the Project by Treatment Group, Year of Implementation and Cohort	63
Table 5.3: Mean Score for Structural Quality of the Observed Environments by Dimension Assessed and Treatment Group	65
Table 5.4: Average Percentage of Fidelity for Structural Elements by Year of Program Delivery and Treatment Group	68
Table 5.5: Mean Scores for Quality of Educational Content by Dimension Assessed and Treatment Group.....	73
Table 5.6: Average Percentages for Fidelity of Educational Content by Year of Program Delivery and Treatment Group	75
Table 5.7: Dosage (Hours of Children’s Daycare Attendance per Week) by Treatment Group and Cohort.....	76
Table 6.1: Content Coverage by Parent Group.....	89
Table 6.2: Content Coverage by Observed Workshop	90
Figure 6.1: Parental Attendance by Workshop and Site	93
Figure 6.2: Rate of Parental Participation by Workshop	94
Table 7.1: Percentage of Cases Missing at Least One Survey or One Assessment in the First Year of the Study (Four Assessment Periods).....	106
Table 7.2: Children Assessed in English during the First Five Assessments by Group	107
Table 7.3: Distribution of Missing Values and ÉPE-AD Imputations by Treatment Group (Second Cohort).....	109
Table 7.4: Comparison of the Readiness to Learn project and the SVOLM	114
Table 7.5: Comparison of the Readiness to Learn project and the SVOLM —Children Categorized by Mother Tongue.....	114
Table 7.6: Comparison of Mothers in the Readiness to Learn project and the SVOLM — Mothers Categorized by Mother Tongue.....	115
Table 7.7: Comparison of Fathers in the Readiness to Learn project and the SVOLM — Fathers Categorized by Mother Tongue.....	115
Table 7.8: Comparison of the Readiness to Learn project and the SVOLM — Families by Income Bracket	116
Table 7.9: Comparison of Mothers’ Level of Education in the Readiness to Learn project and in the SVOLM	117
Table 7.10: Comparison of Fathers’ Level of Education in the Readiness to Learn project and the SVOLM.....	117
Table 7.11: Comparison of Family Size ¹ in the Readiness to Learn project and the SVOLM.....	118
Table 7.12: Comparison of the Readiness to Learn project and the SVOLM — Number of Children per Respondent.....	118
Table 7.13: Comparison of the Readiness to Learn project and the SVOLM — Number of Single-Parent and Two-Parent Families	119
Table 9.1: Unadjusted Program Impact on School Readiness in the First Year — Standardized Scores.....	146
Table 9.2: Program Impact (after Adjusted for Covariates) on School Readiness in the First Year — Standardized Scores	147

Table 9.3:	Unadjusted Differentiation of the Effect of the Number of Hours Spent at Daycare (Dosage) by Program Received (First Year) — Standardized Scores	153
Table 9.4:	Adjusted Differentiation of the Effect of the Number of Hours Spent at Daycare (Dosage) by Program Received (First Year) — Standardized Scores	154
Table 9.5:	Differentiation of the Tested Program's Quality	160
Table 9.6:	Differentiation of the Tested Program's Fidelity	160
Table 9.7:	Inter-correlations of the Daycare Program's Quality and Fidelity Indices.....	161
Table 9.8:	Effect of Daycare Program Quality with Adjustment (First Year) — Standardized Scores	165
Table 9.9:	Effect of Daycare Program Fidelity with Adjustment (First Year) – Standardized Scores.....	166
Figure 9.1:	Diagram Illustrating the Two Aspects of the Program Effect: The Effect Attributable to Quality and Fidelity of the Daycare Program' (a, b) and That not Attributable to Quality and Fidelity (c')....	167
Table 9.10:	Mediation Test Using Quality as the Mediator of Program Effect at 12 Months.....	168
Table 9.11:	Mediation Test Using Fidelity as the Mediator of Program Effect at 12 Months	169
Table 9.12:	The Moderating Effect of Household Type on School Readiness Indicators.....	170
Table 9.13:	Impact of the Tested Program by Household Type: With High or Low Exposure to French — (First Year) — Adjusted Scores	171
Figure 9.2:	Program Effect (DinD) on Communication at Each Assessment.....	176
Figure 9.3:	Program Effect (DinD) on Self-Awareness at Each Assessment	178
Figure 9.4:	Program Effect (DinD) on Cognition at Each Assessment	179
Figure 9.5:	Program Effect (DinD) on Expressive Vocabulary at Each Assessment	181
Figure 9.6:	Confidence Intervals for the Impact of the Standardized Program (ANCOVA) on French Vocabulary: Standardized Receptive Vocabulary (ÉVIP-R) and Expressive Vocabulary (EOWPVT) Scales	183
Table 9.14:	Number of Hours Spent at Daycare in the First Year (Dosage) and Its Unadjusted Effect on the Development of School Readiness in the Second Year for the Two Daycare Groups — Standardized Scores	185
Table 9.15:	Number of Hours Spent at Daycare in the First Year (Dosage) and Its Adjusted Effect on the Development of School Readiness in the Second Year for the Two Daycare Groups — Standardized Scores	186
Table 9.16:	Impact of Daycare Dosage on French Vocabulary: Standardized Receptive Vocabulary (ÉVIP-R) and Expressive Vocabulary (EOWPVT) Scales	187
Table 9.17:	Quality of the Daycare Program in the First Year and Its Adjusted Effect on the Development of School Readiness in the Second Year — Standardized Scores.....	192
Table 9.18:	Fidelity of the Daycare Program in the First Year and Its Adjusted Effect on the Development of School Readiness in the Second Year — Standardized Scores.....	194
Table 9.20:	Mediation Test Using Fidelity as Mediator of the Program Effect at 24 Months	196
Table 9.21:	Mediation Test Using Quality as Mediator of the Daycare Program Effect at 24 Months.....	197
Figure 9.7:	Confidence Intervals for the Program's Standardized Effect (DinD) on Communication by Household Type	199
Figure 9.8:	Confidence Intervals for the Program's Standardized Effect (DinD) on Self-Awareness by Household Type	200
Figure 9.9:	Confidence Intervals for the Program's Standardized Effect (DinD) on Cognition by Household Type	201
Figure 9.10:	Confidence Intervals for the Program's Standardized Effect (DinD) on Expressive Vocabulary by Household Type	203
Figure 9.11:	Confidence Intervals for the Program's Standardized Effect (ANCOVA) on the ÉVIP-R and the EOWPVT-F by Household Type	204
Figure 10.1:	Distribution of Frequency of Literacy Activities at the Baseline	209
Figure 10.2:	Distribution of Frequency of Literacy Activities at 12 Months	209
Figure 10.3:	Distribution of Frequency of Literacy Activities at 24 Months	210
Figure 10.4:	Effect of the Family Literacy Program on Frequency of Literacy Activities	211
Table 10.1:	Odds Ratio for the Standardized DinD Effects of the Adjusted Model for Frequency of Literacy Activities	212
Figure 10.5:	Distribution of Language of Literacy Activities at the Baseline	213
Figure 10.6:	Distribution of Language of Literacy Activities at 12 months.....	214
Figure 10.7:	Distribution of Language of Literacy Activities at 24 Months.....	214

Figure 10.8: Effect of the Tested Program on Language of Literacy Activities	215
Table 10.2: Odds Ratio for the Standardized DinD Effects of the Adjusted Model for Language of Literacy Activities	216
Table 10.3: Summary of the Family Literacy Program's Effect on Parents' Opinions: The Parent as a Child's First Educator	217
Table 10.4: Summary of the Family Literacy Program's Effects on Parents' Opinions: The Importance of Fostering a Francophone Environment in the Home.....	218
Table 10.5: Summary of the Family Literacy Program's Effects on Parents' Opinions: The Parent and Educator Working Together to Prepare the Child for School	218
Table 10.6: Summary of the Family Literacy Program's Effects on Modeling (Two Weeks Post-Program).....	219
Table 10.7: Summary of the Family Literacy Program's Effects on Modeling (Eight Months Post-Program)	219
Table 10.8: Summary of the Family Literacy Program's Effects on Parents' Self-Efficacy.....	220
Table 10.9: Summary of the Family Literacy Program's Effects on Parents' Knowledge.....	220
Table 10.10: Summary of the Family Literacy Program's Effects on Parents' Sense of Belonging to the Francophone Community	221
Table 10.11: Correlations Between Dosage and Three Questions Eliciting Parents' Opinion.....	222
Table 10.12: Correlations Between Dosage and Modeling, Self-Efficacy, Knowledge and Sense of Belonging.....	222
Table 10.13: Correlations Between Quality of Program Delivery and Three Questions Eliciting Parents' Opinion.....	224
Table 10.14: Correlations Between Quality of Program Delivery and Modeling, Self-Efficacy, Knowledge and Sense of Belonging.....	224
Table 11.1: Summary of Results for the Implementation Study	232
Table 11.2: Summary of Results for the Tested Program's Impact Analyses	234
Table A.1: Timetable for Child Assessments.....	255
Table B.1: Expressive Vocabulary Subscale.....	258
Table B.2: Self-Awareness Subscale	259
Table B.3: Phonological Awareness Subscale	259
Table B.4: Numeracy Subscale	260
Table B.5: Memory of Personal Information Subscale.....	260
Table B.6: Alphabetic Knowledge Subscale.....	261
Table B.7: Oral Reading Subscale	261
Table D.1: Decisional Tree for Determining the Language of Test Administration	266
Table E.1: Means and Standard Deviations for Original and Imputed Scores as Well as Their Inter-correlation.....	268
Table F.1: Relevance of Demographic and Family Composition Variables	270
Table F.2: Relevance of Socioeconomic Variables	271
Table F.3: Relevance of Linguistic Variables	272
Table F.4: Relevance of Sociolinguistic Variables.....	273
Table F.5: Relevance of Parenting Variables	274
Table F.6: Relevance of Daycare Environment Characteristics	275
Table F.7: Relevance of Methodological Factors.....	276
Table F.8: Tests Robust to the Instability of the Confounding Variables' Effects over Time.....	277
Table G.1: Program Impact on French Vocabulary: Standardized Receptive Vocabulary (ÉVIP-R) and Expressive Vocabulary (EOWPVT-F) Scales.....	279
Table H.1: Unadjusted Program Impact on School Readiness in the Second Year – Standardized Scores	281
Table H.2: Adjusted Program Impact on School Readiness in the Second Year – Standardized Scores	282
Table I.1: Tests of the Hypothesis that the Household Linguistic Profile Serves as the Moderator of Program Effect in the Second Year.....	283
Table I.2: Impact of the Tested Program by Household Linguistic Profile – (Year Two) – Adjusted Scores.....	284
Table I.3: Impact of the Tested Program on Vocabulary by Linguistic Profile: Standardized Receptive Vocabulary (ÉVIP-R) and Expressive Vocabulary (EOWPVT-F) Scales	285
Table J.1: Impact of the Tested Program on Frequency of Literacy Activities as well as Language of Literacy Activities (Unadjusted)	287
Table J.2: Adjusted Impact of the Tested Program on Frequency of Literacy Activities and Language of Literacy Activities	288

Executive Summary

The purpose of this report is to document the implementation and impact of the Readiness to Learn in Minority Francophone Communities project (short title: Readiness to Learn project; previously the Child Care Pilot Project, or CCPP), a demonstration project funded by Human Resources and Skills Development Canada (HRSDC). The Social Research and Demonstration Corporation (SRDC) was hired to implement, manage, collect and analyze the project data. The project tests a preschool program¹ that combines a childcare component developed specifically to meet the needs of Francophone children in minority settings with a family literacy component targeting the parents of these children. This program aims to develop the child's language skills, knowledge and use of French, knowledge of and engagement in Francophone culture, as well as to foster the child's school readiness and general development.

The program is evaluated using a quasi-experimental methodology with non-equivalent control groups. The methodology involves three treatment groups: the Program Daycare group, consisting of children enrolled in a French-language daycare that offers the new preschool program; the Comparison Daycare group, consisting of children enrolled in a French-language daycare that does not offer the new program; and the Informal Care group, consisting of children whose daytime care is provided at home or at an unregulated family daycare. The purpose of the Comparison Daycare group is to take into account how formal daycare affects child development, which is a subject unto itself. The purpose of the Informal Care group is to factor in how an informal childcare environment affects child development. The project has two cohorts of participants, the first recruited in 2007 and the second, in 2008.

This report focuses on the data collected prior to the intervention and over the two years of program delivery, that is, from May 2007 to October 2009 for the first cohort and from October 2008 to October 2010 for the second cohort. This period corresponds to the time when the children ranged from ages three to five, on average. A mixed research methodology was used for the Readiness to Learn project. This approach favours the use of a range of tools, both quantitative and qualitative, from several sources selected based on research objectives, that is, to determine whether the program has had the desired effect and to understand how it exerts that effect. Analyses were thus conducted using data from, among other sources, child assessments, parent surveys, observations at daycares and Family Literacy Workshops as well as administrative data (for example, children's attendance record at the daycare).

We began by studying project implementation through direct observations at sites where the project was taking place. An analysis of these data shows that the program generally received by children and parents resembled that established by program developers. The fidelity of the Daycare Program and the Family Literacy Program was very good. In fact, the quality of the Daycare Program was greater than that of typical daycare programs in minority language settings. Educators, trainers and parents perceived the tested program as favourable to the children's school readiness.

¹ Officially referred to as *enriched childcare services* in HRSDC documents, the Social Research and Demonstration Corporation (SRDC), in agreement with HRSDC, now refers to the program as the "preschool childcare program."

The results of quantitative analyses show that: (a) the tested program has a positive effect on children's ability to communicate in French as well as on their vocabulary development, (b) the superior fidelity/quality of the tested Daycare Program accounts very well for this effect (results from the mediation analyses), (c) the change in parental attitudes and behaviours does not appear to be associated with the children's observed progress, (d) the greater a child's exposure to the program during delivery, the greater the program's effect on that child's development, and (e) as expected, exposure to the French language partially determines the nature of the program's effect: children with high exposure when the study began benefited more in terms of cognitive development while children with low exposure appear to have benefited more in terms of language-skill development. Two key conclusions stem from these findings. First, the Daycare Program seems to be the primary driver of the program's effect. What's more, all children appear to benefit from the tested program, although the nature of the observed effect seems to depend to a certain degree on the linguistic profile of the child and his or her family. The picture that emerges from all the analyses allows us to conclude that the tested program has a modest impact on school readiness as well as on the language development of children in the Program Daycare group. In terms of social policy, these research findings speak to the priorities identified by the federal government in the *Roadmap for Canada's Linguistic Duality 2008–2013*.

This report is one in a series of reports by SRDC. It follows the *Readiness to Learn in Minority Francophone Communities: Reference Report* (Legault, Mák, Verstraete, & Bérubé, 2014), the final version of which was submitted to HRSDC on October 13, 2009. This first report sketched a profile of the children, their family and the communities participating in the Readiness to Learn project. Moreover, this report supplements the *Readiness to Learn in Minority Francophone Communities: Project Implementation Report* (Bérubé, Legault, Janisse, Carson, Saucier, & Lefebvre, 2014), submitted to HRSDC on May 31, 2010, and the *Readiness to Learn in Minority Francophone Communities: First Cohort Findings Report* (Thompson, Legault, Lalonde, & Bérubé, 2014) submitted to HRSDC on July 31, 2010. Two other reports are planned: one concerning the program's effects one year after the end of the intervention and the other, its effects two years after the end of the intervention.

1. Presentation of the Readiness to Learn in Minority Francophone Communities Project

1.1. PROJECT BACKGROUND

The Readiness to Learn in Minority Francophone Communities project (Readiness to Learn project) is part of the Government of Canada's *Action Plan for Official Languages* 2003–2008 and continues under the *Roadmap for Canada's Linguistic Duality* 2008–2013. The project's guiding principles include a desire to help minority Francophone communities give children a good start in life, as well as to encourage parents to participate actively in their child's education (Human Resources and Skills Development Canada, or HRSDC, 2006). At the community level, the Readiness to Learn project is intended to be a rigorous assessment of a promising intervention that aims to maintain, and even breathe new life into, the ethnolinguistic vitality of the minority Francophone community.²

This project recognizes that developing language and identity is more important for linguistic minorities than majorities. While members of the linguistic majority may take this developmental process for granted, the true complexity of the process stands out in linguistic minority situations. Development of language and identity is the outcome of a socialization process experienced in the family environment, the school or preschool environment, and the socio-institutional environment (Landry and Allard, 1997). Vygotsky (1978) conducted pioneering studies on the importance of culture in child development. The culture in which a child grows up affects the development of his or her language skills and learning in general through integration of the social symbols to which he or she is exposed. Thus, the child's social environment is inseparable from the construction of his or her cultural and linguistic identity, as well as from his or her general development.

In the public sphere, the reality of a minority context means that young Francophones are exposed to two different cultures at a time when their identity is taking shape. According to Gilbert (2003), exposure to French in all social contexts is even more important for a child living in a highly minority Francophone setting where, simply by virtue of demographic weight, the English language predominates in every aspect of day-to-day life. In terms of childhood development, several authors advocate the availability of French-language childcare services and education as key vectors of community vitality (Commission nationale des parents Francophones, 2005; Landry and Allard, 1997; Gilbert, 2003). Ideally, say champions of Francophone community vitality, eligible parents would enrol their children in quality childcare services and at French-language schools. However, the reality is otherwise.

The children of eligible parents are often enrolled in English-language schools. The findings of the 2006 Survey on the Vitality of Official-Language Minorities (SVOLM; Corbeil, Grenier, and Lafrenière, 2007) indicate that only 56% of children with eligible parents attend French-language primary schools. This proportion drops to 44% among teens. Thus, a significant percentage of eligible parents choose to enrol their children in immersion or English-language schools for a number of reasons, including: English is the child's mother tongue or the language

² See Guimond (2003) for an overview of studies on ethnolinguistic vitality in minority settings.

he or she knows best, school proximity, non-availability of a French-language school, and the quality of the program or school.

Several studies show that young Francophone children enrolled in French-language schools score lower in literacy and numeracy than do children in majority linguistic groups. This disparity between the groups can be observed, among other ways, in the results of international testing such as the Program for International Student Assessment (PISA), in which Francophone children attending French-language school in a linguistic minority setting score lower in reading than their English-speaking Canadian peers (Bussière, et al., 2001; Canadian Council on Learning, 2008). The few studies on young minority Francophones indicate that this achievement gap appears at a young age. A recent study of Franco-Manitoban children ages four to six concludes that they score lower on vocabulary tests (Peabody Picture Vocabulary Test – Revised, or PPVT-R, and the communication and general knowledge scales of the Early Development Instrument, or EDI), and this is particularly true of children who live in a majority Anglophone environment every day. The trend continues when these children reach Grade 3 of primary school. Children who grow up in a Francophone family and preschool environment score higher in Grade 3 reading than Francophone children living in a majoritarily Anglophone linguistic environment (Chartier, Dumaine, and Sabourin, 2011).

Limited exposure to French at home, which limits development of the child's French-language skills, is likely the source of these difficulties. In fact, almost two-thirds of young minority Francophones are from exogamous households (67%) and most adopt English as the household language (Landry, 2010).³ Only 20% of exogamous couples choose to raise their children ages zero to four in French (Martel, 2001). Further, the latest data from the 2006 Census indicate that almost 39% of Francophones outside Quebec speak English instead at home, although French remains a language that is used (Corbeil and Blaser, 2007).

This greater use of English in daily life partly explains why 62% of Francophone adults outside Quebec who take a French (instead of English) literacy test score below the level of literacy proficiency deemed necessary to function in society (i.e., a literacy level of over 3 on a scale of 5; Statistics Canada and HRSDC, 2005, Table 3.24). This proportion would no doubt be higher if all Francophone adults outside Quebec took the test in French (65% of them chose to take the English test, despite identifying French as their mother tongue; Statistics Canada and HRSDC, 2005, p. 54). According to these results, there is every reason to shore up language acquisition among minority populations and encourage parental engagement in preserving the Francophone community.

For young minority Francophones, this series of findings is worrisome when we consider that language skills remain the *keystone* of academic achievement, which in turn affects occupational and social success. Early intervention with young minority Francophones in both the family and preschool environments may positively influence their linguistic trajectory, thereby fostering academic achievement and integration within the Francophone community. This intervention should include, among other aspects, a component targeting parents in order to raise their awareness about the challenges of living in a minority environment and the actions they can take to transmit this rich cultural heritage to their children. The purpose of the Readiness to Learn project is to evaluate the impact of such an intervention.

³ Landry, R. (2010). *Petite enfance et autonomie culturelle : Là où le nombre le justifie... V*. Research report for the Commission nationale des parents francophones. Canadian Institute for Research on Linguistic Minorities. Moncton, New Brunswick.

1.2. THE READINESS TO LEARN PROJECT

The Readiness to Learn project tests a preschool childcare program and aims to identify the benefits of that preschool program to children living in a minority Francophone environment. The tested preschool program combines a childcare component developed specifically to meet the needs of Francophone children in minority settings with a family literacy component targeting the parents of these children. This family literacy component seeks to encourage the parents' active participation in their child's development and school readiness, as well as in the transmission of French language and culture. The project itself is one of many studies on preschooler development and on the vitality of the French language in minority settings.

The program's effect on child development is examined by comparing a group of participants who were exposed to the new program (referred to as the Program Daycare group) to comparison groups consisting of participants who *were not* exposed to the new program (referred to as the Comparison Daycare group and the Informal Care group). A Research Advisory Committee, whose members are academics specializing in Francophone early childhood development and representatives from Francophone communities, was created in order to help design, implement, monitor and evaluate the pilot project. The Social Research and Demonstration Corporation manages, implements and evaluates the program, in addition to actively participating in the project's development.

The program was delivered to two cohorts of participants. The first cohort began the program in fall 2007 in six minority Francophone communities (Saint John and Edmundston in New Brunswick; Orleans, Cornwall and Durham in Ontario; and Edmonton in Alberta). Program delivery to the second cohort began in fall 2008 in two communities (Orleans and Cornwall in Ontario). The children are followed from age three to age seven, when they begin Grade 2 of primary school. The length of the study allows us to monitor the development of young minority Francophones from preschool until their education commences. The last data will be collected in fall 2011 for first-cohort families and in fall 2012 for second-cohort families.

In its first phase, the Readiness to Learn project aims to answer the following research question: *Does the new preschool program, consisting of a daycare component and a parent/child workshop component, have a significant impact on children's language skills, Francophone cultural identity and school readiness beyond the development that would occur in the absence of this program, and independently of other external factors that may come into play?* Related issues are also investigated, including: *For whom is this program most beneficial? Is this program profitable? Can the new program be repeated?* In its second phase, the Readiness to Learn project addresses a new research question: *Does the new preschool program better equip Francophone children in minority environments to succeed in reading and mathematics, tasks essential to academic achievement?* While the first research question concentrates on the preschool period, the second research question focuses on the education period encompassing Grades 1 and 2, when the children are ages six and seven. This report concerns the first phase of the project.

1.3. THE PROJECT'S CONTRIBUTION

The project's contribution is that it takes into account the unique environment of minority Francophone communities by including a francization component, a cultural identity component and a parental awareness and engagement component. The project focuses on young Francophone children and is based on Bronfenbrenner's ecological model (1979), one premise of which is that children are influenced by all the environments that surround them. The tested program is designed to strengthen the ties between the child's main environments—where actions in the daycare setting support those in the family setting, and vice versa—so as to optimize the child's learning with respect to his or her general development and school readiness as much as with respect to his or her French language and cultural development.

The many benefits of programs that modify both the child's daycare and home environments were established in several studies on other “vulnerable” populations (see the literature review by Reese, Sparks, and Leyva, 2010; Engle, et al., 2007). It is believed that maximization of these program effects depends on parents and educators adopting the same approaches with the child. This was shown in a study by Pelletier and Corter (2005). Its finding showed that a dual-component (parent and educator) program based on early literacy activities fostered greater child learning in this area than a similar program consisting of only one of these components. Children whose daycare and home environments had changed were more advanced in vocabulary development, early reading and numeracy.

In short, the entire preschool program is based on an integrative perspective and uses a range of approaches to influence the main vectors of child development. The sections that follow begin by presenting the factors that influence early childhood development while factoring in the specificities of living in a minority Francophone environment, then concentrate on defining school readiness. The chapter ends with a section detailing the objectives and content of this report.

1.4. DEVELOPMENT MODELING FOR CHILDREN IN MINORITY SETTINGS

The Readiness to Learn project is derived from an ecological framework in which factors at various levels influence child development. This vision is clearly represented in Bronfenbrenner's ecological model (1979). This researcher was the first to express in words and images the entire system of influences that shapes child development. The model is based on three premises:

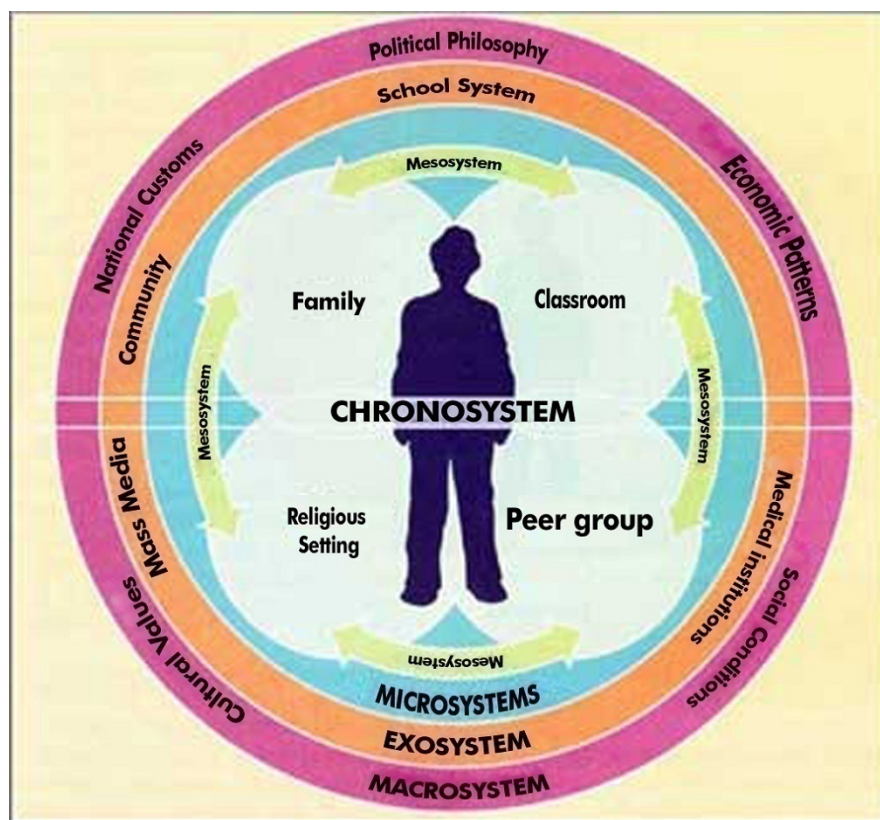
- a) the child is at the centre of the model;
- b) the central role of the child's experiences (which are considered the “drivers” of development); and
- c) the nature of the relationships among the child's various environments.

Bronfenbrenner's model consists of five systems (Figure 1.1):

- **Microsystem:** Immediate environment (family, school, type of childcare, peers, neighbourhood).
- **Mesosystem:** Interactions among the immediate environments (e.g., between home and school).
- **Exosystem:** External environment that indirectly affects the child (such as the parents' work).
- **Macrosystem:** Broader cultural context (western culture versus eastern, national economy, political culture, subculture).
- **Chronosystem:** Structure of the events affecting the environment and life transitions.

In the context of the Readiness to Learn project where the focus is on young children in minority language settings, three systems of the Bronfenbrenner's model are of particular importance. First, the microsystem affects the child's development through the characteristics of his or her family and childcare environment. In terms of family characteristics, we distinguish between contextual variables and family processes (this division is based on the National Longitudinal Survey of Children and Youth, or NLSCY; Statistics Canada and HRSDC, 2006). Contextual variables refer to "factual" data known to influence child development (e.g., family composition). Next, the mesosystem, such as ties between the family and childcare environments, also plays a role in preschooler development. Lastly, the macrosystem is among the influences of interest for the project because it consists of the community in which the child is raised and its linguistic characteristics in particular.

Figure 1.1: Bronfenbrenner's Complete Ecological Model (1979)



Source: In *UW-Extension ABC Project*, Appendix B (November 2004).

1.4.1. The Microsystem: The Family Environment (Contextual Variables)

Child development is influenced by several factors intrinsic to the child, such as fetal history, birth weight, and length of gestation. In addition to the child's inherent characteristics, over the years research has identified a series of environmental factors that may affect his or her development. The contextual variables of the family environment are among the factors that contribute most to child development (Sanders and Morawska, 2006). These include family composition, income, the parents' level of education, and languages used at home.

Family Composition

In the context of the Readiness to Learn project, we examine the family variables that affect child development, particularly those that influence a child's linguistic trajectory. Thus, birth order is an important variable. In fact, studies show that first-born children have a broader vocabulary than do their siblings (Tamis-LeMonda and Rodriguez, 2008).

Family type (intact two-parent, single-parent, blended, etc.) is also a variable that must be considered. A study conducted using longitudinal data from the NLSCY showed that family type (two-parent or single-parent) has a direct influence on several aspects of child development. Among children ages six to eleven, the authors conclude that "[t]wo-parent households [...] had children who were less hyperactive, more academically skilled, less anxious or depressed, and (judged by teachers) good in academic standing." (Adams and Ryan, 2000, p. iii). However, it is

not family composition so much as factors associated with single-parenting, such as maternal stress or a drop in family income, that affect the child *a priori*, hence the importance of carefully interpreting the results.

Gross Family Income

Hundreds of studies have already established a relationship between family poverty and a child's health, school readiness, academic achievement and behaviour. However, few of these studies have examined the effects of the timing, length and degree of poverty. Insufficient family income affects child development in numerous ways: poor nutrition, fewer learning situations, instability of the place of residence, attendance at schools with fewer resources, family violence, etc. (see Duncan and Brooks-Gunn, 1997). Some studies, like the one by Berger and colleagues (2005), have confirmed the hypothesis that a low income influences child development outcomes through its multiple effects on the home environment. However, although the causal relationship between a low income and children's intellectual and behavioural outcomes is quite clear, interpretation of these associations remains debatable, as well as the policy implications.⁴ Recent studies on the subject tend to identify two primary ways in which low income can affect a child: the physical environment and the quality of parenting (Berger, Paxson, and Waldfogel, 2005).⁵ In other words, poverty will affect the purchase of the family's material resources, in addition to its stress level, which in turn influence child development. Several studies based on national data have shown that the level of cognitive stimulation in the home environment (measured by learning material and the parents' approach to learning) accounts for 33% to 50% of the association between income and various outcomes in terms of the child's cognitive and language development (Dearing, Berry, and Zaslow, 2006).

Studies also tend to demonstrate that the negative effects of poverty on a child's cognitive development and academic achievement are greater during the preschool period than at any other time (Dearing, et al., 2006). This datum must therefore be taken into account in evaluating the program's effect as part of the Readiness to Learn project.

Parents' Level of Education

As with family income, the parents' level of education is an important factor in a child's success (see, among others, Haveman and Wolfe, 1995). However, exactly how parental education influences child development has been less well studied. Klebanov and colleagues (1994) have shown that a mother's education and family income are important factors when it comes to the existence of a physical environment conducive to learning, but that education alone is an important factor in "warm" parenting. A series of studies by Davis-Kean (2005) concludes that parental education affects child development not only through the parents' social success, but also through their beliefs and their behaviours toward their child. As such, level of education has an importance to child development that extends beyond the socioeconomic context with which it is often associated.

⁴ For example, direct monetary transfers to families would be enough if there was a clear causal relationship between income, quality of the environment, and children's outcomes (Berger, et al., 2005).

⁵ The first theory was initially developed in economic literature (see Becker, 1993) and the second, in developmental psychology literature (e.g., Dearing, et al., 2006).

Mother's Age at the First Child's Birth

The mother's age at birth of the first child is another factor with numerous effects on child development. Studies show that a mother's age when her first child is born affects the development of that child and of subsequent children, even after parenting styles and family functioning are accounted for (Tremblay, et al., 2004). In the case at hand, note that very young mothers provide their child with less complex language stimulation. They use a more limited range of vocabulary and verbally stimulate their children less (Tamis-LeMonda and Rodriguez, 2008). This behaviour is passed on to the child, who is then at risk of experiencing language development that is insufficient to ensure a successful start at school.

Language(s) Spoken at Home

In minority contexts, the languages most often spoken at home by the parents of the targeted child are related to language transmission and to the vitality of French at home (Forgues and Landry, 2006). Thus, the family linguistic environment is associated with a child's school readiness and academic achievement (Chartier, et al., 2011). In order to find this information, one must determine which languages are usually spoken in the child's environment, particularly by his or her parents and older siblings both at home and outside the home. It is also interesting to know which languages the child usually speaks. This linguistic behaviour is a concrete indication of the languages in which the child feels most comfortable expressing himself or herself. All of this information allows us to sketch an overall picture of the linguistic influences to which the child is exposed.

1.4.2. The Microsystem: The Family Environment (Family Processes)

In addition to the descriptive characteristics of families, family processes have a definite influence on child development. The main elements that affect child development include family functioning, parenting style, and the family's literacy activities.

Family Functioning

Beyond the mother's psychosocial difficulties, the household's general environment is also an element that researchers wish to link to child development. Family functioning is defined as the quality of relationships within the family, in terms of the quality of communication, harmony among family members, and the support available within the family. From a social perspective, poor family functioning is widely associated with aggressive developmental trajectories among children (Tremblay, et al., 2004). Where language development is concerned, a child's vocabulary acquisition is positively related to the family's level of functioning, that is, the level of harmony and communication among family members (Desrosiers and Ducharme, 2006). Family functioning thus has many significant repercussions on children's life paths.

Parenting Style

As their child's first educator, parents play a key role in their child's development and functioning. Parenting style affects the child's social, intellectual, moral and emotional development (Bornstein and Bornstein, 2007). Parenting styles can be broken down into two aspects: *sensitivity*, which measures the parent's attentiveness to the child and ability to respond to his or her needs and interests; and *control* (or *strictness*), which refers to the level of supervision and discipline, as well as the extent to which the parent requires that the child obey

and show self-control (Canadian Council on Learning, 2007). Studies in this regard indicate that children have better language skills and score higher on intelligence tests when their parents are more encouraging and less controlling (Sanders and Morawska, 2006). It is therefore important to measure both aspects when monitoring a child's language and cognitive development.

Literacy Activities

Experiences in early childhood affect children's language skills when they begin school (Doherty, 1997). According to Desrosiers and Ducharme (2006), children whose parents read to them regularly are less likely to present a delay in vocabulary. Reading at home at a young age (before age three-and-a-half) is even associated with improved verbal abilities among children who showed language difficulties. Similarly, parental participation in home-based learning activities is a predictor of children's long-term social and academic adaptation (Izzo, Weissberg, Kaspro, and Fendrich, 1999). Of particular interest for the Readiness to Learn project, exposure to oral and written French in preschool is crucial for the child to develop strong linguistic skills. Activities like story time and borrowing books from the library convey the unique elements of Francophone culture to children (Salerno in Lafrance, 1993).

1.4.3. The Microsystem: The Childcare Environment

The characteristics of the childcare environment influence numerous facets of a child's development. Several years ago, psychologists and educators concluded, based on observations and experience, that non-parental care affects children's cognitive and language development (National Institute of Child Health and Human Development Early Childcare Research Network and Duncan, 2003). More specifically, the quality of a childcare environment affects children's cognitive and language development, school readiness, and behaviour (Cleveland, et al., 2006). Moreover, this impact varies greatly based on certain family factors. For example, the effects of a quality childcare environment are more significant among children living in a less fortunate socioeconomic context (Burchinal, et al., 2000), for which linguistic minority status is sometimes considered an example (Maltais, 2007).

Daycare can be a place of learning that leads to better academic skills. For example, in U.S. studies, children who went to daycare scored higher in reading and mathematics when they began senior kindergarten (see, among others, Howes, et al., 2008). Appropriate material, including quality games and books, an appropriate physical environment and affectionate educators who support child development may be very beneficial for all children, especially those in a difficult family situation. Quality of childcare services can take two forms: structural quality, consisting of factors that can be modified through legislation (educators' qualifications, operating hours, group size, etc.), and process quality, referring to the child's experiences in his or her daycare environment (quality of the available activities and of the relationship with the educator) (Burchinal, et al., 2000). For the purposes of the Readiness to Learn project, this information was gathered through observations. It allows us to compare the daycare component's tested program to the activities offered at comparison daycares. This is how we will determine whether the tested program has truly changed the daycare environment.

In a minority Francophone context, the linguistic aspect is particularly important in studying the impact of a daycare program. Exposing a preschooler to a French-language childcare environment has a positive impact on his or her academic achievement. Chartier and colleagues

(2011) proved this statement in a study using longitudinal data from 217 children in a Manitoba Francophone community.⁶ Children who were exposed to French in both their family environment and their childcare environment scored higher on the ÉVIP-R (which measures receptive vocabulary) and on EDI communication and general knowledge tests than children who were exposed to French only at home.

1.4.4. The Mesosystem: Relationship Between the Family and the Childcare Environment

The mesosystem consists of the ties among the various systems with which the child is in contact. These ties make a unique contribution to child development. Several studies support the idea that the relationship between the school environment and the family contributes to child development. Children whose parents are involved in their school work are better adapted socially and academically than other children. They also have a more positive attitude toward school and display higher aspirations for their future, regardless of family income and their parents' level of education (Connors and Epstein, 1995). The same parallel has been established between parental involvement in the preschool environment and children's early literacy skills. Parents who get involved by talking with the educator, asking her questions about their child's day and participating in daycare activities have children with broader vocabulary, more phonological awareness and better early writing skills (Arnold, Zeljo, Doctoroff, and Ortiz, 2008). The added value of a program that combines daycare services and Family Literacy Workshops highlights the importance of close collaboration between parents and educators, and of adopting common and complementary approaches (Corter and Pelletier, 2005).

1.4.5. The Macrosystem: Community Variables

A growing number of authors recognize the important influence of community characteristics on child development (Hertzman and Kohen, 2003; Moore, 2005). The community environment fosters children's cognitive development as well as their physical and emotional health (Willms, 2007). One key factor affecting child development is the availability of French-only resources and services for families.

Resource Availability and Use

The availability and use of a community's resources for its young families are important factors for child development. According to Connor and Brink (1999), certain categories of community resources are particularly important for child development, particularly the education and health system, entertainment and culture, societal programs, special needs programs, and sports and recreation. In the context of a Francophone minority, French-language early childhood services and resources are seen more as a protective element for Francophone identity and preservation of the French language (Commission nationale des parents Francophones, 2005). The notion of a protective element is based on Breton's (1964) construct of *institutional completeness*. At the extreme end of the continuum, a community with institutional completeness would enable its Francophone population to conduct all daily activities in French. It follows that the presence of French-language institutions in a community encourages the creation of social networks and greater social cohesion within the community. Landry (1994) used this concept of

⁶The study is called the *Tots Study*, also referred to as the *1997 Birth Cohort Study*.

a complete institutional infrastructure in his work on the four types of capital—demographic, political, economic and cultural—with a tremendous influence on the development, preservation, and even renewal of a community’s ethnolinguistic vitality.⁷ It is through this ethnolinguistic vitality that minority communities preserve their sense of belonging and pride, thereby supporting an integration, rather than assimilation, of Francophone language and culture into the majority community. One of these forms of capital, cultural capital, “...refers to the resources and information that act as agents in passing on culture” (Landry, 1994, p. 18, quoted in Guimond, 2003). The author postulates that this cultural capital can be assessed in part through the diversity of educational institutions and access to cultural resources in the community. Landry, Allard, and Deveau (2007b) state that French-language schools—especially the existence of a school system that enables young Francophones to study in French from preschool to the post-secondary level—play a crucial role in developing, preserving, and enhancing institutional completeness. Within the preschool realm, formal daycare centres, junior kindergarten, family daycare services, extracurricular programs, resource centres and play groups pave the way to French-language schooling (Gilbert, 2003).

1.5. A DEFINITION OF SCHOOL READINESS

As all school-age children must, young Francophones in minority environments must prepare for the start of school. This transition is important, as studies in the field indicate. In fact, there is a strong association between a child’s school readiness and his or her academic achievement (Lemelin and Boivin, 2007). Some indicators can predict a child’s aptitude for learning in school as early as age three (Thomas, 2006).

The term “school readiness” is a multidimensional concept, one that, according to Doherty (1997, p. 25), refers mainly to a child’s ability to handle the typical tasks requested of him or her at school, such as staying seated and learning the material. It also encompasses the skills the child must learn from birth to age six in order to ensure not only academic achievement, but also [TRANSLATION] “success in all aspects of adulthood, particularly in the job market....” Five aspects of school readiness are common to all research (Meisels and Atkins-Burnett, 2006):

- 1) physical well-being and appropriate motor development;
- 2) emotional health and a positive approach to new experiences;
- 3) age-appropriate knowledge and social skills;
- 4) age-appropriate language skills; and
- 5) age-appropriate general knowledge and cognitive skills.

Although factors 1, 4 and 5 are aspects generally associated with success in the specific tasks necessary for academic achievement, points 2 and 3 are also predictors of that success. A child’s emotional control, general attitude in the classroom (staying seated all day or curiosity in learning, for example) and social skills are essential to academic achievement. The National Education Goals Panel (Meisels and Atkins-Burnett, 2006) recognizes that these five factors are inseparable from family, school and community and that school readiness must be understood in relationship to these levels of influence.

⁷ See Guimond (2003) for an overview of studies on ethnolinguistic vitality in minority settings.

1.6. ABOUT THIS REPORT

This report details two studies on both cohorts. The goal of the first study is to examine implementation fidelity and quality for the Daycare Program and for a family literacy program, referred to as the Family Literacy Workshops. The goal of the second study is to determine the short-term effects of the daycare component's new preschool program on child development, particularly children's language skills and school readiness. It also aims to determine how the Family Literacy Program affects parents' attitudes and behaviours. These analyses use data collected from May 2007 to October 2009 for first-cohort participants and from September 2008 to October 2010 for second-cohort participants.

Chapter 2 details the implementation activities for both components of the tested program. Chapter 3 concerns the project's methodological aspects, including the experimental design, study sample, measurement tools, and study hypotheses. Chapter 4 describes the methodology used to review the implementation of the tested program's two components. Chapters 5 and 6 report, respectively, the implementation results for the Daycare Program and the Family Literacy Program. Chapter 7 concerns the strategies used to ensure data quality and reliability as well as the results of preliminary analyses for imputation of missing data, specification of error terms, and identification of confounding variables. The chapter concludes with the results of analyses designed to determine the representativeness of the sample for the Readiness to Learn project compared with a sample of the minority Francophone general population from the Survey on the Vitality of Official-Language Minorities (2006). The report goes on to detail the approach adopted for the impact analyses in Chapter 8. Chapter 9 covers the results of impact analyses for children's developmental outcomes in the first and second years of project implementation, while Chapter 10 describes the results of analyses on how the family literacy component impacts participating parents. Chapter 11 concludes with a review of the main findings and a discussion of all the reported results.

This report is one in a series of reports by SRDC. It follows the *Readiness to Learn in Minority Francophone Communities: Reference Report* (Legault, Mák, Verstraete, & Bérubé, 2014), the final version of which was submitted to HRSDC on October 13, 2009). This first report sketched a profile of the children, their families and the communities participating in the Readiness to Learn project. Moreover, this report supplements the *Readiness to Learn in Minority Francophone Communities: Project Implementation Report* (Bérubé, Legault, Janisse, Carson, Saucier, & Lefebvre, 2014) and the *Readiness to Learn in Minority Francophone Communities: First Cohort Findings Report* (Thompson, Legault, Lalonde, & Bérubé, 2014). The former concerns the analysis results for data collected to review project implementation for the first cohort, while the latter details the results of analyses examining the program's impact on first-cohort child developmental outcomes as well as parental attitudes and behaviours. Two other reports are planned: the first concerning the program's effects one year after the end of the intervention and the second, its effects two years after the end of the intervention.

2. Implementation Activities

This chapter describes the elements required to implement and deliver the program's two components. The chapter begins with information about how communities and families were recruited for both cohorts, followed by a description of the human capital required to implement the project. The chapter continues with information on the specific training sessions given to educators and family literacy practitioners. It ends with a description of each component and its delivery.

2.1. RECRUITMENT OF THE COMMUNITIES

First cohort: The first step in implementing the project was to contact interested communities with the potential to support the project for the years over which the tested program would be delivered. To ensure the pilot project's success, HRSDC drew up a list of the criteria that applicant organizations must meet to participate in the Readiness to Learn project. First, the applicant organization had to represent a local community and conduct activities in the field of early childhood development in a minority Francophone environment. Second, the applicant organization had to be a non-profit childcare provider or be associated with such a provider. Third, the community had to be characterized by a pool of eligible Francophones large enough to recruit at least 40 three-year-old children and provide a description of their linguistic profile. Fourth, to qualify, a community had to identify key individuals willing to get involved in the program. Finally, the applicant organization had to prove that the community had the necessary infrastructures and equipment to deliver the program throughout the pilot project, as well as the staff required to provide the childcare services at the daycares.

HRSDC organized an invitation to tender in order to recruit interested communities. Of the seven communities that applied, two were rejected. While the first initially seemed to have great potential, subsequent information on existing infrastructures revealed shortcomings. The second already had a junior kindergarten program at its daycares, a program that had inspired the study's new preschool program. Early on in the project, it became clear that a sixth community would be needed to achieve the statistical power required for reliable findings. A decision was made to develop a new site, but through the relationship formed with the school board in a previously recruited community; this sped up implementation so that the program could be delivered at the same time as in other project communities. This community was also familiar with the Readiness to Learn project, since it had participated in the pilot study earlier that year.

Second Cohort: In August 2008, the coordinators for the two communities began the pre-implementation activities. They started by contacting the champions to obtain their support for a second cohort to be studied in their community. The next step was to contact the childcare providers and the daycare coordinators that were already participating in the project (first cohort) to ask whether they would cooperate if the program were to be implemented and evaluated for a second cohort.

Once HRSDC authorization was received, the coordinators approached all the daycare educators and coordinators to properly explain the goals and the terms of their participation in

the Readiness to Learn project and, in particular, to obtain their invaluable help in recruiting families. In Orleans, two daycares were added to the comparison group, for a total of four daycares, to ensure that the target number of second-cohort participants was reached; in Cornwall, two new educators were assigned to the Daycare Program.

2.2. RECRUITMENT OF THE PARTICIPATING FAMILIES

A number of communication methods or strategies were used to reach the population in each selected community. In four of the six first-cohort communities, participants were recruited to form three treatment groups: the Program Daycare group, the Comparison Daycare group and the Informal Care group. No Comparison Daycare group could be formed in the two other communities for lack of a second Francophone childcare environment. In the two second-cohort communities, participants were recruited for all three treatment groups.

Steps were taken with the French-language daycares in each community to reach parents in the Program Daycare group and the Comparison Daycare group. The community coordinators contacted the parents of eligible children attending daycare directly to explain the research project and the benefits of participation. This led to as many children as possible being recruited in several communities. The coordinator then made an appointment with each parent to present the Readiness to Learn project more completely, before asking him or her to sign the informed consent and to fill out the baseline survey. Community coordinators also organized information sessions at some daycares in the program and comparison groups to provide all the necessary information about the Readiness to Learn project and to answer parents' questions.

A wider range of approaches was used to recruit parents for the Informal Care group. Among other means, SRDC used television and radio reporting, the distribution of information leaflets at various locations (e.g., school, medical clinics, early childhood centres, etc.), as well as advertisements in local and regional newspapers. In the end, word-of-mouth was by far the most effective method of recruiting members for this comparison group.

Thanks to these strategies, the required number of participants was recruited for both cohorts in all three treatment groups, i.e., the Program Daycare group, the Comparison Daycare group and the Informal Care group. What's more, SRDC took advantage of the networks established in the project's first two years to speed up second-cohort recruitment. For example, first-cohort parents and educators were asked to help recruit parents for the second cohort.

2.3. HUMAN CAPITAL REQUIRED FOR IMPLEMENTATION

Project implementation required numerous resources, both human and material. These resources made it possible to both implement and evaluate the project. The sections that follow explain the role of each stakeholder.

2.3.1. Champions

A champion was identified in each community to promote the project. The champion could be an individual or a local organization. Champions had to have in-depth knowledge and understanding of life and challenges in a minority Francophone community. They also had to be good communicators and be trusted by the parents. Concretely, champions served as project

ambassadors vis-à-vis community partners and families. Their support for the project helped establish the credibility of the Readiness to Learn project and promote it as an important initiative that was beneficial for the community.

2.3.2. Community Coordinators

The first responsibility of the community coordinators was to recruit children and their families. They went on to ensure that the positive relationship forged with parents during recruitment was maintained. Community coordinators kept parents informed about the project's latest developments. Note that this positive relationship led to a high rate of retention. The coordinators were also responsible for keeping community partners informed about project delivery, including data collection and follow-up. This exchange of information helped maintain the interest and support of project partners, which helped overcome the few challenges encountered.

In terms of implementation, the coordinators were responsible for the logistics of basic and follow-up training. They were also asked to play an active role in the family literacy component. They helped to recruit family literacy practitioners as well as to prepare and, where necessary, buy the material required for workshop delivery; they were also responsible for organizing meals, the drop-in childcare service, and the Resource Centre.

Finally, the community coordinators were responsible for coordinating data collection, including the children's quarterly assessment. They were responsible for the parent surveys conducted three times a year. The coordinators helped train the evaluators and conducted the follow-up necessary to ensure that quality data were collected from children. They also collected attendance data for the daycares and for Family Literacy Workshops, as well as observing the daycare activities and Family Literacy Workshops. Finally, they were responsible for interviewing the educators and community representatives.

2.3.3. Educators

The educators' involvement in the project and their enthusiasm for the new program were, unquestionably, crucial to its success. They were certainly the persons responsible for delivering the Daycare Program. A high level of educator engagement in the Readiness to Learn project inevitably contributed to improved quality of program delivery. In addition, educator engagement made it easier to recruit and retain parents and children for the Readiness to Learn project. Educators often encouraged parents who were hesitant to enrol their children or who were unsure about the program's terms and implications. The additional explanations and answers they gave often made a difference in the parents' decision to participate.

Under the Daycare Program, educators had to have an in-depth knowledge of the program and the target learning outcomes. They had to learn the program content and apply it in a manner that respected the program's approach and philosophy. This meant that educators had to be familiar with the stages of child language development, as well as with activities that enrich a child's vocabulary while consolidating his or her sense of belonging to the Francophone culture and community. The head educator was responsible for planning activities related to the month's theme. These play-based activities had to stimulate the child's five senses and develop his or her multiple intelligences. The francization techniques used gave children an opportunity to improve and master their vocabulary, as well as to develop an interest in reading and early writing. The

program focused on literacy activities, including interactive reading and learning nursery rhymes. Educators ensured that children became autonomous in terms of both social contact with peers and the choice of their own activities.

Educators were also asked to deliver the program for the Family Literacy Workshops in order to give children a sense of security. The workshop process involved keeping parents and children together in the same room when the workshop began and ended, but separating them in the interim. During that time, parents headed to another room to follow a workshop designed for them. The educator's presence with the child helped him or her feel secure and optimized his or her learning.

The educators' support for parents who attended the Family Literacy Workshops was also key in motivating them to continue attending. Parents were encouraged to participate regularly with regular reminders of meeting dates, the topic for discussion, and the activities that would be carried out with the children. In fact, the pilot project insisted on the importance of daily contact and communication between the parent and the child's educator. The discussions and communication that occurred during the family literacy component forged a tie between parents and educators and allowed them to share observations and discoveries about the child's developmental progress.

2.3.4. Family Literacy Practitioners

The parent component of the Family Literacy Workshops was led by practitioners with experience leading workshops for adults or with an adult education background. They had to properly understand the program content and the targeted learning outcomes. When delivering workshops, they were expected to use an adult education approach and to respect the program's philosophy. Practitioners had to create bonds with parents in order to understand the challenges they encountered at the family and community level. They encouraged discussions so that parents would actively participate and openly discuss the topics presented. They encouraged parents to participate in their child's language and vocabulary development by giving them activities to do at home.

In addition, the Readiness to Learn project required that practitioners be present at meals; this enabled them to support parents and talk with those who had more questions either before or after the workshop. An attentive ear and regular support for parents encouraged them to attend the Family Literacy Workshops regularly.

2.3.5. Trainers

SRDC hired the early childhood consultant who adapted the Daycare Program to train and monitor the daycare educators. The trainer's main responsibility was to make the necessary changes and adjustments so that educators could faithfully deliver the program. She also provided ongoing training for educators by advising them and providing them with targeted resources based on the needs she observed herself or those identified by the educators.

SRDC also hired the designers of the Family Literacy Program to oversee basic training and the monitoring of family literacy practitioners. The trainers were primarily responsible for ensuring that the practitioners properly understood the program's approach as well as its values and objectives. Practitioners could contact trainers as necessary to obtain clarifications on program content and advice on program delivery.

2.3.6. Evaluators

The hiring criteria for evaluators required that they have an aptitude for establishing a rapport with children and for communicating easily with them. Evaluators were first required to take basic training so that they would have the necessary skills to conduct the child assessments. The evaluators then worked in pairs for the first four children assessed, which gave them an opportunity to master the tool and make sure the collected data was valid.

Evaluators also had to know and apply requirements with respect to confidentiality and privacy, data handling, and transfer of the completed assessments and surveys. Before each assessment period, they took follow-up training to review how to apply the measurement tool and, where necessary, learn how to use new measurement tools.

2.4. TRAINING

2.4.1. Basic Training for Educators in the Daycare Program

Educators in every community were given a three-day basic training session in April and May 2007 for the first cohort. For the second cohort, the trainer gave educators a mini training session during a scheduled follow-up in September 2008. This was decided because most educators involved had already received basic training in the summer of 2007, in anticipation of delivering the Daycare Program for the first cohort. The training presented the theoretical and practical aspects of the concepts presented. Among these theoretical aspects were the pilot project's objectives and characteristics, the program's content as well as its values and basic principles, francization techniques, and communication with parents. Practical concepts concerned the presentation of a book and song as well as thematic planning. Each participant was given a manual detailing the content of the Daycare Program.

Discussions among participants led to identification of the differences and similarities between the existing daycare program and the tested program. These discussions also identified the aspects of the new program that would be problematic, as well as any resistance by educators; this information was used to plan follow-up sessions based on these elements. The training also gave the trainer an opportunity to determine each educator's specific needs.

The training went very well in general, as the educators were receptive and enthusiastic about the program. The educators' engagement in the Readiness to Learn project was also easier in communities where the staff already knew the community coordinator or where the community champion was directly involved in the childcare environment.

2.4.2. Follow-up Training for Educators in the Daycare Program

After basic training, follow-up training was given every two months in the six communities, beginning in August 2007 for the first cohort and in September 2008 for the second cohort. This follow-up training was given regularly over a period of one to two years, for as long as there were children in the Readiness to Learn project enrolled at the daycare. No follow-up training was given in summertime because regular programming was interrupted at all daycares.

For the first cohort, the purpose of the first follow-up training session was to facilitate the spatial reorganization of the rooms into learning centres and to help educators plan the following

month's activities. Room reorganization was not necessary for the second cohort since program delivery continued on the same premises as those used for the first cohort. The trainer then went on to observe classroom activities, give educators feedback and provide targeted training. Among the subjects addressed in the follow-up were thematic planning, children's portfolios, developing a literacy centre, integrating literacy and numeracy activities into daily activities, and encouraging creative drawing by children. At her last visit, the trainer asked educators to share their impressions on the program and to identify its strengths and weaknesses. She also awarded a certificate of participation to the educators who gave the Daycare Program under the Readiness to Learn project. Note that implementation of the Daycare Program failed in Edmonton.⁸ The trainer's last visit there was in September 2008.

2.4.3. Basic Training for the Family Literacy Program

A three-day basic training session was given in October 2007. Each participant received a manual that detailed the programming of the family literacy component that was to be delivered to the communities of the Readiness to Learn project. This guide was intended to be a roadmap and provided significant ongoing support for the practitioners who delivered the Family Literacy Workshops. The training helped participants better understand the minority Francophonie, introduced them to the basics of family literacy in a minority Francophone context and explored the specificities of families living in that context. The trainers gave an overview of the 10 workshops and explained the child, parent and parent-child components. The trainers actively involved the practitioners in illustrating each workshop's mandatory activities. Finally, the trainers had a long discussion with participants on how to adjust workshop content to reflect the unique characteristics of each community. For example, some practitioners from communities with a number of exogamous families expressed worries and fears regarding the fact that workshops were given only in French. The exchanges were an opportunity to identify the strategies that would allow practitioners to meet the needs of Anglophone parents while ensuring that the workshops remain a place where French is favoured and emphasized.

Basic training was also given to new practitioners for the second cohort in October 2008. A review booklet was distributed to practitioners who had worked with the first cohort and who would be delivering the workshops for the second cohort. The purpose of this booklet was to remind the experienced practitioners of the Family Literacy Program's key aspects, including workshop content as well as the approach for children and parents living in a minority Francophone environment.

Note that community coordinators also participated in basic training (in October 2007 and October 2008) to become familiar with the program's objectives and principles as well as with the workshops' content. Training served to jump-start teamwork by the coordinator, the practitioners working with parents and the practitioners working with children.

2.4.4. Follow-up Training for the Family Literacy Workshops

Soon after the first series of workshops ended, the trainers organized two teleconferences. These sessions were to obtain progress reports on workshop delivery, answer the practitioners' questions, and suggest solutions to any difficulties encountered. The trainers began by addressing

⁸ For an analysis of this failure, see the *Readiness to Learn in Minority Francophone Communities: Project Implementation Report* submitted to HRSDC in August 2010.

the most difficult aspects of workshop delivery and offered potential solutions to those barriers. The issue of accommodating Anglophone parents was addressed. One practitioner working with parents offered a few ideas to facilitate the inclusion of an Anglophone parent. The trainers were available throughout the workshop to answer any questions from practitioners about the program's delivery or content as well as to suggest solutions to any challenges that came up.

2.5. DESCRIPTION OF THE DAYCARE COMPONENT

The program delivered as part of the Readiness to Learn project uses a preschool or school readiness approach that views child development as a holistic process; a number of developmental dimensions are therefore targeted. With this type of program, children are led to achieve specific development objectives directly linked to school readiness and intended to facilitate academic achievement. This approach stands opposed to a “social” teaching approach where the program gives general orientations and where each environment adopts elements based on the specificities of its community. Since the goal of the Readiness to Learn project is to optimize the linguistic and general development of children living in minority Francophone contexts, the daycare component particularly emphasizes children's exposure to verbal communication through books and songs, while giving them many opportunities to express themselves and develop their thinking in French.

The daycare component of the Readiness to Learn project favours a play-based approach. Based on this approach, [TRANSLATION] “children are considered independent beings who can actively shape their learning environments.” (Bertrand, 2007, p. 4). Finally, this approach encourages creativity, which fosters process, exploration and experimentation. With a play-based approach, children learn through literacy activities as well as through a range of recreational activities. Children are encouraged to explore and learn through play with the ongoing support of the educator, who encourages them and closely monitors their development. The program is designed to influence the child's socialization, francization and general development (*Programme des prématernelles en garderies, Manuel de formation* [training manual for junior kindergarten daycare program] 2007). The fundamental principles of the Daycare Program include:

- Making the child's needs central to the program and activities;
- Stimulating the child's five senses and multiple intelligences;
- Providing an environment rich in oral and written language;
- Using and applying francization techniques to promote the use of French;
- Promoting autonomy;
- Providing the child with learning centres that interest him or her and encourage the child to make choices;
- Offering the child appropriate learning material and setting up that material at his or her height;
- Fostering positive interactions with the child and his or her parents;
- Supplementing the daycare program with a family literacy program.

The programming of the daycare component is based on the Franco-Saskatchewanian junior kindergarten program developed by the Ministry of Education of Saskatchewan (2001) for four-year-olds. An early childhood consultant hired by HRSDC adapted the program for the younger population of the Readiness to Learn project, that is, three-year-olds. She also made several changes to the Daycare Program by adding elements from the *Jouer c'est magique* program, particularly in terms of the daily schedule and the length of the day in a childcare environment. Finally, with the creative approach to learning concept, the consultant:

- created or modified the assessment tools, such as the observation grids and the child's portfolio;
- specified the monthly themes and simplified the thematic weekly programs;
- specified the method of presenting a book and song;
- reduced the number of learning centres and simplified their content; and
- drew up a list of the resources and material needed to deliver the program.

Consistent program delivery is ensured through 15 kits containing the necessary resources and material. The content of 12 of the 15 kits is tied to a specific monthly theme. There is also an art kit, a science kit and a music kit. The kits include toys, puzzles, books, CDs, music instruments and plastic arts material. All the material is in French.

2.5.1. Delivery of the Daycare Program

First cohort: The first year of the Daycare Program covered the months of September 2007 to August 2008. The program was inaugurated on September 1, 2007, in the communities of Cornwall, Edmonton, Edmundston and Saint John. It was launched in October 2007 in the communities of Durham and Orleans. The program's second year ran from September 2008 to August 2009 in only four of the six communities. Children in Edmonton, Edmundston and Saint John were exposed to the Daycare Program on a full-time basis while children in Cornwall were exposed to it on a part-time basis. Children in Durham and Orleans were not exposed to the Daycare Program in the second year. In these communities, parents enrolled their child in the free full-time junior kindergarten that was available. Note that the tested program was not faithfully applied in the summertime, either in program daycares delivering the Readiness to Learn project or at the comparison daycares delivering another program.

Second Cohort: The first year of the Daycare Program for the second cohort covered the months of September 2008 to August 2009. The program simply continued to be applied in the communities of Cornwall and Orleans for the benefit of a second cohort. As with the first year, we observed that the tested program was not faithfully applied in the summertime at program daycares delivering the Readiness to Learn project and at comparison daycares. The program delivery schedule is presented in Table 2.1.

Table 2.1: Delivery Schedule for the Preschool Program

Community	Date launched	Date ended	Degree of exposure
Cornwall – Cohort 1	August 28, 2007	August 24, 2009	Full-time the first year and part-time the second year
Cornwall – Cohort 2	September 1, 2008	June 21, 2010	Full-time the first year and part-time the second year
Durham	September 24, 2007	July 31, 2009	One year full-time for most children (2007 to 2008) and two years full-time for a few children
Edmonton	September 17, 2007		Failed implementation
Edmundston	September 4, 2007	June 19, 2009	Two years full-time
Orleans – Cohort 1	October 1, 2007	August 29, 2008	One year full-time
Orleans – Cohort 2	September 1, 2008	August 28, 2009	One year full-time
Saint John	September 4, 2007	August 27, 2009	Two years full-time

2.6. DESCRIPTION OF THE FAMILY LITERACY COMPONENT

The Family Literacy Program was developed specifically for the pilot project by Éduk, in collaboration with HRSDC and SRDC. The program was designed to meet the objectives of the Readiness to Learn project and the special needs of minority Francophones. Thus, its goal is child development rather than improved literacy for parents. The emphasis is on the children's knowledge development, experiences and socialization periods. The program is intended to equip parents so that they can support their child's development in terms of the French language, culture and identity, whether they live in a unilingual, bilingual, trilingual or multicultural family context. It also aims to raise parents' awareness about the work of educators and the importance of the complementary parent-educator roles in supporting the child's learning.

The content is based on the strengths of Canada's major recognized literacy programs, including:

- *Grandir avec mon enfant* (2002) and its adaptations, particularly Nova Scotia's *J'apprends en famille*, especially as concerns activities related to parenting skills and children's needs;
- *Chansons, contes et comptines*, as well as *Grandir avec des livres*, since the accent is on pre-reading skills and the Francophone cultural component;
- the English-language "Learning Together" program, which was the subject of a longitudinal study in Alberta (from 2001 to 2005), and whose effectiveness with children and families are well documented (Phillips, Hayden and Norris, 2006); and
- the *Programme des prématernelles fransaskoises* (2001).

2.6.1. Harmonization of the Two Program Components

To maximize the impact of the Readiness to Learn project on families, it was essential to harmonize the Daycare Program and the Family Literacy Program with respect to the two programs' values, fundamental principles and approaches. The Daycare Program focuses on the francization and early literacy of preschoolers (including the subthemes of reading and writing). The Family Literacy Workshops complete the Daycare Program with exchanges that raise parents' awareness of their role as their child's first educator and of the specificities of life in a linguistic minority setting. The main themes concern how to support their child's development, stimulate his or her learning, and convey French language and culture to him or her. The content also aims to raise parental awareness regarding the work of educators and to foster partnership among those working with the child (educator, parents and others).

To tie in these components, the designers of the Family Literacy Workshops worked closely with the early childhood consultant who adapted the Daycare Program. The advantage of this strategy was to ensure consistency among the various aspects of the Readiness to Learn project and continuity in the learning by child and parent. The designers factored in:

- the Daycare Program's underlying approach and values;
- francization strategies for use with children;
- the themes explored every month at daycare ;
- the list of resources (games, toys and books purchased for the daycare component); the proposed resources for the Family Literacy Workshops round out those of the daycare component; and
- the developmental stage of preschoolers.

The harmonization of the two components is itself an important contribution to child development. Parents who get involved by talking to the educator, asking her questions about their child's day and participating in daycare activities have children who show a broader vocabulary, more phonological awareness and better early writing skills (Arnold, et al., 2008). The Daycare Program and the Family Literacy Program highlight the importance of good cooperation between parents and educators. Harmonization of the two components fosters partnership among those working with the child (educator, parents and others) and aims to strengthen the child's learning in various areas of his or her life, such as at daycare, at home, and in the community.

2.6.2. Delivery of the Family Literacy Program

Family Literacy Workshops were given only in the first year of the tested program's implementation. The Family Literacy Program consisted of 10 workshops presented in two series. For the first cohort, an initial series of four workshops was offered in November and December 2007 and a second series of six workshops was given in January and February 2008. For the second cohort, an initial series of five workshops was offered in November and December 2008 while the second series of five workshops was presented in January and February 2009. Only the families in the Program Daycare group were asked to participate in the family literacy component.

The workshops were given on the daycare premises to ensure that children benefited from a safe environment suited to their needs (e.g., appropriately sized tables and chairs). The daycare is also a place that is familiar to the children, which gives them the sense of security so crucial to optimizing their learning. Each workshop lasted two hours and consisted of five components: the greeting component, the parent component, the child component, the parent-child component, and the closing component. Parents and children were together in the same room for the greeting component, the parent-child component and the closing component. The parent component and the child component happen at the same time in different rooms. The core program takes place during the parent component, when parents are given information, followed by discussions and exchanges when parents share their experience in order to learn from each other. This approach builds on family strengths and reinforces parents' existing practices. The practitioner serves as a facilitator who encourages discussions, identifies important points, and refers families to the information they need.

A learning kit is lent during the closing component. Parents bring the fabric bags home and then return them at the next workshop. Every kit contains the two following French-language resources: a book for the parent and a music CD, a DVD or an educational game. Families returned the kit at the next workshop and exchanged it for another. The kit also contained the materials and written instructions for a creative activity to do at home with the child. There are a total of 10 different creative activities, one to be added to the kit at the end of each workshop. Children kept the material. The purpose of the creative activity was to reinforce what the children learned at daycare and during the workshop. Families were also invited to visit the Resource Centre set up specifically for the workshops. This centre provided French-language resources, including books for children and parents, CDs, DVDs, books/CDs and games. Every child could choose one book or other resource and add it to his or her kit.

Several elements were introduced to reduce barriers to family workshop participation, particularly a meal before or after the workshop. Offering a meal before the workshop makes it easier for families to participate, since they no longer have to head home for supper then return for end-of-day workshops. Similarly, for morning workshops, the option of a family lunch before leaving was less of a disturbance to the schedule of young children, who often nap in the afternoon. Where numbers justified, a drop-in daycare service was offered for siblings of the children in the Program Daycare group who were age two-and-a-half or older. Families could also claim a set amount for babysitting expenses incurred for children under age two-and-a-half. Finally, the coordinators were in regular contact with the families to confirm their participation in the workshop. For example, coordinators made calls, sent e-mails or handed out letters to all parents enrolled in the Family Literacy Workshops to remind them when workshops resumed in January.

2.6.3. Resource Centre — Purchase of Material

Over 300 French-language resources were purchased to ensure consistency in program delivery. The families of the first and second cohorts were invited to borrow these resources during and after the end of Family Literacy Workshops. Note that in 2008–2009, parents and children in both cohorts could borrow books from the Resource Centre. About 60 more French-language resources were thus purchased in September 2008 to better serve the higher number of families visiting the Resource Centre. Most of the resources purchased were intended to replace

material that was missing or damaged after the Family Literacy Workshops were delivered to the first cohort. SRDC also increased the number of books for parents.

Once Family Literacy Workshops ended, the community coordinators came up with a book-lending system for all the resources used as part of the workshops. These resources could be borrowed on two occasions during the week. During the day, the coordinators visited the classrooms of children in the Program Daycare group so they could borrow a book from the Resource Centre. At the end of the day, they dropped in at the Daycare Program so that parents could visit the Resource Centre with their child. This gave families an opportunity to borrow a resource for children and parents. Note that the Resource Centre was closed over the summer.

The activities of the Resource Centre continued in the second year in every community of the Readiness to Learn project, whether or not the children in the Program Daycare group were attending daycare. Thus, families in Edmonton, Durham, Edmundston and Saint John were able to borrow resources until June 2009. Families in Cornwall and Orleans (where there was a second cohort) were able to borrow resources until June 2010. The Resource Centre was, however, closed over the summer. The coordinators carried out an inventory of the Resource Centre several times during and after delivery of the Family Literacy Program and the creation of the Resource Centre. At the end of the Readiness to Learn project, the content of the Resource Centre was given to the provider responsible for services at the Daycare Program.

3. Methodology

This chapter concerns the methodological aspects used in the Readiness to Learn project. The first section describes the eligibility criteria and the target population while the second presents the project's experimental design. A third section explores certain threats to internal validity and the strategies introduced to counter them. The fourth section details the measures used in the impact analyses for the tested program. The last section concerns the series of hypotheses tested as part of the impact analyses.

3.1. TARGET POPULATION

Parents and children were recruited based on specific eligibility criteria. The *first criterion* was that the child be born between January 1, 2004, and January 31, 2005, for the first cohort, and in 2005 for the second cohort. This criterion corresponded to the eligibility criterion of enrolment for kindergarten in September 2009 for the first cohort and in September 2010 for the second cohort set by the Ontario Ministry of Education and the New Brunswick Department of Education and Early Childhood Development. The *second criterion* required that one of the child's parents be an "*ayant droit*" as defined by section 23 of the *Canadian Charter of Rights and Freedoms*.⁹ Since the purpose of the tested program was to enhance children's language skills and school readiness, this criterion ensured that the target population, that is, children entitled to attend French school, was reached.

The *third criterion* concerned the parent's intention to send his or her child to a French-language school. This criterion was in fact rarely applied since the parents of such young children had generally not made a decision regarding choice of school. However, if the parents answered that English-language school was their final choice, SRDC made the decision not to obtain the parent's informed consent since the new preschool program was designed in part to better prepare children for French-language school.

3.2. EXPERIMENTAL DESIGN

The program was evaluated using a quasi-experimental design with non-equivalent control groups. As with experimental studies in the field, a quasi-experimental design is intended to test, by means of comparison groups and pre-intervention measures (i.e., measures taken before the intervention), the causal hypothesis that an intervention has a significant effect on the variables in question beyond changes that would occur in the absence of that program and independently of other external factors that may come into play.

The methodology involves three treatment groups: a Program Daycare group, consisting of children enrolled in a French-language daycare that offers the new preschool program; a Comparison Daycare group, consisting of children enrolled in a French-language daycare that does not offer the new program; and an Informal Care group, consisting of children whose daytime care is provided at home or at an unregulated family daycare. The Comparison Daycare group takes into account how formal daycare affects child development, which is a subject unto

⁹ *Canadian Charter of Rights and Freedoms* available at <http://laws.justice.gc.ca/eng/Charter/page-1.html>.

itself. The purpose of the Informal Care group is to factor in how an informal childcare environment affects child development, especially with respect to the French language. Children enrolled in an English-language daycare were added to the Informal Care group for the impact analyses. Although these children were exposed to a formal daycare environment, they were not exposed to French. By not classifying these children in the Comparison Daycare group, we preserved language homogeneity in the Comparison Daycare group and exposure to a French-language daycare program.

3.3. INTERNAL VALIDITY

Since random assignment was not used, there are likely inherent differences between the Program Daycare group and the comparison groups from the outset. This makes it even more important to implement conditions to ensure the study's internal validity, thereby eliminating from the start any *plausible* alternative explanations for results. In fact, it is less advantageous to use control techniques, often statistical, after the study ends.

Under the Readiness to Learn project, the conditions necessary for internal validity are:

- sample size;
- a sampling strategy that ensures equitable distribution among the groups;
- pre-intervention measures of the program's expected effects (hereinafter referred to as "outcomes") and associated factors;
- a verification of how sample attrition affects group distribution;
- special attention to situations that may lead to contamination of the comparison groups.

3.3.1. Sample Size

To ensure the internal validity of the impact study with three groups, a sample must consist of at least 156 children broken down equally into the Program Daycare group ($n = 52$), the Comparison Daycare group ($n = 52$), and the Informal Care group ($n = 52$). This number of children per group provides the statistical power necessary to detect a moderate impact, with a very good confidence level that the real population value of the impact estimate falls within a specified range (i.e., 19 times out of 20, the same results would be obtained with other samples).

3.3.2. Targeted Sampling Strategy

A targeted sampling strategy was chosen to create comparison groups that were highly similar to the Program Daycare group so as to neutralize as much as possible the influence of unmeasured factors on the study's outcomes. Previous studies have shown that certain characteristics, like family income and the parents' level of education, accounted for part of a child's developmental trajectory. It follows that if the Program Daycare group differs considerably from the comparison group with regard to one of these characteristics, any differences observed between the two groups on outcome measures may well be explained by these initial differences rather than by the program. Thus, a special effort was made to target participants for the comparison groups with a sociodemographic profile similar to that of the Program Daycare group (e.g., socioeconomic level) and living in the same neighbourhood, thus

ensuring they had access to the same French-language resources and services as the Program Daycare group¹⁰. A first control for group composition was conducted when families enrolled in the project, particularly as regards home location for potential members of the comparison groups. A second control consisted of using pre-intervention measures, which is discussed in the next section.

3.3.3. Pre-Intervention Measures

The causal inference arising from a quasi-experimental methodology is facilitated by the use of pre-intervention measures of outcomes and associated factors. In the case of the Readiness to Learn project, the outcome in question is children's school readiness as measured by the *Évaluation de la petite enfance — Appréciation directe* (ÉPE-AD) and its expressive and receptive vocabulary subscales. The ÉPE-AD is the validated French translation of the Early Years Evaluation — Direct Assessment (EYE-DA). Given that the results presented in this report are based exclusively on the French version of the scale, it will henceforth be referred to exclusively by its French acronym, the ÉPE-AD.

Pre-intervention measures allow for a better understanding of how the program affects child development since the school readiness of all children prior to project implementation can be compared with their school readiness after project implementation. These measures also enable us to verify whether children in the Program Daycare group and those in the comparison groups follow similar developmental trajectories before the start of the intervention. This information is useful because the program impact estimates would be biased if any of the treatment groups consisted of children who were more developmentally advanced than those in another group before the intervention began.

The purpose of pre-intervention measures of sociodemographic and socioeconomic variables is also to establish whether children in all three treatment groups experienced similar conditions before the intervention. The choice of variables for the pre-intervention measures was based on previous study findings showing that these significantly influence a child's school readiness. Statistical control of these variables in the analyses will allow for separating the effects of the new preschool Daycare Program on child development from the effects of other variables for which empirical evidence has already been established.

The initial profile of the children, their family and the communities participating in the Readiness to Learn project was reported in a document entitled *Readiness to Learn in Minority Francophone Communities: Reference Report* (Legault et al., 2014). This report also contains the result of analyses establishing the homogeneity of the treatment groups before the intervention. Note that the results detailed in the report stem from analyses of the six communities. These analyses had to be redone for this report so as to include only the four communities included in the impact analyses. The list of identified variables is found in Section 7.3 of this report while the results of the comparative analyses are found in Section 7.4 of this report.

3.3.4. Attrition

Loss of participants in the course of the study jeopardizes its external validity (i.e., the degree to which the sample represents the population it is supposed to represent) as well as its internal

¹⁰ For more information, refer to the Revised Work Plan and Methodology Report submitted to HRSDC on March 30, 2007.

validity (i.e., by changing the composition of treatment groups). There can be several reasons for a loss, including a move or a participant's loss of interest. Whatever the reason for withdrawal, it is important to determine whether the group of participants that has withdrawn from the study differs from the group of participants remaining in the study as regards certain sociodemographic aspects. Differences between the two groups may signal the presence of a subgroup of the target population with particular characteristics (which constitutes a threat to external validity). Moreover, a significant change in the composition of the treatment groups may affect the validity of the impact estimates (i.e., may affect internal validity). A control for the reasons of withdrawal from the project was therefore introduced from the project's outset in order to better assess the threat to the study's internal validity arising from the circumstances related to withdrawals.

3.3.5. Prevention of Comparison Group Contamination

Contamination occurs when changes in the scores for the outcomes of the comparison groups are attributable to the application of the tested program's modalities and conditions. In other words, contamination happens when comparison group members are exposed to the components of the Daycare Program and/or those of the Family Literacy Program.

Contamination of the Comparison Daycare group is more probable in communities where more than one of the region's French-language daycares is participating in the project, making it possible for information to circulate between the program daycare and the comparison daycare. Several strategies were introduced to minimize opportunities for information transfer (a list is provided below). In addition, childcare service providers were warned about situations that may lead to contamination of the comparison groups and alerted to the importance of avoiding or minimizing these situations for the study. This information was mainly targeted to communities with several French-language daycares overseen by the same childcare service provider. These discussions led to close cooperation between SRDC and the childcare service providers.

The strategies implemented to minimize contamination opportunities were:

For the daycare component:

- Basic and follow-up training were given only to educators and assistant educators involved in delivering the Daycare Program.
- Movement of educators trained in the Daycare Program to a comparison daycare was limited. In fact, staff movement from a program daycare to a comparison daycare occurred in only one community. In one case, the educator was assigned to a group of children younger than children in the Readiness to Learn project. In the other cases, the assistant educators were assigned to daycares that were not participating in the Readiness to Learn project. Any staff leaving the program daycare was informed of the importance of not sharing the knowledge and methods with other educators and was required to return the training manual to the coordinator. Finally, the coordinator closely monitored the situation at the comparison daycare.
- The classes of the Comparison Daycare group were located in a building separate from that where the classes in the Program Daycare group were given.
- The resources and material of the Daycare Program had to remain in the program daycare until three months after delivery of the tested program ended.

For the family literacy component:

- Basic and follow-up training were given only to the practitioners involved in delivering the Family Literacy Workshops.
- After delivery of the family literacy component ended, practitioners had to return the training manual to the coordinator. They were also informed of the importance of not sharing the knowledge and methods with others.
- The childcare service providers and the school boards were prohibited from delivering the Family Literacy Program or a similar program in the community of the study.
- The resources and material of the Resource Centre had to remain in the program daycare until three months after delivery of the tested program ended.

3.4. SAMPLE BY COMMUNITY AND BY TREATMENT GROUP

The first cohort of project participants was from the minority Francophone communities of Cornwall, Orleans and Durham, Ontario, and Edmundston, New Brunswick.¹¹ The recruitment period for this first cohort extended from May to October 2007. The second cohort of project participants resides in the minority Francophone communities of Cornwall and Orleans, Ontario. The recruitment period for this second cohort took place in fall 2008. Note that the statistics describing the sample reported in this section concern the participants of both cohorts combined.

Gross sample: At the time of enrolment, the project involved 356 children from 352 families. As observed in Table 3.1, the communities of Edmundston and Cornwall (first cohort) had the highest proportion of participants (23.9% and 20.2%, respectively), followed by the communities of Orleans (15.5% for the first cohort and 15.7% for the second cohort), Cornwall (12.9% for the second cohort) and Durham (11.8%). The sample consisted of 341 children at the end of the first year of intervention (both cohorts combined), of which 167 were boys and 174, girls. Sample size was down slightly to 338 children at the end of the second year of intervention, of which 166 were boys and 172, girls.

The average age of children at enrolment was 38 months. At that time, the Readiness to Learn project had a balanced number of boys (49.4%) and girls (50.6%). The mother tongue of children in the sample (according to the most knowledgeable respondent) was majoritarially French (71.1%) followed by English or another language (19.4%).

According to the baseline survey, the mother's age at birth of the child in question was 27.5 on average. Moreover, 80% of these mothers had at least a college diploma, and half of them had a university degree. Average household size was four members. 8.4% of families were headed by single parents. Over half of the participating families (59.1%) had an annual income of over \$70,000; median annual income ranged from \$80,000 to \$99,999.

With respect to the linguistic profile of the participating families, over half of the mothers (61.7%) and fathers (55.0%) spoke only French to their child. Most children were from Francophone endogamous homes (49.3%), followed by exogamous homes (39.2%).¹²

¹¹ It was impossible to properly measure the program's impact in the communities of Edmonton and Saint John due to the lack of French-language daycares that could be used as counterfactuals.

¹² Homes were categorized based on the combination of the mother's and the father's first official language spoken (FOLS).

Table 3.1: Participant Breakdown by Community since Enrolment

	Enrolment	Year 1	Year 2
Cornwall - Cohort 1	72 (20.22%)	69 (20.23%)	68 (20.12%)
Cornwall - Cohort 2	46 (12.92%)	45 (13.2%)	45 (13.31%)
Durham	42 (11.8%)	36 (10.56%)	35 (10.36%)
Edmundston	85 (23.88%)	83 (24.34%)	83 (24.56%)
Orleans - Cohort 1	55 (15.45%)	55 (16.13%)	54 (15.98%)
Orleans - Cohort 2	56 (15.73%)	53 (15.54%)	53 (15.68%)
Total	356 (100%)	341 (100%)	338 (100%)

Sample by treatment group: At the time of enrolment, the project involved 110 children enrolled in the Program Daycare group (G1), 135 children in the Comparison Daycare group (G2) and 111 children in the Informal Care group (G3). A review of Table 3.2 shows a reduction in G1 and G3 along with an increase in G2 sixteen months after enrolment (i.e., +16 months).¹³ These changes in the composition of treatment groups occurred because children in Ontario registered for school at the start of the project's second year. These fluctuations were also greater in Cornwall where several children in G1 and G3 enrolled at a school where a comparison daycare was also located. Changes in the composition of the treatment groups were therefore factored into the analyses for the program's second year.

Table 3.2: Participant Breakdown by Treatment Group

	Enrolment	+ 4 months	+ 8 months	+ 12 months	+ 16 months	+ 20 months	+ 24 months
G1	110 (30.9%)	104 (29.7%)	97 (28.0%)	97 (28.4%)	94 (27.6%)	91 (26.9%)	91 (26.9%)
G2	135 (37.9%)	132 (37.7%)	133 (38.3%)	132 (38.0%)	159 (46.6%)	160 (47.3%)	160 (47.3%)
G3	111 (31.2%)	114 (32.6%)	117 (33.7%)	113 (33.0%)	88 (25.8%)	87 (25.7%)	87 (25.7%)
Total	356 (100%)	350 (100%)	347 (100%)	342 (100%)	341 (100%)	338 (100%)	338 (100%)

3.4.1. Retention Rate

The Readiness to Learn project has an excellent retention rate, with only 18 withdrawals (5.1%) since the project began in 2007. The main reason for a child's withdrawal from the project is a family move outside the community. Note that some children continued to be followed in cases where the move was to another community participating in the project. From that point on, these children were considered participants in the new community. Table 3.3 breaks down the reasons for withdrawal from the project.

¹³ Although the children registered for school 12 months after enrolment (+12 months), changes in the composition of the treatment groups do not appear immediately in the data collection due to the assignment rules. These are detailed in Section 8.4.

Table 3.3: Reasons for Child Withdrawal from the Readiness to Learn Project

Reason for Withdrawal	Year 1	Year 2	Total
Move to a non-participating community	6	2	8
Child changed to an English-language daycare/not enough English at the program daycare	3	0	3
Families' lack of availability	2	2	4
Bothered by questions in the baseline survey	1	0	1
Loss of contact with the family	2	0	2
Total withdrawals	14	4	18

3.5. MEASURES

This section presents the plan for data collection from parents and children, as well as for the Family Literacy Workshops. It continues with a description of the quantitative measurement tools used from project commencement until the end of the second year, that is, from summer 2007 to October 2009 for the first cohort and from October 2008 to October 2010 for the second cohort.

3.5.1. Data Collection Plan

Data were collected from parents and children on a quarterly basis. Baseline data for the first cohort was collected from May to October 2007. The next collections took place in February, June and October 2008, as well as in February, June and October 2009. Data collection for the second cohort followed a similar pattern, but one year later. More specifically, the baseline collection was conducted in October 2008, and the following collections were made in February, June and October 2009, as well as in February, June and October 2010. The last data collection for both cohorts was carried out post-program, that is, at a time when all children were attending school full-time.

Parent Surveys

In total, seven surveys were conducted of parents for both cohorts since the start of the Readiness to Learn project. The baseline survey was carried out when the child enrolled in the project, that is, from May 1 to October 31, 2007, for participants in the first cohort and in October 2008 for those in the second cohort. The survey lasted about an hour and was conducted in person by the community coordinator with the parent most knowledgeable about the child (PMK).¹⁴ The six surveys that followed were short, ranging from 10 to 30 minutes. They were administered over the phone by the coordinators or by a member of SRDC. The average response rate for the follow-up surveys was 94.7% (see Table 3.4).

¹⁴ A second baseline survey was administered only to first-cohort parents who enrolled in the project prior to September 1, 2007. This survey served to update information on the type and frequency of literacy activities and non-daycare activities performed with the child when he or she was just starting daycare. For impact analyses, the data from the second baseline survey replaced the data from the first baseline survey, if any. The two surveys are thus considered a single survey for the purposes of the analyses.

Table 3.4: Response Rates for Parent Surveys

	Baseline survey	+ 4 months	+ 8 months	+ 12 months	+ 16 months	+ 20 months	+ 24 months
Surveys administered	356	341	343	335	338	335	331
Absent	0	9	5	7	3	3	7
Withdrawals	0	6	8	14	15	18	18
Response rate	100%	95.8%	96.3%	94.1%	94.9%	94.1%	93.0%

For the Family Literacy Workshops

Since the new tested program included a family literacy component targeting the parents of the Program Daycare group, we created an assessment to determine this component's impact on parental attitudes and behaviours. Two surveys were developed to gather information before and after the Family Literacy Workshops were delivered. The pre-intervention survey was administered by phone to all parents of the Program Daycare group in November 2007 for the first cohort and in November 2008 for the second cohort. This survey measured expectations, opinions and certain behaviours of parents in the Program Daycare group when the Family Literacy Workshops began. A post-workshop survey was administered two to three weeks after the last workshop in order to measure any changes in these outcomes. This survey was also administered by phone. Note that certain pre-intervention measures were collected retrospectively during the post-workshop survey.

A third anonymous survey was administered at the last workshop to elicit parents' opinion on the logistical aspects of the workshops, particularly regarding the physical environment, the quality of exchanges, and the quality of the practitioners' delivery. All these data served to evaluate the quality of implementation for the Family Literacy Workshop component, the results of which are detailed in Chapter 6 of this report. The response rate was excellent for the pre- and post-workshop surveys, but relatively low for the logistical survey due to the low rate of participation in that workshop (see Table 3.5).

Table 3.5: Response Rates for Family Literacy Workshop Surveys

	<i>Pre-intervention</i>	<i>Logistical Survey</i>	<i>Post-intervention</i>
Surveys administered	135	92	134
Absent	8	50	8
Withdrawals	6	7	7
Response rate	90.6%	61.7%	89.9%

Note: These rates include the communities of Edmonton and Saint John.

Child Surveys

Every survey period for parents was paired with a children's assessment. In all, seven assessments were administered to children every four months for the first two years of the Readiness to Learn project with an average response rate of 95.5% (see Table 3.6).

Table 3.6: Response Rate for Child Assessments

	Assessment periods						
	Baseline assessment	+ 4 months	+ 8 months	+ 12 months	+ 16 months	+ 20 months	+ 24 months
Children assessed	346	347	342	337	339	334	334
Absent	4	3	6	5	2	4	4
Withdrawals	6	6	8	14	15	18	18
Response rate	97.2%	97.5%	96.1%	94.7%	95.2%	93.8%	93.8%

3.5.2. Child Measures

Throughout the waves of data collection, a range of tools was used to monitor child development for several important dimensions associated with school readiness. The schedule and tools for each assessment period are found in Appendix A.

Évaluation de la petite enfance — Appréciation directe (Willms, 2007)

The ÉPE-AD measures four domains of school readiness as well as a fifth domain designed specifically for the purposes of the Readiness to Learn project:

- (Domain A) Awareness of Self and the Environment;
- (Domain B) Cognitive Skills;
- (Domain C) Language and Communication;
- (Domain D) Physical and Motor Development; and
- (Domain E) Awareness and Engagement in Francophone Culture.

Each domain entails a number of questions presented in ascending order of difficulty. Two versions of the ÉPE-AD were used as part of this project. The initial ÉPE-AD was used until the fifth assessment (+16 months) while the “modified ÉPE-AD,” the version reworked by SRDC, was used for the sixth and seventh assessments (+20 and +24 months). See Appendix B for a comparison of content for the tool’s three versions.

Domain A, Awareness of Self and the Environment, consists of 16 questions that determine the degree to which the child can recognize and identify the elements in his or her environment. For example, the child is asked to name colours, parts of the body, certain occupations and their role, and his or her date of birth.

Domain B, Cognitive Skills, consists of 17 questions measuring various logico-mathematical aspects. For example, the child is asked to count, form groups using various objects, compare different shapes, and distinguish shape sizes. The child is also asked to name a few letters of the alphabet, to identify their sound, and read eight words.

Domain C, Language and Communication, consists of 14 items measuring the child’s ability to communicate and understand. The evaluator asks the child to point out the image representing a word that she says, repeat seven-syllable sentences, answer questions with complete sentences, use pictures to tell a story, and show the meaning of four action words. This is the only domain

administered to all children in French since this domain is part of the decisional tree used to determine the language in which the assessment is administered.¹⁵

Domain D, Physical and Motor Development, consists of 16 items measuring the child's ability to carry out certain fine and gross motor skill activities. Fine motor skills encompass several elements of early writing, such as the ability to trace letters, draw a straight line or colour between the lines while holding a pencil correctly. Gross motor skills concern the child's ability to move his or her body, for example, by jumping over an object, balancing on one foot or hopping up and down on one foot three times in a row. Note that a Canadian study based on NLSCY data indicates that after age three, there is little difference in this domain among children who are developing normally. These findings were confirmed by the fifth assessment (+16 months), which is why Domain C was not measured at the sixth and seventh assessments conducted at +20 months and +24 months.

Items are ranked based on a five-point scale for Domains A, B and C and on a 4-point scale for Domain D. The first three domains require the use of test charts or objects that the child must point to or manipulate. When a test is being administered, a child sometimes reaches a "plateau" in a given domain when the questions become too difficult. A domain assessment is stopped when the child scores "0" or "1" for three consecutive items. The evaluator then moves on to the next domain. It takes about 45 minutes to an hour to administer the test.

Up until the +16 months assessment, a child can be assessed either in English or in French, depending on his or her scores for Domains E and C. Domain E, which was designed specifically for the Readiness to Learn project and measures the child's awareness of and engagement in Francophone culture, was used more to establish a friendly rapport with the child and determine the language of testing.¹⁶ Its use was discontinued at the fifth assessment (+16 months) because the children were too young to answer the questions clearly, making the domain unreliable.¹⁷

The ÉPE-AD has good psychometric properties. Results of factorial analyses for the ÉPE-AD confirmed the unidimensionality of each domain in the French version of the test, as theorized.¹⁸ The internal consistency of each dimension ranged from acceptable to very good, with Cronbach alphas ranging from 0.61 to 0.94 for the French version of the test and from 0.79 to 0.94 for its English version (see Table 3.7).

¹⁵ See Appendix D for the ÉPE-AD administration procedure as well as the decisional tree used to determine testing language.

¹⁶ Domain E includes questions on the child's favourite books, television shows and songs, the language of these resources (English or French), as well as the languages spoken with their parents and friends.

¹⁷ See the *Readiness to Learn in Minority Francophone Communities: Reference Report* (Legault et al., 2014) for an explanation of the reasons for this decision.

¹⁸ A factorial analysis of the test's English version was impossible because too few individuals took the test in English. At least 160–170 cases per domain would have been ideal to ensure the statistical validity of calculations (Tabachnick and Fidell, 2006).

Table 3.7: Cronbach Alpha for ÉPE-AD Domains by Assessment Period

	Cronbach alpha [α (n)]						
	Baseline assess't	+ 4 months	+ 8 months	+ 12 months	+ 16 months	+ 20 months	+ 24 months
<i>Domains administered in French</i>							
A	0.92 (301)	0.91 (300)	0.91 (301)	0.90 (311)	0.90 (329)	0.89 (334)	0.89 (334)
B	0.86 (301)	0.87 (300)	0.86 (301)	0.84 (311)	0.85 (329)	0.82 (334)	0.85 (334)
C	0.92 (346)	0.93 (347)	0.94 (342)	0.93 (337)	0.92 (339)	0.63 (334)	0.61 (334)
D	0.89 (301)	0.92 (300)	0.93 (301)	0.93 (311)	0.93 (329)	n/a	n/a
<i>Domains administered in English</i>							
A	0.85 (45)	0.89 (47)	0.88 (41)	0.91 (26)	0.86 (10)	n/a	n/a
B	0.79 (45)	0.89 (47)	0.89 (41)	0.89 (26)	0.86 (10)	n/a	n/a
C	0.87 (45)	0.85 (47)	0.89 (41)	0.90 (26)	0.85 (10)	n/a	n/a
D	0.83 (45)	0.91 (47)	0.88 (41)	0.94 (26)	0.94 (10)	n/a	n/a

Épreuve de dénomination de Gardner (Ska, 1995)

The *Épreuve de dénomination de Gardner* (1979) is the validated and standardized French translation of the Expressive One-Word Picture Vocabulary Test (EOWPVT) for children from kindergarten to Grade 2. This tool measures a child's expressive vocabulary. Thus, instructions and the items' order of difficulty were adjusted for a Francophone population. The exercise consists of showing the child a series of pictures and asking him or her to name the word associated with the image. The test is stopped after six consecutive errors and takes between 10 and 15 minutes to administer. The advantage of this test is that it is very sensitive to differences in French-language proficiency. The total score on the test is the number of pictures named correctly by the child in French. The test also takes into account certain incorrect responses by the child, that is, English words or regionalisms. These incorrect answers are gathered for information purposes only and are not used to calculate the child's score. This tool was used for the sixth assessment (+20 months).

Échelle de vocabulaire en images Peabody — Révisé (ÉVIP-R; Dunn, Thériault-Whalen, and Dunn, 1993)

The *Échelle de vocabulaire en images Peabody — Révisé* (ÉVIP-R) is the validated French translation of the Peabody Picture Vocabulary Test – Revised (PPVT-R), which measures a child's receptive vocabulary. The test resembles a game: the child hears a word spoken out loud and must identify the correct picture from among four alternatives. The test stops once a child makes six errors in eight tries. According to test rules, the child's starting point for the test is based on his or her age and performance. For the purposes of the Readiness to Learn project, a common starting point (i.e., the 15th question, which is the starting point for children at age

three-and-a-half) was established for all children, whatever their age or performance level. Since the project's goal is not to compare children's performance to a "standard,"¹⁹ using a common starting point that is relatively easy enabled us to capture the receptive vocabulary level for children from exogamous families.

The ÉVIP-R score represents the total correct answers given by the child as of the 15th question. This method of calculating the ÉVIP-R score differs slightly from that recommended in the test manual (Dunn, et al., 1993, p. 13–15). We observed a correlation greater than .99 between SRDC's chosen method and that proposed in the manual. This means there is no loss of information. The SRDC method of calculating the child's ÉVIP-R score has several advantages, in particular, it: a) maximizes the variance; and b) allows for the inclusion of children who were unable to take the test because they failed the practice runs (by scoring zero). This tool was used for the seventh assessment (+24 months).

Vocabulary Subscales (ÉPE-AD)

In winter 2009, SRDC reworked the structure of the ÉPE-AD scales to create two vocabulary subscales in order to detect subtle differences in children's language skills. The first subscale measures expressive vocabulary, that is, the child's ability to say the word associated with the picture he or she is shown. This scale contains six items from Domain A and two items from Domain C. Internal consistency is very good for Expressive Vocabulary items with Cronbach alphas ranging from 0.77 to 0.88 for the English and French versions.

The second subscale measures receptive vocabulary and consists of two items from Domain A and three items from Domain C. The child is shown a series of pictures and must identify the picture associated with the word said aloud by an evaluator. Internal consistency for items measuring Receptive Vocabulary in the French version is acceptable with Cronbach alphas ranging from 0.61 to 0.74. However, internal consistency for the English version of the Receptive Vocabulary subscale is less acceptable, with several alphas of less than 0.50 (see Table 3.8).

Table 3.8: Cronbach Alphas for ÉPE-AD Vocabulary Subscales by Assessment Period

	Cronbach alpha [α (n)]						
	Baseline assess't	+ 4 months	+ 8 months	+ 12 months	+ 16 months	+ 20 months	+ 24 months
<i>Domains administered in French</i>							
Expressive Vocabulary	0.86 (301)	0.87 (300)	0.88 (301)	0.87 (311)	0.86 (329)	0.86 (334)	0.84 (334)
Receptive Vocabulary	0.74 (301)	0.68 (300)	0.68 (301)	0.61 (311)	0.64 (329)	n/a	n/a
<i>Domains administered in English</i>							
Expressive Vocabulary	0.81 (45)	0.85 (47)	0.86 (41)	0.84 (26)	0.77 (10)	n/a	n/a
Receptive Vocabulary	0.41 (45)	0.73 (47)	0.38 (41)	0.54 (26)	0.46 (10)	n/a	n/a

¹⁹ Note that current standards for the test's French version were established in 1990 and have not been updated. Their validity is therefore unknown.

Early Years Evaluation — Direct Assessment, Modified Version (SRDC, 2009)

In winter 2009, SRDC found there was a high potential for a ceiling effect after the first analyses of ÉPE-AD results for the first cohort and the score projections for the fifth assessment. A ceiling effect would hinder the ability to continue following the evolution in child development and, by extension, would make it impossible to test for differences among children in the three treatment groups. HRSDC therefore hired the test developer to create new, more difficult questions so that the children's developmental trajectory could be tracked in Domains A, B, and C until the end of the project's second year (the seventh assessment). The developer also eliminated the test's easier questions to maintain the test's administration time. The "extended version" of the ÉPE-AD (Willms, 2009) was ready in May 2009.

At the same time, SRDC restructured the questions that make up the extended version of the ÉPE-AD. The purpose of the restructuring was to obtain a more subtle measurement of the children's language skills while maintaining the ability to track their developmental trajectory based on three of the four ÉPE-AD domains: Awareness of Self and the Environment (Domain A), Cognitive Skills (Domain B), and Language and Communication (Domain C).²⁰ The new tool, called "the modified ÉPE-AD" also led to the Expressive Vocabulary and Receptive Vocabulary measuring presented earlier. The modified ÉPE-AD consists of:

- 15 questions in Domain A mainly on expressive vocabulary and general knowledge;
- 12 questions in Domain B measuring phonological awareness (a child's ability to play with the sounds that make up words) and numeracy concepts (a child's ability to play with numbers). Two other items in Domain B were administered in a slightly different way than in the extended ÉPE-AD. Thus, knowledge of the sounds of all alphabet letters was measured using a booklet showing uppercase letters in ascending order of difficulty for young Francophones. The knowledge of the name of all alphabet letters was measured using a booklet showing lowercase letters in ascending order of difficulty for young Francophones.
- Finally, two questions in Domain C were kept to continue the Expressive Vocabulary scale and two new, more difficult questions from the extended ÉPE-AD were added to measure the child's phonological awareness.

The initial ÉPE-AD was used until the fifth assessment (+16 months) while the "modified ÉPE-AD," the version reworked by SRDC, was used for the sixth and seventh assessments (+20 and +24 months). See Appendix B for a comparison of content for the tool's three versions.

As part of the Readiness to Learn project, it is important to follow the children's developmental trajectory over the two years of the program. Establishing continuity in the dimensions measured by two tools is based on an examination of the pattern of correlations between the subscales measuring the same dimension in each tool. A strong correlation (i.e., of 0.75 or higher) between subscales means they respectively measure the same concept. An examination of the correlation pattern in Table 3.9 confirms the presence of strong correlations between Domains A and B of the initial ÉPE-AD (fifth assessment) and those of the modified ÉPE-AD (sixth assessment), that is, 0.83 and 0.75 respectively. The exception is the

²⁰ For more about the reworking of the ÉPE-AD and its theoretic bases, refer to the report presenting a review of the direct measurement tools for children's developmental dimensions submitted to HRSDC in July 2009.

Communication domain (Domain C), for which a correlation of 0.63 is observed. However, this lower than expected result is offset by the good correlations found between Domain C of the initial ÉPE-AD and the Expressive Vocabulary subscale created using the ÉPE-AD (0.82), as well as between Domain C of the initial ÉPE-AD and Gardner's Expressive Vocabulary scale (0.77). There is also a strong correlation between Domain C of the initial ÉPE-AD and the Domain A scale of the modified ÉPE-AD. These results combined allow us to conclude that use of the modified version entails a negligible loss of information. We are therefore able to monitor the children's developmental trajectory over the two years of the program.

Table 3.9: Pearson Correlation Coefficients for ÉPE-AD Scales and for Expressive and Receptive Vocabulary Scales

		<i>Initial ÉPE-AD: Fifth assessment (+16 months)</i>				
		Scales	A	B	C	Expressive Vocabulary Subscale (ÉPE-AD)
Modified ÉPE-AD: Sixth assessment (+20 months)	A		0.83 (309)	0.63 (322)	0.83 (330)	0.80 (311)
	B		0.51 (309)	0.75 (322)	0.47 (330)	0.45 (311)
	C		0.61 (309)	0.61 (322)	0.63 (330)	0.60 (311)
	Expressive Vocabulary subscale (ÉPE-AD)		0.79 (309)	0.49 (322)	0.82 (330)	0.81 (311)
	Expressive Vocabulary (Gardner)		0.75 (306)	0.52 (319)	0.77 (327)	0.76 (308)
	Receptive Vocabulary (ÉVIP-R)		0.68 (310)	0.53 (323)	0.71 (330)	0.68 (312)

Note: The modified ÉPE-AD does not include a Receptive Vocabulary subscale, which the test developer eliminated in the extended version of the ÉPE-AD because it was too easy. Data analysis for the Readiness to Learn project confirmed that these data were too easy and therefore did not allow for comparison of the children.

3.5.3. Parent Measures

The parent follow-up surveys allow us to determine the degree to which the child's environment changed in the four months prior to the survey. First, the surveys measure the Frequency of Literacy Activities and Language of Literacy Activities as well as other factors that affect children's school readiness and the development of their Francophone identity. Second, they systematically verify changes in type of childcare and the language used in the new type of childcare so that the analyses can factor in these changes. Third, they aim to document changes in family composition to determine whether the child's usual language environment at home has been affected by those changes.

The scales used in the parent surveys of the Readiness to Learn project are derived from studies on Canada's Francophone populations, such as the NLSCY, the *Étude longitudinale du*

développement des enfants du Québec [longitudinal study of child development in Quebec] (ELDEQ; Institut de la statistique du Québec, 2003) and the Survey on the Vitality of Official-Language Minorities (SVOLM, Statistics Canada, 2006). The questions selected for the surveys of the Readiness to Learn project were those relevant to the children's age group.

Note that only the scales included in the impact analyses are detailed in this section. Of course, this means that the list of scales differs somewhat from those in the *Readiness to Learn in Minority Francophone Communities: Reference Report* (Legault et al., 2014). Several other concepts were measured but were not retained for the impact analyses. The decision to include them or not is based on preliminary analyses, the results of which are found in Chapter 5.

Sociodemographic Characteristics

The sociodemographic characteristics of children and parents stem from questions asked on the parental consent form (for example, the child's gender and date of birth) and in the baseline survey. The follow-up surveys were an opportunity to make any necessary changes to this initial profile. This section identifies the sociodemographic variables used in the impact analyses.

Information such as the mother's and father's level of education, the mother's age at birth of the child and family income were gathered from parents in the baseline survey. Family income was subsequently requested on two other occasions. An open-ended question on income was first asked at the fifth assessment (+16 months). Since a number of parents refused to disclose that information, the same question was asked again at the seventh assessment (+24 months) specifically to the parents who had refused to answer, but this time they were given a choice of income brackets for their response.

Several questions were intended to establish the linguistic profile of participating families. Among others aspects, respondents had to identify their mother tongue, knowledge of the official languages, as well as the languages spoken with the child and at home. Spouses were asked the same questions. There are two standard definitions of Canada's Francophone population. Statistics Canada²¹ calculates the population that reports French as their mother tongue, i.e., for whom French is the first language learned and still understood. The second method is to calculate the "first official language spoken" (FOLS), a score that takes into consideration knowledge of both official languages, mother tongue and the language most often spoken at home (Forgues and Landry, 2006).

Every follow-up survey begins with a series of five questions exclusively for parents who experienced a change in family composition since the last survey was completed. Changes in family composition in comparison with answers in the baseline survey (e.g., household size, single-parent families, older siblings, twins, younger siblings) are factored into the impact analyses.

Parents' immigrant status was established by means of a question in the seventh assessment (+24 months). This survey also updated the sociodemographic data for the respondent's spouse. Note that in the baseline survey, the information obtained defined the characteristics of the spouse who was the child's biological parent, even if that parent had no contact with the child. In order to clarify the results, these questions were asked again at the seventh assessment (+24

²¹ See the definition of the Francophone population on the Statistics Canada website at <http://www.statcan.gc.ca/concepts/definitions/language-langue-eng.htm>.

months) to obtain information on the second adult living at home with the child and who is actively involved in caring for the child.

Language Habits

Several questions on language habits were included in the baseline survey and follow-up surveys. Respondents chose the category most representative of their experience from among several options (for example, if they speak English only, English and French, French more than English, etc.). These categories were combined to create a five-point scale representing a language continuum where a score of 1 means that the respondent speaks “English only,” a score of 3 means that he or she speaks “French and English and/or another language,” and a score of 5 means that he or she speaks “French only.” Impact analyses used this language continuum based on a five-point scale. For comparative analysis purposes (Readiness to Learn project compared with the SVOLM), parents’ language habits were grouped over a three-point continuum due to the few cases in certain cells. In these analyses, a score of 1 corresponds to “French only,” a score of 2 means “French and English equally OR French and another language,” and a score of 3 means “English only OR English and another language OR another language.”

A first language continuum was generated representing the language usually spoken by the mother to the child. A second language continuum was calculated by taking the mean scores for four questions on the languages spoken by the child with his or her mother, father, friends and siblings at home, as well as outside the home (Cronbach alpha of 0.95).

Two indices representing the language usually spoken at home were derived. A first index was derived by combining the mother’s FOLS with the father’s. A second index was created by combining the language usually spoken by the mother to the child with the language usually spoken by the father to the child. These two indices were used to determine the child’s home language environment. In addition, the FOLS was used to determine whether the child lives in a home that is endogamous Francophone, endogamous Anglophone, or exogamous.

Family Processes

The baseline survey gathered information on family processes. Some of these processes were measured again in follow-up surveys. A total of five family processes were measured using four scales: parenting style, family functioning, depression, and literacy activities.

The ***Positive Parenting*** scale consists of five items measuring the frequency of positive contact between parent and child (for example, the number of times the parent praises the child, laughs with him or her or physically expresses affection). Items are scored on a five-point scale, where 1 means “never” and 5 means “several times a day.” The scale’s internal consistency is acceptable at 0.62.

The ***Authoritarian Parenting*** scale consists of four items measuring the degree of supervision and discipline. It tells us, among other aspects, whether the child complies with the punishments imposed or whether punishments vary depending on the parent’s mood. Items are scored on a five-point scale where 1 means “never” and 5 means “almost always.” Internal consistency is acceptable with a Cronbach alpha of 0.61.

The ***Family Functioning*** scale, which consists of eight items, assesses the quality of bonds within the family. The content of the items concerns emotional openness, expression of feelings,

and interaction within the family. The choice of answers for the items on the scale ranges from 1, “absolutely disagrees,” to 4, “fully agrees.” The Family Functioning scale shows excellent internal consistency with a Cronbach alpha of 0.81.

The **Depression** scale measures, for example, the number of times the respondent feels depressed or unhappy, cries or feels alone, or does not enjoy life. The scale’s eight items cover a four-point range, where 1 means that the respondent felt this way “rarely or never” and 4 means that he or she feels this way “most of the time or all of the time.” The Depression scale has a very high internal consistency with a Cronbach alpha of 0.80.

The **Literacy Activities** scale consists of five items measuring the frequency of reading, writing and numeracy activities. More specifically, the items measure the frequency at which parents tell stories to their child without a book, sing songs with them, teach them to write or trace letters or numbers, teach them to read words, and encourage them to use numbers in their daily activities. These items are scored on a five-point scale, where 1 means an activity is never done and 5 means an activity is performed several times a day.

Every question on literacy activities is followed by a question verifying the language used during that activity. The respondents choose the category most representative of their experience among several options (for example, whether they speak English only, English and French, French more than English, etc.). These categories were combined to create a five-point score representing a language continuum *for this activity* where a score of 1 means the activity takes place in “English only,” a score of 3 means the activity takes place in “French and English and/or another language,” and a score of 5 means the activity takes place in “French only.” A **Language of Literacy Activities** scale is obtained by averaging the scores for each question. This scale’s internal consistency is very high with a Cronbach alpha ranging from 0.88 to 0.94.

The questions that make up the Literacy Activities scale and the Language of Literacy Activities scale were asked at each survey, including the baseline survey. Internal consistency for these scales was acceptable enough at each survey period to allow for confidence in their homogeneity (see Table 3.10).

Table 3.10: Cronbach Alpha for the Literacy Activities Scale and the Language of Literacy Activities Scale for Each Survey Period

	<i>Cronbach alpha [α (n)]</i>						
	Baseline	+4 months	+ 8 months	+12 months	+ 16 months	+20 months	+24 months
Literacy Activities	0.52	0.51	0.61	0.72	0.73	0.72	0.69
Language of Literacy Activities	0.94	0.93	0.93	0.92	0.89	0.91	0.88

Community Dimension

The Social Capital scale arises from questions asked in the baseline survey. This scale consists of five items that measure the level of mutual assistance among neighbours in the participant’s community. These items establish the extent to which neighbours are willing to help neighbours, work together to solve a common problem, and ensure children’s safety. Items are

scored from 1, “absolutely disagrees,” to 4, “absolutely agrees.” The scale’s internal consistency is excellent with a Cronbach alpha of 0.80.

Identity-related Dimension

The baseline survey included a number of questions on perceptions regarding the *Subjective Vitality of the Francophone Community, Engagement in Francophone Culture*, and the *Sense of belonging to Francophone Culture*. Among the many dimensions of identity, only the sense of belonging to Francophone culture was deemed useful for the impact analyses. The purpose of the question was to determine the cultural group with which parents identify, that is, Francophones, Anglophones, both, or another group.

3.5.4. Parents’ Knowledge, Attitudes and Beliefs

Surveys before and after the Family Literacy Workshops involved a series of questions intended to identify the nature and scope of changes to the perceptions and behaviours of parents in the Program Daycare group. For the purposes of evaluating the Family Literacy Program, several measurements were created based on the questions used in family literacy studies. Some of these studies were conducted by the Centre for Interdisciplinary Research on Citizenship and Minorities (CIRCEM; LeTouzé, 2006). Another source of inspiration was the study on how family literacy affects families in minority contexts, conducted by the Coalition francophone pour l’alphabétisation et la formation de base en Ontario (Benoît, n.d.). Measurements were also created using the post-intervention questionnaire for the school transition program of Edmunstun’s School Board District 3 (Gauthier St-Onge, n.d.) and using the NLSCY (Statistics Canada and HRSDC, 2006). Finally, the retrospective approach used to come up with the questions was inspired by a retrospective survey created by Lamb and Tschillard (2005).

The choice of constructs for studying the effects of the Family Literacy Program is partly based on the reasoned action model developed by Fishbein and Ajzen (for an introduction to the model, see Brigham, 1991). This model links beliefs and attitudes with behaviour prediction. According to the theory, people’s beliefs about a behaviour’s influence both their attitudes toward that behaviour and their perceptions of control (a concept similar to sense of self-efficacy). Attitudes and perceptions of control in turn influence the intention to adopt that behaviour. Finally, intention determines the probability that the behaviour will occur. In the Readiness to Learn project, we measured participants’ beliefs and attitudes regarding the dimensions addressed by the Family Literacy Workshops as well as their sense of self-efficacy. We also measured participants’ knowledge of the main topics of the Family Literacy Workshops. Although it is not part of the theory of reasoned action, level of knowledge also affects people’s attitudes about an action. Finally, a modeling scale was added to capture the importance of parents’ literacy activities on children’s attitudes and behaviours. In fact, several studies have shown the positive effect of a rich family literacy environment, such as the presence of written or writing material, as well as parental behaviours that value reading or writing, like reading or writing in front of children (Dionne, 2007; Sénéchal and LeFevre, 2002; Jordan, Snow, and Porche, 2000). The four concepts specific to pre- and post-workshop surveys are detailed in the following paragraphs.

Four questions measure certain parental *Beliefs and Attitudes* addressed in the Family Literacy Workshops. These beliefs correspond to the program’s key themes, specifically: the

parent as the child's first educator, educator-parent cooperation, the Francophone cultural environment at home, and the appropriate age for reading to a child. The respondent indicates the degree to which he or she agrees with each statement using a four-point scale where 1 means "absolutely disagrees" and 4 means "absolutely agrees." These questions were asked in the pre- and post-workshop surveys. However, it should be noted that these questions do not form a true scale and, therefore, the information gathered is descriptive in nature.

The ***Perceived Self-Efficacy*** scale assesses parents' perceived ability to carry out certain actions discussed in the Family Literacy Workshops. Thus, parents are asked how confident they feel about their ability to: manage their child's emotions and behaviours, speak with him or her in French, help him or her learn, and help him or her develop a Francophone cultural identity. Each of these dimensions is the subject of two questions. The first is retrospective in nature, asking parents how confident they felt *before* the Family Literacy Workshops. The second measures the parents' confidence *after* the Family Literacy Workshops end (i.e., at the time of the survey). These items are scored on a five-point scale, where 1 means "little or no confidence" and 5 means "complete confidence." Internal consistency is good with a Cronbach alpha of 0.70 for the subscale measuring retrospective perceived self-efficacy and of 0.60 for the subscale measuring perceived self-efficacy after the workshops.

The ***Knowledge*** scale assesses parents' level of knowledge for five aspects addressed in the Family Literacy Workshops: child development, how to enhance child development and learning at home, how to help a child prepare for school, the work of daycare educators, and French-language community services or resources. As with the Perceived Self-Efficacy scale, each aspect was the subject of two questions: one retrospective (*How would you describe your level of knowledge on the topics before the Family Literacy Workshops?*) and the other after the Family Literacy Workshops (*How would you describe your level of knowledge on the topics now, after the Family Literacy Workshops?*).²² These items were scored on a five-point scale, where 1 means "no knowledge" and 5 means "excellent knowledge." Internal consistency is acceptable with a Cronbach alpha of 0.79 for the subscale measuring the perceived level of knowledge before the workshops and of 0.67 for the subscale measuring the perceived level of knowledge after the workshops.

The ***Modeling Behavior*** scale consists of five questions based on the literacy activities model. It aims to measure the parent's example to his or her child with respect to the importance of reading and writing. Item content concerns, among other aspects, the frequency at which parents read and write in front of their child, as well as the frequency at which parents leave paper and pencils at their child's disposal. Answers to questions are scored on a four-point scale where 1 means "never" and 4 means "several times a day." The scale was filled out by parents in the Program Daycare group at the time of the pre-intervention survey, the post-intervention survey administered three weeks after the workshops ended (i.e., in March 2008 for the first cohort and March 2009 for the second cohort) and a final survey administered seven months after the workshops ended (i.e., in October 2008 for the first cohort and October 2009 for the second cohort). Internal consistency is acceptable with a Cronbach alpha of 0.75, 0.65 and 0.64 observed respectively for each administration.

²² For an informed discussion of the advantages of using the retrospective aspect as part of training, see Lamb and Tschillard (2005).

3.5.5. Dosage

The number of hours of child care per week is an important variable to consider in measuring the program “dosage” that children receive at daycare. To gather this information, SRDC used the attendance record that the educators already fill out as part of their duties, including the children’s time of arrival and departure. The number of hours the child spent at daycare is then compiled on a weekly basis and sent once a month to SRDC’s Ottawa office. For the purposes of the first-year impact analyses, the hours spent at daycare were analyzed using the average hours spent at daycare per week for the four months immediately prior to the children’s assessment. The second-year impact analyses use the average hours spent per week in the first eight months following the intervention (see Section 8.3).

As regards the Family Literacy Workshops, the attendance of parents, children and other participating adults was compiled for each workshop. Dosage for workshops is calculated using the total workshops taken by at least one adult (see Section 6.3 for the dosage results of the Family Literacy Workshops).

3.6. STUDY HYPOTHESES

3.6.1. Impact of the Daycare Component on Children’s Language Skills

The main objectives of the Daycare Program are school readiness, enrichment of child language skills, and the development of a Francophone cultural identity. These are the areas in which we expect to observe the program’s greatest effects. The hypotheses in this respect are:

1. Children in the Program Daycare group will have higher scores in three of the four domains measuring school readiness, including Language and Communication, Cognitive Skills, and Awareness of Self and the Environment, than children in the comparison groups. There is expected to be no difference between the Program Daycare group and the comparison groups as regards Physical and Motor Development, since the tested program does not target this dimension.
- This hypothesis aims to directly test the main research question. The study’s internal validity is assured by inclusion of the Comparison Daycare group since the influence of formal daycare on child development is a subject unto itself. It is also assured by adding the second comparison group, which was designed to factor in how an informal childcare setting affects the development of a child’s French-language skills.
- Finally, this hypothesis takes into account other existing conditions in the community that may affect child development. For example, the community environment plays an important role in child development to the extent that it offers resources (e.g., parks, wading pools, bike paths) and services (e.g., bookstores, libraries, swimming lessons, etc.).
2. Children in the Program Daycare group will perform better in terms of language skills than children in the comparison groups.
3. The tested program will have a greater impact on the language skills of children in the Program Daycare group from exogamous homes or Anglophone endogamous homes.

4. The magnitude of the Daycare Program's impact on dimensions of children's school readiness will be influenced by the children's level of exposure to the program. Thus, greater exposure to the tested program will be associated with better performance by children in the Program Daycare group in three of the four domains measuring school readiness, that is, Language and Communication, Cognitive Skills, and Awareness of Self and the Environment, than by children in the comparison groups. No difference between the Program Daycare group and the comparison groups is anticipated in terms of Physical and Motor Development since the tested program does not target this dimension.
5. Daycares in the Program Daycare group will be characterized by program fidelity and quality indices that are higher than those of comparison daycares.
6. The magnitude of the Daycare Program's impact on the dimensions of school readiness and language skills will vary based on the extent of the tested program's fidelity and quality.

3.6.2. Impact of the Daycare Component on Children's Identity and Culture

The tested program aims to influence the various contexts necessary to develop the identity and culture of young children. The importance of measuring cultural identity stems from research findings to the effect that children who are exposed to French in several spheres of their life have a stronger sense of identity and of belonging to the Francophone community (Landry and Allard, 2000). However, these studies measured the concept of cultural identity at adolescence. This choice is appropriate according to Erik Erikson's theory of self (1994), which states that identity begins to form around age 12 and matures around age 25. As such, a direct measure of cultural identity at this very young age is impossible.

Instead, children's degree of *exposure* to French can be measured through their parents. In fact, parents are in the best position to inform us about their child's language environments in the settings that he or she frequents and that are of interest for the Readiness to Learn project, that is, family, daycare, and access to French-language services and resources. The collected information will serve as control variables in the impact analyses in order to better identify how the tested program affects the development of young children's language skills.

3.6.3. Impact of the Family Literacy Component on Parents' Attitudes and Behaviours

The content of the Family Literacy Program focuses on four aspects:

- the parent as his or her child's first educator;
- early reading, writing and numeracy activities that foster a child's school readiness;
- the importance of maximizing the child's exposure to French; and
- parent-educator cooperation in the child's learning.

These are the aspects for which we expect to observe the program's greatest effects. The hypotheses are as follows. After the Family Literacy Workshops and in comparison with parents in the comparison groups, parents in the Program Daycare group will report that they:

- much more frequently perform various literacy activities with their child;
- speak to their child in French much more often during literacy activities.

Moreover, we expect that after the workshops, parents in the Program Daycare group will report that they have:

- significantly more knowledge about their child's development dimensions, how to help him or her prepare for school, and French-language community services and resources;
- a higher sense of self-efficacy, especially as regards helping their child prepare for school;
- much more frequent opportunities to model literacy activities;
- experienced a significant change in their beliefs, especially as regards the elements addressed by the family literacy component, in particular: the parent as his or her child's first educator, educator–parent cooperation in child learning, the importance of the Francophone cultural environment at home, and the right age to be reading with their child.

4. Methodology for the Implementation Review

This chapter contains the evaluation matrix used to review the integrity of implementation for the Daycare Program and the Family Literacy Program. A review of program implementation serves a number of objectives. Foremost, it shows how a program developed on paper translates to reality. Typically, an implementation study is used to understand barriers, facilitators, and adjustments required to facilitate the field taking ownership of the program. It is also an important tool to better understand and nuance the program's effects. Thus, a growing number of studies show that an intervention's degree of implementation is linked to the program's observed effects on participants (Charlebois, Brendgen, Vitaro, Normandeau, and Bourdreau, 2004; Conduct Problem Prevention Research Group, 1999; Dane and Schneider, 1998; Durlak and DuPre, 2008). The program's impact will differ if several elements were not implemented or if the quality of implementation is mediocre. In this regard, Durlak and DuPre (2008) concluded, using data from five meta-analyses, that the magnitude of an intervention's effect is two to three times greater when the program is implemented as planned. These findings support the importance of properly documenting and measuring the degree of implementation for the program being assessed. A rigorous evaluation of program implementation is crucial when the intervention in question will be the basis of future practices and policy. Results thus obtained must be tied in closely with both the quantity and the quality of services to the population, in addition to considering other factors, such as any program adjustments done in order to adapt to the field.

In addition to the study's approach, data from various sources was triangulated to come up with an overall portrait of implementation for the Readiness to Learn project. Triangulation was made possible by adopting a three-tiered methodological approach (Patton, 1990). In other words, we took advantage of various measurement tools, both quantitative (as with the fidelity and quality scales) and qualitative (as with the participant interviews). The benefit of this approach, also known as the *mixed-method model*, is that it documents program effects in several ways and allows for a more in-depth program assessment (Patton, 2008). This methodology also gave different stakeholders a voice, resulting in a range of perspectives on project implementation. Data were then gathered using a number of measurements instruments to evaluate a given phenomenon. These diverse data enhance the validity and reliability of findings due to the complementarity of the data collected.

4.1. ANALYSIS MODEL FOR THE QUALITATIVE DATA OF THE IMPLEMENTATION REVIEW

A great many factors describe a program's degree of implementation. The first model to structure these elements was that of Dane and Schneider (1998). This model verifies a program's integrity based on the five following dimensions: **fidelity**, that is, correspondence between the applied intervention and the planned program; **quality of program implementation**, for example, the practitioner's preparation and enthusiasm; **dosage**, that is, participant exposure to the program; **participation**, that is, participant response to the program, measured by their level of participation and enthusiasm; and, finally, **differentiation between the implemented**

program and other interventions (i.e., the program's new contribution versus existing benefits).

Two other studies have suggested adding further elements to this model. Carroll and colleagues (2007) suggest including program complexity and **strategies that facilitate implementation**. One final dimension added to the model concerns the **participants' perceived benefits**. We have included this dimension because it provides information about the benefits perceived by stakeholders involved in the program. This perception will affect willingness to ensure that the intervention continues.

4.2. DATA COLLECTION TOOLS

Within the framework of the Readiness to Learn project, an array of quantitative and qualitative tools was used to gather information on the implementation of the tested program from a wide range of people.

4.2.1. Implementation Review for the Daycare Component

The implementation review identified gaps between theory and practice, bottlenecks or other implementation problems that require correction, under-used program elements, as well as participant movement among program elements. It also identified differences and similarities in program delivery at the six program daycares for the first cohort and the two program daycares for the second cohort, in addition to allowing for comparative analyses of the tested program and the programs delivered at comparison daycares.²³ Several measures are used, each completing the others in terms of the type of information elicited, the level of detail, and the source of information. This allows for triangulation of the information from a range of perspectives, whether the educator's, the observer's, or the trainer's. The sections below describe in detail, respectively, the measurement instruments developed to evaluate the implementation of the Daycare Program, including the daycare observation grid, the protocol for in-depth educator interviews, the educator's journal, and children's daycare attendance.

Daycare Observation Grid

The goal of daycare observations was to collect relevant data on program delivery (as defined by the objectives of the Readiness to Learn project). According to Durlak and DuPre (2008), observation measures are more likely to relate to intervention results than self-reported measures. Among the first elements to observe is the fidelity of project implementation. This consists of ensuring that all the program's essential elements were implemented at the daycares. The observation protocol thus required that observers verify the presence of certain elements specific to the tested program (such as whether parents could see a weekly program or whether a routine chart depicting the day's routine was posted for children). The observers also had to describe how activities were conducted, as well as provide information about children's reactions during activities.

Observations also provided information on the quality of the various childcare settings. To this end, certain subscales were borrowed from the scale most used in studies administered to

²³ Note that data from one of the six daycares in the first cohort were not included in the implementation analyses because the program was not implemented in that community.

assess the quality of childcare settings in North America, that is, the Early Childhood Environment Rating Scale, Revised, or ECERS-R (Harms, Clifford, and Cryier, 1998). Its French version, the *Échelle d'évaluation de l'environnement préscolaire — Révisée* (ÉÉEP-R), was validated as an instrument for assessing the quality of childcare services as defined by the early childhood education specialists of the National Association for the Education of Young Children (NAEYC). The accreditation criteria set by the NAEYC in the 1980s are based on scientific knowledge of the factors that influence the physical, social, intellectual and emotional development of young children (National Association for the Education of Young Children, 1984; Bredekamp and Copple, 1999). The ÉÉEP-R was used several times within the scope of Francophone projects to assess the quality of childcare services. This scale, the reliability of which is well established in terms of internal validity and fidelity, measures the numerous aspects of a daycare that contribute to its overall quality. The ECERS-R (or its French version, the ÉÉEP-R) covers both the structural and process aspects of a daycare.

Quality is usually measured based on two dimensions: structural quality and process quality. Structural quality refers to a series of regulations under current legislation and includes elements such as the child/educator ratio, group size, the staff's level of education, the rate of staff turnover, as well as the educator's professional development and salary.²⁴

Process quality refers instead to children's social and educational experiences arising from their interactions with educators. This dimension explicitly recognizes the educator's important role in creating a rich, stimulating environment where children are encouraged to learn. Some studies argue that interactions with educators are the most important aspect of daycare quality (Committee for Economic Development Research and Policy Committee, 1993).

As part of the Readiness to Learn project, the implementation review for the Daycare Program required the use of certain indicators representative of structural and process quality. Since the study focused on the assessment of specific processes inherent to program delivery, the observers filled out only the subscales most relevant to the program's goals. As a general indicator of quality, the observers gathered information on the quality of indoor space (lighting, ventilation and room for every child), the elements in place to ensure the children's health and safety, the quality of the children's greeting upon arrival, as well as the quality of educator-child interactions.

In order to measure the Educative Quality of the childcare environment, the observers had to describe the activities observed, namely which components of child development were targeted and children's reactions during the activity. Special attention was given to the use of communication to foster children's language development. Thus, the observations allowed for the following ECERS-R subscales to be filled out: *Informal Use of Language*, *Encouraging Children to Communicate*, and *Using Language to Develop Reasoning Skills* or *Reasoning*. Given the program's emphasis on exposing children to reading, the *Books and Pictures* subscale was also filled out.

At the outset of the Readiness to Learn project, it was decided not to use the structured grid of the ECERS-R, since this type of observation could be perceived as too intrusive and impede

²⁴ Several elements of Structural Quality, such as salary, training and level of education, are difficult to ascertain through observation. This information was instead obtained through in-depth interviews or by referring to the educator data collected when the consent form was signed.

the full cooperation of program and comparison daycares. Note in this regard the difficulty of recruiting comparison daycares in certain communities (see Section 5.6 of the *Project Implementation Report*, Bérubé et al., 2014). It was therefore decided that the observers would take less formal notes on the ECERS–R elements for the targeted subcategories. Before the observations, each observer was given training on the elements to be observed, as well as a detailed observation protocol reminding them of the observation objectives, the elements to note, and concrete examples of excellent, good or inadequate practices. Observations were then validated by comparing the observers' notes with those of the trainer and by ensuring that the descriptions of each environment were in agreement. Analysis results were also presented to the observers for confirmation that the portrait of every childcare environment was representative.

To complete the profile of daycare activities, information was collected on circle time, when educators in the Program Daycare group had to read the children a story. Observations provide information about the style of reading (whether interactive or traditional), as well as the use of French-language songs and nursery rhymes. Observations were also made during supervised activity and free play periods. In these cases the observers focused on how much room the children were given to develop their autonomy and creativity.

Observations were conducted at the daycares in the Program Daycare group, as well as at comparison daycares. These data enable us to better understand the preschool program's new contribution to childcare settings. Certain practices may already have been in place in the childcare environment and the program may not have resulted in any major changes to existing practices. In this case, the children of the two conditions may present a similar development. Comparing what happens in the Program Daycare group and in the Comparison Daycare group allows us to determine how the program differs from the activities that occur naturally in childcare settings.

Protocol for Educator Interviews

Educators were interviewed in order to obtain their opinions on and experience with project implementation, namely, the daycare component and, for participating educators, the Family Literacy Workshops. The interview protocol is based on a review of existing literature on factors that may influence a program's implementation but are quantitatively neither observable nor measurable. Interview questions were therefore intended to factor in the difficulties that educators experienced in applying program components, adjustments, as well as day-to-day assimilation and integration of new practices arising from the program. The interview questions were also intended to elicit feedback from educators regarding training and follow-up. Finally, a few questions concerned the organizational environment, an element key to the success or failure of a new program. In fact, it is well known that without administrative support, the implementation of a new program is often destined to fail (Chen, 2005).

The subsections below describe how the in-depth interview was administered, as well as its content.

In-depth Interview: The Administration Process

The educators responsible for applying the program were interviewed twice during the Readiness to Learn project at all six sites for the first cohort and at both sites for the second cohort. Interviews were recorded as MP3 audio files. They lasted an average of 41 minutes, with

an interval ranging from 22 minutes to 65 minutes. They were then transcribed, after which the transcription was validated by an external judge.

Content of the Educator Interview Grid

SRDC developed the educator interview grid based on existing documentation concerning factors that may impede a program's implementation but are quantitatively neither observable nor measurable. It allows for examining the appraisal of basic training and follow-up visits, the experience of educators in learning and implementing the new program, the perceived effects on children and on parents, as well as the educators' perception of their role with respect to the children. There is also emphasis on the general work environment. Program quality is clearly affected by various factors associated with management of the daycare, and these questions are designed to give educators their say on this topic. Thus, two main questions were asked in this regard, namely, they were asked to describe their current working conditions and the support of their immediate supervisors at the daycare for the new program.

Educators' Journal

The purpose of the journal was to give educators a way to share their thoughts and feelings about project implementation (unexpected benefits, frustrations, etc.). The journal was made available from the very start of the daycare component's implementation. Information was gathered regularly, that is, educators were asked to jot down their thoughts a few times a month. The journal was confidential and filled out anonymously. To encourage participation, the coordinators regularly reminded educators about the journal. This optional activity was in the form of stapled sheets of paper or notebooks, depending on the community. The journal was an opportunity for educators to share their ideas on four topics formulated as statements:

- What I've learned and what I like so far with the Daycare Program;
- What's been hardest to implement or do so far;
- My suggestions to improve the program;
- Other comments I'd like to make.

This information fleshed out the data collected through educator interviews.

Daycare Attendance

The number of childcare hours per week is an important variable to consider in measuring children's program "dosage" or "exposure" at program and comparison daycares. This information was already being gathered through the attendance records used by the daycares. Educators had to fill out an attendance sheet (with every child's time of arrival and departure) every week. SRDC could have used this sheet to obtain information on the children in the project. However, use of the sheet was deemed inappropriate due to the risk of noting information on children who were not participating in the Readiness to Learn project (whose parents had not consented to share this information with SRDC). To be diligent in this regard, a new sheet for recording attendance (date and time) was created specifically for the Readiness to Learn project. This attendance sheet specified only the personal identification number (PIN) of participating children and space for recording the date and times of arrival and departure.

Community coordinators were responsible for visiting the daycare two or three times a month to take account of attendance by the participating children. They transferred the attendance information recorded in daycare documents onto the Readiness to Learn project attendance sheet. They then forwarded the duly completed attendance sheets to the Ottawa office on a monthly basis.

4.2.2. Implementation Review for the Family Literacy Workshops

The implementation study for the Family Literacy Workshops entailed several complementary measures for the data collection method and the source of information. This approach allows for triangulating the information from several perspectives, including those of the practitioner working with the parents, the practitioner working with the children, the observer, or parents participating in the workshops. The sections below describe in detail, respectively, the measurement instruments developed for the Readiness to Learn project, that is, the observation grid for the Family Literacy Workshops, the practitioners' journal, the in-depth practitioner interview grid, the parents' workshop appraisal, the logistical survey filled out by parents, the post-intervention survey of parents, as well as workshop attendance by families and staff.

Observation Grid for the Family Literacy Workshops

Observation of the Family Literacy Workshops allowed for collection of relevant data on program delivery (as defined by the objectives of the Readiness to Learn project and those of the Family Literacy Program). The work began with a literature review to identify validated tools for measuring the degree of implementation for the Family Literacy Program and its effects on participants.

To achieve the goals of the Readiness to Learn project, a semi-structured grid was created as a guide to the characteristics to be observed, although the observations themselves were in the form of informal notes. The information gathered told us about the topics addressed during the workshops, the practitioner's skills in delivering the workshop, and parents' reactions to the workshop topics.

The community coordinators and the members of the SRDC research team were responsible for observing the workshops. Because the program was new, all workshops were observed. The observation protocol gave greater emphasis to the parent component than to the child or parent-child components.

Practitioners' Journal

The purpose of the practitioners' journal was the same as with the educators' journal, that is, it was used as a tool for communication between the practitioners and the SRDC research team. Although this activity was optional, the practitioners were strongly encouraged to regularly jot down their observations and reactions concerning workshop delivery. The journal asked practitioners to share their reactions after the workshops, as well as their observations about parents' reactions to the workshops. It addressed four topics formulated as seven questions:

1. What is working well so far with the Family Literacy Workshops?
2. What has been the hardest aspect to implement and do so far with the Family Literacy Workshops?
3. What did parents like or appreciate?
4. What did parents like less?
5. If I had to give the workshop again, what would I do differently?
6. My suggestions for improving the program (in general)...
7. Other comments I'd like to make.

The journal was confidential and filled out anonymously. Practitioners were only asked to indicate the name of the community, the date of the workshop for which the journal was filled out and the group the practitioner was working with (parents or children).

Practitioner Interview Grid

The purpose of interviewing the practitioners was to gather their opinions and suggestions in order to improve the program and find out which aspects worked well and which were harder to implement. Since no interview grids have been validated and/or published, SRDC developed an interview grid based in large part on the interview grid for educators. Questions on the grid were taken from tools developed for other research projects conducted by SRDC and the University of Ottawa. Finally, certain questions were inspired by the training manual for the Family Literacy Program.

In-depth Interview: The Administration Process

All practitioners working with parents, as well as a few practitioners working with children, were invited for an interview one or two weeks after the workshops ended. This decision was made because the program is new and primarily targets parents, hence the importance of interviewing all the practitioners working with parents. Interviews were recorded as MP3 audio files and were conducted either in person or by phone. Interviews with the practitioners who worked with parents lasted an average of 1 hour and 8 minutes (ranging from 47 minutes to 1 hour and 35 minutes). Interviews with the practitioners who worked with children lasted an average of 48 minutes (ranging from 19 minutes to 1 hour and 9 minutes). The recordings were then transcribed with the respondents' consent. Transcription was subsequently validated by an external judge.

Content of the Practitioner Interview Grid

SRDC developed the practitioner interview grid so as to capture the factors that may affect a program's implementation but are neither observable nor quantifiable. The interview grid examined the following points:

- Knowledge about the Readiness to Learn project, family literacy programs and the Family Literacy Program;
- Initial reactions to the Family Literacy Program;

- Workshop implementation and logistics (i.e., workshop organization, duration and format; facilitators and barriers encountered; benefits and drawbacks related to delivery of each component [parent, child, and parent-child]; relevance of the material provided);
- Program content (i.e., relevance to parents; relevance of the content to children; usefulness of the material distributed to parents);
- Training and support (i.e., the relevance of basic training; needs in terms of consultations and identification of the additional resources consulted);
- The program's effects on the practitioner at the occupational level; and
- The program's effects, as observed by practitioners or reported by parents, on parental beliefs, attitudes and habits.

Parents' Workshop appraisal

Parents were encouraged to fill out a short assessment of workshop content. This appraisal concerned their reactions, such as what they liked at each workshop, what they liked less or found useful, and topics they would like to know more about. The workshop appraisal was designed to inform the practitioners and the SRDC research team about parents' reactions, thereby enabling them to make appropriate adjustments at the next workshop. For example, if several parents said they wanted to know more about a given topic, the practitioner could provide references or resources at the next workshop. The workshop appraisal was administered at the end of every workshop and took about five minutes to fill out.

Parents' Logistical Survey

At the last workshop, parents filled out a short logistical survey on workshop scheduling and delivery, facilitation by the practitioner, the aspects they liked, workshop content, and their reactions to the workshops. The goal of this logistical survey was to shape the delivery of new workshop series in terms of logistics, process and content coverage. Note that this survey was filled out anonymously in order to encourage honest feedback from parents. The survey, which consisted of about 15 questions, took roughly 15 minutes to fill out. It was administered to parents at the beginning of the last workshop. Obviously, only those parents who attended the tenth workshop were able to fill out the survey. Note that certain questions from the logistical survey were repeated in the post-intervention survey due to their significance in determining the workshops' impact on parents and in improving the Family Literacy Program.

Parents' Post-Intervention Survey

Parents were asked to answer a post-intervention survey after workshops ended. For the purposes of the implementation review, two versions of the post-intervention survey were created: one for participating parents (i.e., those who attended three or more workshops) and the other for non-participating parents (i.e., those who attended two or fewer workshops). The rule of three or more workshops to create a first group identified as participating in workshops was established arbitrarily. It was deemed that as of three workshops, parents would have enough experience with workshop delivery and content to provide informed feedback regarding these aspects.

Of particular interest for the implementation review of the Family Literacy Workshops, the post-intervention survey included questions on the experience of parents participating in the Family Literacy Workshops (e.g., effects of the workshops on the parents' day-to-day habits; suggestions or comments for improving the program; participation facilitators or barriers), while non-participants answered questions to identify topics of interest and the reasons for their non-participation (e.g., areas of child development and school readiness they would have liked to learn about; suggestions or comments for improving the program; participation facilitators or barriers). Project implementation was therefore documented through these responses, which explained the barriers or reasons why parents chose to participate or not in the workshops.

Attendance at Family Literacy Workshops

The number of sessions attended by each family is an important variable to consider in measuring parents' program "dosage" or "exposure." Community coordinators were responsible for collecting this information. Attendance was noted for parents, children, practitioners working with parents, practitioners working with children, assistant practitioners and any other person involved in delivering the services. The attendance of the individuals responsible for service delivery was important to consider since their presence ensured continuity in the practitioner-participant relationship as well as stability in service delivery.

4.3. METHOD OF DATA ANALYSIS

4.3.1. Analysis Strategies

Data analysis is based on the methodology of grounded theory, as presented by Strauss and Corbin (1998). Grounded theory allows researchers to study a complex object and to understand how it fits into a given reality. This makes it possible to organize data and identify their meaning, while remaining very close to the statements gathered in the interview process or to the responses made to open-ended questions. Moreover, grounded theory gives particular emphasis to the social context of the studied object (Laperrière, 1997).

The N-Vivo software was used to facilitate management of the gathered material. Data were analyzed using axial coding, as defined by Strauss and Corbin (1998). A first vertical analysis was conducted, that is, an analysis of the content from each measurement instrument. This content analysis was used to create the dimensions. A cross-sectional analysis was then conducted, that is, analysis results were compared for every dimension of each measurement instrument. This analysis led to a more detailed definition of the specificities of each set of data based on the measure and to formulation of the key themes addressed in all data from various instruments.

For example, each observation of the Daycare Program was read in its entirety and the elements of the observation related to one dimension of the studied implementation were grouped together (e.g., all passages indicating that the educator posted children's crafts on the walls at the daycare were placed in the *structural fidelity/crafts* category). Once all the observations were analyzed, a matrix was created using the N-Vivo software. Each line of the matrix represents the observations of a class in a given daycare environment. The columns represent each category for the studied dimension (e.g., for the Structural Fidelity matrix for daycares, the columns represent each element of fidelity selected for study). The matrix results

tell us whether elements are present or missing for each class (score 0/1). Finally, the scores obtained can be combined to establish a portrait for the daycare environment of the Program Daycare group versus those of the comparison group. Table 4.1 uses dummy results to illustrate a matrix for Structural Fidelity.

Table 4.1: Dummy Results for the Structural Fidelity Matrix

Class	Structural Fidelity in Childcare Settings Period 2: November 2007 – February 2008						
	Properly identified centres	Picture/word cards present	Weekly program present	Theme-specific activities	Routine chart present	Routine chart used	Crafts
1	0	1	0	0	1	1	1
2	1	1	1	0	1	1	0
3	1	0	1	0	1	1	1
4	1	1	1	1	1	1	1
5	0	1	1	1	1	1	1
6	1	1	1	0	1	1	1
7	1	0	1	0	1	1	1

In total, the analyses examined seven dimensions as part of the project's implementation study:

1. fidelity;
2. quality of intervention;
3. differentiation between the tested program and other programs;
4. participation (i.e., participants' response);
5. dosage;
6. facilitators and barriers; and
7. participants' perceived benefits.

All analyses were validated at several levels, primarily to ensure that the content grouped into the same category actually addressed the same theme and that the category title was representative of the content. Each category's content and title were then validated by a third party. Finally, the matrix results were validated by comparing the results for whether various elements were present or missing in the daycare environment to the observation notes of the early childhood consultant to ensure there was agreement in the profiles of a given environment established by two different sources. In the case of elements to be implemented in the environment, the photos taken of the childcare settings were consulted to make sure that an

element identified as missing was not actually found in the photos. Finally, analysis results were presented to the observers for confirmation that the portrait of each daycare environment was representative.

To obtain an overall picture of the tested program's implementation, we triangulated data from several sources using a three-tiered mixed methodology (Patton, 1990). First, data on a given element were from various sources. We asked participants, practitioners and observers to tell us about their experience with the program using the same instrument. Next, data were gathered through a range of measurement instruments to evaluate the same phenomenon. Finally, the study allowed for triangulation of the methodologies. In fact, the information was gathered through different media, both quantitative and qualitative.

5. Implementation Results for the Daycare Program

This chapter presents the findings of the implementation study for the Daycare Program. The chapter begins with a section detailing the *analysis strategy for daycare observations*, including treatment of the missing data. Section 5.2 details the *Structural Quality* results for the observed daycares, an element that concerns the overall quality of the daycare environments.

The next sections concern elements related directly to the Daycare Program. More specifically, Section 5.3 describes the results for fidelity of the Daycare Program's implementation. These results are for *Structural Fidelity*, that is, implementation of the program elements in the environment. Section 5.4 concerns *Fidelity of Educational Content*, that is, elements that must be present in the activities to which children are exposed. To facilitate reading, the results for quality of implementation for program content elements, particularly *Educative Quality* elements (i.e., the educator's preparation), are presented at the same time as fidelity. All data in these sections are sourced from the observation notes taken at the daycares. Since observations were made at both program daycares and comparison daycares, the results also tell us about *differentiation*, that is, the difference between what the program contributes to the environments and what takes place naturally in non-program environments.

Section 5.5 concerns the *Quality of Educator-Child Interactions*. Section 5.6 addresses *Dosage*, that is, program frequency or quantity at various daycares. Section 5.7 details *Educator Reactions* to the program. Data were obtained from educator interviews. Section 5.8 presents the *Facilitators and Barriers* encountered during project implementation. Section 5.9 discusses the program's *Perceived Effects* on children and educators. The data in Sections 5.8 and 5.9 are drawn mainly from educator and trainer interviews, which were subsequently validated by observations at the daycares. Results in these sections also stem from observation notes taken by the trainer during her visits to daycares in the Program Daycare group. Finally, a summary of findings is presented in Section 5.10.

The following points should be noted:

1. The implementation review for the Daycare Program is based on the data collected in five of the six communities in the first cohort and the two communities that make up the second cohort. Data from one community in the first cohort were not considered because the program was not implemented in that community.
2. Analyses revealed that program delivery was similar for the first and second cohorts. No marked difference was noted in terms of the quality and fidelity indices. The results are therefore presented for both cohorts combined. This approach maintains the anonymity of educators in the second cohort, of whom there were few in the second year of the program (fewer than five).
3. Although the second cohort consisted of two first-cohort communities (Cornwall and Orleans), it is easier to present the results by dealing separately with each of the five communities in the first cohort and the two communities in the second cohort, for a total of seven sites.

5.1. ANALYSIS STRATEGY FOR DAYCARE OBSERVATIONS

In the first year of the Readiness to Learn project, the daycare observations for the Program Daycare group were gathered from August 2007 to June 2008 for the first cohort and from October 2008 to June 2009 for the second cohort. These observations give us an overview of how the program is delivered in the course of a typical year. Observations were grouped into three 4-month periods corresponding to the child assessment periods. In actual fact, some daycares in the Program Daycare group were observed more than once during a given period. The **first period** corresponds to the months from *August to October*. The **second period** extends from *November to February* and constitutes the program's core. Finally, the **third period** lasts from *March to June* (see Table 5.1 for a summary of the observation periods).

Observations at comparison daycares were less frequent since no particular classroom intervention was planned. Comparison daycares were all observed during the **second period** (from November to February) and most (13 of the 19 classes) were observed a second time during the **third period**. As a result, the number of observations per comparison daycare varies from site to site.

Delivery of the Daycare Program continued for a second year at five of the seven sites in the study. At two sites (Orleans/first cohort, Orleans/second cohort), the program was halted because all the children had begun school full-time.²⁵ Given the high turnover in educators giving the tested program, a decision was made to continue monthly observations in the program's second year. As in the first year, the observations were grouped into three 4-month periods corresponding to the child assessment periods. Thus, the **fourth period** corresponds to the months from *August to October*, the **fifth period** extends from *November to February* and, finally, the **sixth period** lasts from *March to June* (see Table 5.1 for a summary of the observation periods). In the second year, most comparison daycares were observed only once, either during the fifth or sixth period.

Table 5.1: Observation Periods by Year

	First Cohort	Second Cohort
First Year		
Period 1	August to October 2007	October 2008
Period 2	November 2007 to February 2008	November 2008 to February 2009
Period 3	March to June 2008	March to June 2009
Second Year		
Period 4	August to October 2008	August to October 2009
Period 5	November 2008 to February 2009	November 2009 to February 2010
Period 6	March to June 2009	March to June 2010

In the first year, observations were conducted by community coordinators in 18 classes forming the Program Daycare group (12 classes in the first cohort and 6 classes in the second

²⁵ In Durham, only a few children participated in the Daycare Program in the second year because most children had begun school. Daycare observations were nevertheless conducted during that period and the data were included in the analyses.

cohort) and 19 classes forming the comparison group (12 classes in the first cohort and 7 classes in the second cohort). The program continued for a second year in Cornwall (two cohorts), Durham, Edmundston and Saint John. In the program's second year, community coordinators conducted observations in 10 classes forming the Program Daycare group (eight classes in the first cohort and two classes in the second cohort) and 14 classes forming the comparison group (10 classes in the first cohort and 4 classes in the second cohort; see Table 5.2).

Table 5.2: Number of Classes in the Project by Treatment Group, Year of Implementation and Cohort

	Year 1		Year 2	
	Program Daycare Group	Comparison Group	Program Daycare Group	Comparison Group
Cohort 1	12	12	8	10
Cohort 2	6	7	2	4
Total	18	19	10	14

Treatment of Missing Daycare Observations

It was impossible to conduct all the planned observations for classes at daycares in the Program Daycare group and comparison group during the program's two years due to scheduling conflicts and certain reservations expressed by daycare staff. As a result, the number of observations varies from site to site for a given period (e.g., from November to February), which required treatment based on context. Percentages were calculated based on the observations. Missing data are identified in the text where necessary.

Management of Multiple Observations

More than one strategy was used to manage multiple observations, in particular:

- Use of the ***observation average*** (e.g., if an element is present for one assessment and missing for the other, a value of 0.5 is assigned to the class). The advantage of this strategy is that it fully represents the fluctuations observed in applying each element.
- Application of a ***flexible rule***, where a score of "1" is given if the element is present for at least one observation during the specified period. This approach allows for reporting information in a manner that expresses the results for *classes* that met a given criterion (e.g., 11 of the 18 classes met criterion X).

The following points should be noted:

1. In the tables, results are reported by calculating the average observations. On the one hand, Structural Fidelity (Table 5.4) and Fidelity of Educational Content for program elements (Table 5.6) are presented as a percentage. On the other hand, Structural Quality (Table 5.3) and Quality of Educational Content (Table 5.5) are presented as mean scores.

2. In the text, the results are generally presented based on the flexible rule, thus allowing us to describe the number of classes that met a given criterion during the first and second year of Daycare Program delivery.
3. Because the percentages in the text and tables are calculated using various methods, the numbers in the tables will not necessarily match those presented in the text. The two sources of information reflect separate yet complementary aspects of implementation.

5.2. STRUCTURAL QUALITY AND DIFFERENTIATION OF THE OBSERVED DAYCARES

To determine whether the program and comparison daycares of both cohorts offer an equivalent general level of quality based on certain measurable aspects, quality indices were calculated for every class at each daycare. The general quality measured as part of this study focuses on the environments' structural aspect. These indices are quality of indoor space, health and safety measures, as well as measures of children's and parents' greeting and departing by educators. These quality indices were measured by the ECERS–R subscales (Harms, Clifford, and Cryer, 2010).

The results of each ECERS–R subscale are compiled on a seven-point scale defined as follows: **inadequate** (1) means conditions or care that are well below basic requirements for childcare; **minimal** (3) means conditions or care that meet basic requirements and, to a low degree, basic developmental needs; **good** (5) means conditions or care that stimulate child development to a certain degree; and **excellent** (7) means personalized, superior-quality conditions or care (Harms and Clifford, 1990). The results of this section are presented in Table 5.3.

For the “*Indoor Space*” subscale of the ECERS–R, the average for the Program Daycare group was 6.89 out of 7 (SD = 0.32) and that of the comparison group was 5.89 out of 7 (SD = 2.28). Most educators in both treatment groups (16 in the Program Daycare group and 15 in the comparison group) scored 7 (excellent). Overall, the scores of both treatment groups ranged from good to excellent for this dimension of quality.

As regards quality of “*Health Practices*,” the Program Daycare group scored an average of 6.22 out of 7 (SD = 1.17). Quality for the comparison group was 6.00 out of 7 (SD = 1.25). Several educators scored 7 (12 in the Program Daycare group and 11 in the comparison group). A few educators scored 5 (four in the Program Daycare group and five in the comparison group) or 4 (two in the Program Daycare group and three in the comparison group). In general, all daycares in the project had a high quality of hygiene.

Finally, for the assessment of “*Greeting/Departing*,” the average for the Program Daycare group was 6.11 out of 7 (SD = 1.53) and for the comparison group, the average was 5.68 out of 7 (SD = 0.95). Several educators scored 7 (ten in the Program Daycare group and three in the comparison group) or 6 (five in the Program Daycare group and ten in the comparison group), showing a good quality of greeting at the daycares in the Readiness to Learn project.

In short, the observation notes revealed that there is a very slight difference in overall daycare quality, to the advantage of daycares in the Program Daycare group. However, these

differences are minor and both groups show good quality (scoring 5 or higher on a scale of 7) for each dimension assessed. As such, we can say that there is no differentiation between daycares in the Program Daycare group and the comparison group in terms of Structural Quality. To give context to the following results, Table 5.3 also presents the averages for a study conducted by Japel and colleagues (2005). This study assessed the level of quality at 296 daycares in the province of Quebec in connection with the *Étude longitudinale du développement des enfants du Québec* (ELDEQ). The daycares observed were accredited for-profit, non-profit and home-based daycares. The findings of the study by Japel, Tremplay, and Côté (2005) show that for each ECERS–R subscale in this study, the daycares in the Readiness to Learn project score higher than the Quebec daycares.

Table 5.3: Mean Score for Structural Quality of the Observed Environments by Dimension Assessed and Treatment Group

Overall Quality of Childcare Settings (based on ECERS-R subscales)	Classes in the Program Daycare Group n = 18	Classes in the Comparison Daycare Group n = 19	Daycares in the study by Japel et al. (2005) n = 296
Indoor Space	6.89 (0.32)	5.89 (2.28)	2.80 (2.09)
Health Practices	6.22 (1.17)	6.00 (1.25)	3.06 (2.23)
Greeting/Departing	6.11 (1.53)	5.68 (0.95)	5.14 (2.14)

Note: Indices are calculated using observations for the first year of implementation for the first and second cohorts combined. There were too few observations in the second year to calculate these indices.

5.3. STRUCTURAL ELEMENTS IN THE OBSERVED ENVIRONMENTS: FIDELITY AND DIFFERENTIATION

The observations by community coordinators verified the extent to which program elements were implemented in all classes over the two years of the project. This section reports the analysis results for Structural Fidelity by year for the Program Daycare group and the comparison group.

The section begins with a profile of the elements implemented in the first and second years of the tested program’s delivery. The percentages indicate whether the assessed elements were present or missing (i.e., based on the flexible rule). Wherever possible, the results are reported by observation period for a better appreciation of how implementation evolved over time. We end with a general profile of the classes studied broken down by year of program delivery and treatment group. This profile allows for evaluating the fidelity and stability of the structural elements implemented at daycares in the Program Daycare group and in the comparison group. In all cases, we deal with classes as units of analysis and we distinguish between those from daycares in the Program Daycare group and those from daycares in the comparison group. Results for the fidelity of structural elements are presented based on the flexible rule in the text as well as on the average observations (as percentages) in Table 5.4.

The Structural Fidelity assessment includes seven elements: (1) clearly identified learning centres, (2) cards displaying a picture and a word, (3) weekly programs, (4) theme-based activities, (5) the presence of a routine chart, (6) the use of a routine chart, and (7) displayed crafts.

Program's Structural Fidelity

Among the first structural elements of the program involved implementing, on the premises of the daycares in the Program Daycare group, a number of learning centres such as a block centre, an arts centre, an imagination centre, a board game centre, etc. The presence of learning centres is fairly widespread in childcare settings. The specificity of the Daycare Program was its requirement that these ***learning centres be clearly identified with the help of a picture accompanied by the matching word***. This requirement meant that educators had to clearly define each learning centre, thereby exposing children to the world of writing by supporting written symbols with images. This element of the program was implemented at all daycares of the group participating in the intervention from the program's very outset (11 of the 18 classes were observed during the first period) except at one site (4 classes). At that site, educators properly identified the learning centres only in the third period (March to June). Conversely, the data indicate that the classes forming the Comparison Daycare group were less in the habit of clearly identifying their learning centres. Although all classes in the comparison group had learning centres, the centres were identified using a picture or word card at least once in the first year in only 8 of the 19 classes (42%). The second year, the centres were clearly identified at least once in all the classes in the Program Daycare group (100%), while the practice was observed in 8 of the 14 classes (57%) of the Comparison Daycare group.

The program's second element concerned the display of different ***cards displaying a picture and a word*** in program daycares classes. These cards exposed children to written language every day. Eight of the 11 classes observed in the Program Daycare group (73%) implemented this element of the Daycare Program in the first period. Over the course of the year, this proportion increased to 78% and 82% in the second and third periods respectively. In comparison, this practice was present in seven classes of the Comparison Daycare group (37%) during the second period. In the second year, 7 of the 10 classes in the Program Daycare group (70%) had introduced the cards displaying a picture and a word. Cards displaying a picture and a word were found in 10 of the 14 classes at comparison daycares (71%), thus reducing the difference between the two groups in this regard.

The third element of the tested Daycare Program was that educators had to use a ***weekly program*** presenting the activities that would be conducted that week with the children. This grid had to be visible for parents at all times at the class entrance and had to be up-to-date. In 17 of the 18 classes (94%), the weekly program was used systematically. For classes at comparison daycares, 14 of the 19 classes (74%) used the weekly program on a regular basis. In the second year, it was observed that 8 of the 10 classes in the Program Daycare group (80%) and half of the 14 classes in the Comparison Daycare group (50%) used these grids at least once.

Fourth, in addition to using the weekly programs, educators had to ***base their activities on a theme*** selected for the month. The data show that 16 of the 18 classes in the Program Daycare group (89%) focused their activities on a theme in the first year. This practice was found in 14 classes at 19 comparison daycares (74%). The second year, all classes in the Program Daycare group (100%) and half of the 14 classes in the Comparison Daycare group (50%) had linked their activities to a theme.

The program's fifth element involved ***display of a routine chart***. The routine chart shows children the day's activities in their scheduled order. Educators had to post the routine chart at child level in an unobstructed area. The routine chart was displayed in 15 of the 18 classes in the

Program Daycare group (83%) as of the second observation period. Note that as of the third period, all classes in the Program Daycare group (100%) displayed the routine chart. However, no classes in the Comparison Daycare group had posted a routine chart. In the second year of activities, 9 of the 10 classes in the Program Daycare group (90%) displayed the routine chart. In the classes of the Comparison Daycare group, a routine chart was displayed at least once in 4 of the 14 classes (29%).

The sixth element of the program concerned *using the routine chart with the children*. This practice helps children understand how the day unfolds, anticipate upcoming activities, and, especially, feel secure. This element was applied only in 11 of the 18 classes of the Program Daycare group (61%) in the first year. No classes in the Comparison Daycare group used a routine chart at that time. In the second year, only 4 of the 10 classes in the Program Daycare group (40%) used the routine chart. Only a single class in the Comparison Daycare group used a routine chart (7%).

The last element to be implemented as part of the Daycare Program was the *display of children's crafts* on the daycare premises. Craft display helps children develop a sense of belonging to the daycare, in addition to celebrating their creations. According to observations, crafts were displayed at least once in all classes of the Program Daycare group (100%) and in 16 of the 19 classes in the Comparison Daycare group (84%). Crafts also had to be placed *at children's eye level*. Fifteen classes in the Program Daycare group (83%) introduced this practice in the first year, while at comparison daycares, the practice was found in seven of the classes observed (37%). Finally, educators had to *write the child's name* on the front of the crafts so that children could learn to recognize their name and its letters. Children's names were written on crafts at least once during the year for all classes in the Program Daycare group (100%). This practice was in place for 16 of the 19 classes in the Comparison Daycare group (84%). The second year, all classes in the Program Daycare group except one (90%) displayed the crafts in accordance with program requirements while the practice was observed in 5 of the 14 classes in the Comparison Daycare group (36%).

All observations were compiled by year to sketch a general portrait of the two groups studied. An examination of Table 5.4 shows that on average, in the first year, the classes in the Program Daycare group implemented 78% of the Daycare Program elements (ranging from 46% to 91%), while all these elements were found in less than half of all classes at comparison daycares (average of 40%, ranging from 0% to 71%). In the second year, the level of implementation at daycares in the Program Daycare group was very similar, with an average of 73% (ranging from 40% to 90%). For comparison daycares, the average was 42%, similar to the first year (ranging from 4% to 71%).

In short, in the first year, all the program elements were present and practiced in the Program Daycare group more often than in the Comparison Daycare group. In particular, the learning centres were more frequently identified, cards displaying a picture and a word were more present, and routine charts were present and used during the day. In the program's second year, all the elements were implemented more often by classes in the Program Daycare group than by classes in the Comparison Daycare group, except for the presence of picture/word signs, for which the rate of implementation was similar for both treatment groups (70% versus 71%). All these results suggest that program elements were, for the most part, implemented in the classes of the Program Daycare group in a fairly stable manner.

Table 5.4: Average Percentage of Fidelity for Structural Elements by Year of Program Delivery and Treatment Group

Fidelity of Structural Elements	Year 1		Year 2	
	Classes in the Program Daycare Group	Classes in the Comparison Daycare Group	Classes in the Program Daycare Group	Classes in the Comparison Daycare Group
	n = 18	n = 19	n = 10	n = 14
Clearly identified learning centres	89%	34%	78%	45%
Cards displaying a picture and a word	79%	39%	70%	71%
Weekly programs	91%	71%	73%	46%
Theme-based activities	89%	71%	90%	46%
Presence of routine chart	75%	0%	80%	21%
Use of routine chart	0%	0%	40%	4%
Crafts displayed	79%	65%	81%	58%
Average for all 7 elements	78%	40%	73%	42%

5.4. EDUCATIONAL CONTENT: FIDELITY, QUALITY AND DIFFERENTIATION OF ACTIVITY CONTENT

Given the importance of language skills development for children living in a minority Francophone context, the Daycare Program particularly emphasized children’s exposure to oral communication through books and songs, while giving them many opportunities to express themselves and develop their thoughts. Observations determined the degree to which the program’s educational content was implemented over the two years of the project. The paragraphs that follow examine each element assessed as well as its results. Three main aspects are addressed: (1) **Reading**, (2) **Conditions Conducive to French-language Communication**, and (3) **Pre-writing**. Three other elements were also assessed: **Creativity**, **Autonomy** and **Transitions**. The section ends with a summary of the results. Note that results for the elements in this section are reported mainly based on the flexible rule in the text, as mean scores in Table 5.5 for Quality of Educational Content, and as percentages in Table 5.6 (i.e., based on the average observation notes) for Fidelity of Educational Content.

Reading

First, emphasis was placed on the importance of reading to children and of making quality books available to children. Several studies have shown the importance of reading for young children. Moreover, as Snow and colleagues (1998) explain, access to books exposes children to new vocabulary, in addition to initiating dialogue between the adult and child. To foster a stimulating environment for oral communication, daycares in the Program Daycare group were given almost 300 books in addition to other French-language resources (e.g., CDs, DVDs). These new resources, as well as the program’s emphasis on reading to children, is reflected in the significant difference in quality scores for the Program Daycare group and the Comparison Daycare group for the “**Books and Pictures**” subscale of the ECERS–R. This subscale measures

the quantity and accessibility of books in the classroom as well as the quality of the literacy activities led by the educators.

Classes in the Program Daycare group obtained a mean score of 5.0 ($SD = 2.06$) for this dimension, a score that describes the literacy activities in the Program Daycare group as “good.” Eight classes scored 7, another eight scored 4, and two classes scored 1, indicating inadequate quality. In comparison, the classes of the Comparison Daycare group scored 2.42 ($SD = 1.98$), which means their environment was only slightly better than inadequate for this dimension. Seven classes scored between 4 and 6, while 12 classes had a score of 1. The main shortcoming of educators in the Program Daycare group was that they did not initiate reading spontaneously throughout the day. Educators in the Comparison Daycare group did not often read books to children and did not have enough books available for the number of children in the class.

However, access to material is not an end in and of itself: the educators must actually use it. The observations of the community coordinators confirmed that, in the first year, educators in 17 of the 18 classes in the Program Daycare group (94%) ***read books during circle time***. The observations also showed that during the second observation period, 5 of the 18 classes in the Program Daycare group (28%) ***read books to children outside*** circle time. This proportion rose during the year such that by the third period, educators in 8 of the 18 classes in the Program Daycare group (44%) had made it a habit of reading to children at other times of the day. What’s more, as of the third period, educators in 7 classes of the Program Daycare group (39%) exposed children to reading in at least two different contexts. As for educators in the comparison group, the coordinators’ observation notes for the first year revealed that educators in 11 out of 19 classes (58%) had read to children during circle time and that educators in 5 classes (26%) had read to children as part of an organized activity. This means that in all classes of the Program Daycare group (100%), children were exposed to reading at least once a day, whereas this was only the case for 6 of the 19 classes in the Comparison Daycare group (32%).

In the second year, educators in all classes of the Program Daycare group (100%) read to children during circle time, and in six of these classes (60%), educators also read to children as part of organized activities. Thus, educators in 6 out of 9 classes in the Program Daycare group (67%; the observation for one class is missing) exposed children to reading in more than one context. As concerns the Comparison Daycare group, educators in 5 of the 14 classes (36%) took advantage of circle time to read to children, and educators in 9 classes (64%) read during organized activities. Educators in 4 out of 12 classes in the Comparison Daycare group (33%; the observations for two classes are missing) exposed children to reading in at least two different contexts.

According to research on child development, reading is not always enough to improve children’s language skills. In fact, as shown by Hargrave and Sénéchal (2000), children benefit more from ***interactive reading***, where the adult involves the child in a discussion about the story and asks questions to encourage the child to give the story more thought and in so doing enhance his or her understanding of it. Researchers found that children exposed to this type of reading had a much broader range of vocabulary than children exposed to traditional reading. A study conducted by Wasik and colleagues (2006) took the intervention even further. Daycares were given books in thematic boxes that also contained material for other activities based on the books’ topic. Educators were trained to reuse the vocabulary encountered in the books as part of thematic activities. They were also told about the importance of asking open-ended questions and

of explaining to children the meaning of the targeted words before reading to them. The children who participated in this program learned more vocabulary than did children exposed to a conventional style of reading.

The Daycare Program of the Readiness to Learn project also advocated the interactive reading method. This method was in fact part of the training for educators in the Daycare Program. As in the study by Wasik and colleagues (2006), books were distributed to daycares in thematic kits that also contained material for developing activities on the selected topic. This arrangement allowed educators and children to reuse and consolidate the vocabulary to which they had been exposed during reading. When reading a book, educators had to ask open-ended questions and initiate discussions about the book. Educators also had to use the material (e.g., puppets, pictures, objects, accessories, etc.) to support the reading, thereby enabling the children to better follow the story while maintaining their interest.

The data revealed that, in the first year, all educators in the Program Daycare group (100%) read a book at least once in accordance with the requirements of the tested program. Note that we were unable to compare the implementation of this element in the Comparison Daycare group due to incomplete data; several educators in the Comparison Daycare group did not read children a book during the observations. The second year, the interactive reading style was found in 9 of the 10 classes in the Program Daycare group (90%). In the Comparison Daycare group, this style of reading was observed in only 1 of the 14 classes (7%).

Conditions Conducive to French-Language Communication

A second important aspect of the Daycare Program was to give children numerous and varied opportunities to communicate in French. This objective was achieved in various formal and informal contexts. It should be noted that all educators were asked to use francization techniques, including speaking to children only in French and repeating the child's words in French when he or she spoke in English. Based on the observation notes, educators in both treatment groups often asked questions that encouraged children to speak French. They also often made their requests simpler instead of using English when children did not understand the French words. Finally, all educators in both treatment groups spoke only French with the other adults in the class.

For a more in-depth examination of educator-child communications, we made our observations at times that favoured communication by the children and used the ECERS-R subscales to assess that communication. To stimulate children's communication, educators in the Program Daycare group were encouraged to hold circle time, that is, to gather the children together in order to discuss various topics based on the day's events and the current theme. From the very first year, ***circle time*** was implemented at all daycares in the Program Daycare group. At daycares in the Comparison Daycare group, 17 classes (89%) held circle time during the second observation period, but three classes stopped the practice during the third period. The second year, circle time was observed in all the classes in the Program Daycare group and those of the Comparison Daycare group (the observations for three classes of the Comparison Daycare group are missing).

The quality of informal verbal communication between educators and children was measured using the "***Informal Use of Language***" subscale of the ECERS-R. The criteria of this subscale measure the degree to which educators encourage children to express themselves by initiating conversations with them and by encouraging the children to talk with each other. Observations

show that daycares in the Program Daycare group were more encouraging of spontaneous communication than comparison daycares. The average for classes in the Program Daycare group was 5.33 out of 7 ($SD = 1.53$) while the average for classes in the Comparison Daycare group was 4.68 out of 7 ($SD = 1.67$). Note that 8 of the 18 classes in the Program Daycare group scored the maximum of 7, which ranks their practices as excellent, in comparison with five classes in the Comparison Daycare group. The other classes in the Program Daycare group, as well as 11 classes in the Comparison Daycare group, scored 4, which means their practices range from a ranking of minimal (3 out of 7) to good (5 out of 7). The main shortcoming of these educators was that they gave children little encouragement to participate more in discussions, either by adding new elements to their statements or by asking questions to further explore the topic.

Observations also concerned support for communication through more formal techniques. For example, educators in the Program Daycare group were encouraged to use visual material in order to help the children communicate (such as the use of puppets or small figures to help replay a story or tell a new one). Moreover, they were encouraged to form small discussion groups centering on a specific topic. This dimension was measured using the “*Encouraging Children to Communicate*” subscale of the ECERS–R. Fifteen of the 18 classes in the Program Daycare group achieved the top score for this subscale versus 11 of the 19 classes in the Comparison Daycare group. The other educators in both treatment groups scored 4 or higher, indicating a level of quality above the minimum score (3), except for one educator in the Program Daycare group. The average for both treatment groups in the study was 6.39 ($SD = 1.58$) for the Program Daycare group and 5.84 ($SD = 1.42$) for the Comparison Daycare group. The main shortcoming of these educators was that they introduced few spontaneous conversations into the activities and free play outside of circle time to encourage children to communicate their ideas.

Songs and nursery rhymes were another unique opportunity to expose children to new vocabulary and raise their awareness about the sounds that make up words, as with songs and nursery rhymes. Data for the first year revealed that educators in all classes of the Program Daycare group (100%) used songs during circle time. In seven classes (39%), educators had children sing as part of a organized activity and in two-thirds of classes (67%), they used songs for transitions. The general average for this dimension across the various times was 58%. Also, in the first year, educators in 15 of the 19 classes in the Comparison Daycare group (79%) had children sing during circle time, in 4 classes (21%), educators encouraged children to sing during organized activities and in a high proportion of classes (79%), educators used songs during transitions. The general average for this dimension across the various times was 52%. In the second year of implementation, all educators in the Program Daycare group (100%) used songs during circle time; in four classes (40%), children sang during organized activities and in 8 out of 10 classes (80%), educators sang systematically with children during transitions. The general average was 68%. For the Comparison Daycare group, all the educators observed sang during circle time (100%; the observations for four classes are missing). In one class (8%), the educator sang with children during organized activities (the observations for two classes are missing) and in half of the 14 classes (50%), educators facilitated transitions with songs. The general average for classes in the Comparison Daycare group at all these times was 50%. In total, the results show that both treatment groups in the study used songs to a fairly similar degree.

Activities that encourage the child's *reasoning* are linked to child language development and logical thought. This type of activity was promoted in program and comparison daycares, since it was an opportunity to encourage children to express themselves and explore their ideas by talking. The observations determined the frequency at which daycares conducted activities to promote reasoning by children, such as classification activities and activities involving colours or shapes, without regard to children's use of language during those activities. Educators took advantage of various times throughout the day to develop this ability among children. Observations during the second period (of the first year) show that educators in 15 classes of the Program Daycare group (83%) used circle time to expose children to reasoning activities, versus 14 classes in the Comparison Daycare group (74%). Educators in 11 classes of the Program Daycare group (61%) used organized activities to do so, versus 14 classes in the Comparison Daycare group (74%). Seven educators (41%) used snack time (the observation for one class is missing), versus two classes in the Comparison Daycare group (13%; the observations for four classes are missing). What's more, educators in two classes of the Program Daycare group (12%) introduced various concepts to children during transitions (the observation for one class is missing), versus three educators for the Comparison Daycare group (16%). In total, all educators in the Program Daycare group (100%) and 14 educators in the Comparison Daycare group (74%) used reasoning in more than one context. Both treatment groups had the same rate in the second year, when children perform reasoning activities on more than one occasion in roughly two-thirds of the groups observed in both the Program Daycare group (7 out of 10 classes, or 70%) and the Comparison Daycare group (10 out of 14 classes, or 71%).

Daycare observations allowed us to study the quality of reasoning activities by reconstituting the "*Using Language to Develop Reasoning Skills*" subscale of the ECERS-R. This subscale evaluates how educators discuss logic with children when playing with them and provide them with material to develop their reasoning. The scale also assesses the extent to which educators encourage children to explain their reasoning during an activity. On this sub-dimension of the ECERS-R, classes in the Program Daycare group scored an average of 4.67 out of 7 (SD = 1.68), meaning that their practices range from minimal to good. Four educators were given a score of 7 (excellent), most (eight) scored 4, and one educator scored 1 (inadequate). Classes in the comparison group obtained a mean score of 3.37 (SD = 2.06), that is, a quality index just above minimal. Only two of the 19 educators scored 7, nine educators were given a score of either 3 or 4, and five educators scored 1. Almost all educators experienced difficulty at several levels of this subscale. They did not initiate or encourage conversations with children in the context of reasoning games or reason with children when a problem arose.

Table 5.5 presents the average Quality of Educational Content by childcare setting, i.e., the Program Daycare group or the Comparison Daycare group. This table also presents the results obtained by Japel and colleagues (2005) for the mean quality of educational content at 296 daycares in Quebec. The results presented in this table suggest that quality is higher in the Program Daycare group for all subscales measured than in the Comparison Daycare group or the Quebec daycares.

Table 5.5: Mean Scores for Quality of Educational Content by Dimension Assessed and Treatment Group

Quality of Educational Content (based on ECERS–R subscales)	Classes in the Program Daycare Group n = 18	Classes in the Comparison Daycare Group n = 19	Daycares in the Japel et al. (2005) Study n = 296
<i>Books and Pictures</i>	5.00 (2.06)	2.42 (1.98)	3.71 (2.14)
<i>Informal Use of Language</i>	5.33 (1.53)	4.68 (1.67)	4.03 (1.95)
<i>Encouraging Children to Communicate</i>	6.39 (1.58)	5.84 (1.42)	4.47 (1.96)
<i>Using Language to Develop Reasoning Skills</i>	4.67 (1.68)	3.37 (2.06)	3.74 (2.00)

Note: Indices are calculated based on observations for the first year of implementation for the first and second cohorts combined. The number of observations in the second year was too low to calculate these indices.

Pre-Writing

The third dimension observed concerns children’s exposure to *pre-writing*. This dimension measures the frequency at which children trace letters or write their name. These exercises help children refine their fine motor skills, in addition to helping them recognize the letters of the alphabet. This ability is associated with preschoolers’ capacity to decode a text, which is necessary for learning to read (Lonigan, Schatschneider, Westberg, and the National Early Literacy Panel, 2008). During the second observation period, children in 11 of the 18 classes in the Program Daycare group (61%) were observed while tracing or writing their name. A slight increase was observed during the third observation period (March to June), when educators in 12 classes of the Program Daycare group (71%) encouraged children to write letters (the observation for one class is missing). Conversely, 5 of the 19 classes in the Comparison Daycare group (33%) encouraged children to write (the observations for four classes are missing). In the second year, all classes observed in both treatment groups except one class in the Comparison Daycare group (92%) had children write during this period (the observations for two classes in the Comparison Daycare group are missing). It should be noted that the children were older, and therefore more skilled and more interested in this task.

Creativity, Autonomy and Transitions

Three final program elements had to be introduced in order to implement the entire curriculum of the tested program. First, the program placed special emphasis on children’s *creativity*. In this context, the children were encouraged to explore and to build on their own based on their understanding of the environment. The children’s creativity was particularly encouraged during crafts where children were free to choose among various materials and where the finished product was different from one child to the next. This encouraged the child’s autonomy and enhanced the child’s self-esteem. Observations revealed that educators in all classes of the Program Daycare group (100%) allowed children to explore and be creative in their crafts during the first year, versus 14 of the 19 classes (78%) in the Comparison Daycare group (the observation for one class is missing). In the second year, all classes in the Program Daycare group except one (90%) gave children a chance to be creative, while this situation was found in seven classes of the Comparison Daycare group (58%; the observations for two classes are missing).

Second, the program attributed special importance to children's *autonomy*. However, observations showed that there were as many opportunities for autonomy development in the Program Daycare group as in the Comparison Daycare group. In fact, all educators observed over the two years encouraged children to be autonomous by coaxing them to dress themselves or wash their own hands, and allowing them to choose their activity during free play.

Finally, some educators in both groups introduced strategies to facilitate *transitions* between activities, another important element of the tested program. In the first year, at least 75% of transitions went smoothly for 15 of the 18 classes in the Program Daycare group and for 14 of the 19 classes in the Comparison Daycare group. In the second year, although certain educators had trouble with transitions, all the transitions observed (100%) went smoothly for 7 of the 10 classes (70%) in the Program Daycare group and for 12 of the 14 classes in the comparison group (86%).

In short, as presented in Table 5.6, the program's educational content was respected to a degree of 74% by classes in the Program Daycare group for the first year of implementation (with a range of 29% to 100%). On average, 54% of the program's key elements were applied at comparison daycares (ranging from 5% to 97%). All program elements were present at a level equal to or higher in the classes of the Program Daycare group than in those of the Comparison Daycare group in the first year. Nevertheless, certain elements were implemented much more often in the Program Daycare group. In particular, reading was more frequent during circle time, the interactive reading style was identified more frequently, and there was greater emphasis on early writing. In the second year, the level of application rose for both groups. For the Program Daycare group, the average increased to 83% (ranging from 50% to 100%), while it reached 63% for comparison daycares (ranging from 9% to 100%). In other words, both environments shared certain points in common, although the Program Daycare group retained certain specificities. In fact, during the program's second year, daycares in the Program Daycare group implemented all the elements as often as or more often than those in the Comparison Daycare group, except for one element (i.e., effective transitions).

Table 5.6: Average Percentages for Fidelity of Educational Content by Year of Program Delivery and Treatment Group

Fidelity of Educational Content	Year 1		Year 2	
	Classes in the Program Daycare Group n = 18	Classes in the Comparison Daycare Group n = 19	Classes in the Program Daycare Group n = 10	Classes in the Comparison Daycare Group n = 14
Reading				
Reading during Circle Time	93%	53%	100%	45%
Reading More than Once during the Observation	29%	5%	65%	35%
Interactive Reading	87%	24%	78%	9%
Conditions Conducive to French-language Communication				
Circle Time	100%	84%	100%	100%
Songs and Nursery Rhymes	58%	52%	68%	50%
Reasoning	44%	45%	50%	47%
Pre-writing				
Pre-writing	63%	33%	100%	91%
Three Other Elements				
Creativity	84%	69%	90%	64%
Autonomy	97%	97%	100%	100%
Transitions	84%	80%	75%	85%
Average for all 10 elements	74%	54%	83%	63%

5.5. QUALITY OF EDUCATOR–CHILD INTERACTIONS

Observations allowed us to compare the quality of educator-child interactions (i.e., educators' sensitivity) in the Program Daycare group with that in the comparison group. An ECERS–R subscale called *Staff–Child Interactions* was used to determine the educators' sensitivity to children. This subscale measures the degree to which staff members meet children's needs, are warm, respect children, and like to be with them.

The scale revealed that educators in the Program Daycare group and in the Comparison Daycare group were very sensitive to children's needs. The average for the Program Daycare group was 6.22 out of 7 (SD = 1.63), while that of the comparison group was 5.26 out of 7 (SD = 2.31). For the Program Daycare group, 13 of the 18 classes scored 7 (excellent), two scored 6 (from good to excellent), and one scored 1 (inadequate). With respect to the Comparison Daycare group, 10 classes scored 7, two scored 6, and three were given a score of 1. These observations show that there are minor differences between the treatment groups. Both groups show a good quality of interaction between educators and children at the daycares.

5.6. DOSAGE

As mentioned earlier, the program was delivered for only one year in certain communities, but continued for a second year in other communities (see Table 2.1 in Chapter 2 of this report for the program delivery schedule for communities in both cohorts).

Dosage, which is calculated using the average number of hours that children spent at daycare per week, is described based on treatment group and cohort for the two years of the program. On average, children in both treatment groups attended daycare for a similar number of hours per week over the two years (see Table 5.7). Specifically, in the first year, the children in the first cohort of the Program Daycare group attended daycare on average 29.1 hours per week, while children in the second cohort attended on average 26.5 hours per week (average for both cohorts combined = 27.8 hours per week). First-cohort children in the Comparison Daycare group attended daycare on average 28.2 hours per week, while those in the second cohort attended on average 31.5 hours per week (average for both cohorts combined = 29.8 hours per week).

In the second year of the program, first-cohort children in the Program Daycare group attended daycare on average 24.8 hours per week, while second-cohort children attended on average 11.1 hours per week (average for both cohorts combined = 18.0 hours per week). In the second year, first-cohort children in the Comparison Daycare group attended daycare on average 26.9 hours per week, while second-cohort children attended on average 11.2 hours per week (average for both cohorts combined = 19.0). As shown in Table 5.7, in the program's second year, second-cohort children spent less time in daycare than those of the first cohort because they went to kindergarten either part-time or full-time. However, the number of hours spent at daycares per week in the second year is similar for both treatment groups, that is, 11.1 hours for children in the Program Daycare group and 11.2 hours for those of the Comparison Daycare group.

Table 5.7: Dosage (Hours of Children's Daycare Attendance per Week) by Treatment Group and Cohort

	Year 1		Year 2	
	Program Daycare Group	Comparison Daycare Group	Program Daycare Group	Comparison Daycare Group
Cohort 1	29.1	28.2	24.8	26.9
Cohort 2	26.5	31.5	11.1	11.2
Total	27.8	29.8	18.0	19.0

5.7. EDUCATOR REACTIONS IN THE PROGRAM DAYCARE GROUP

Educators' reactions to the program were measured through mid-term and exit interviews of the educators participating in the Program Daycare group. Community coordinators interviewed a targeted sample of head educators or educators responsible for delivering the program in the classroom. The interviews were conducted with the educators either six months after they began participating (mid-term interview) or at the end of program delivery (exit interview). In total, 20 educators from both cohorts agreed to be interviewed about their participation. Four of them were interviewed twice, that is, they participated in both the mid-term interview and the exit

interview. The results show that positive, mixed and negative reactions were reported by educators in both cohorts. Note that the five communities of the first cohort and the two communities of the second cohort are dealt with separately, for a total of seven sites.

Positive Reactions to the Program

Two elements were identified as those most appreciated by educators participating in the project: circle time and the routine chart. First, although circle time is not a new practice, the educators at four of the seven sites reported that they really liked this element of the program.

“I really liked circle time, the way it was done, as well as how the trainer taught us to do it, with lots of songs and storytelling. I really enjoyed that.” (Educator)
[Translation]

Second, the routine chart was unique to the tested program. Section 4.2 describes how the routine chart works. At three of the seven sites, educators identified the routine chart as being a fantastic element of the program. Section 5.9 explains the routine chart’s effects on children.

“The routine chart is excellent; if there is a change during the day, you can do it with the routine. You can explain that one activity is being switched with another, and no one says “Oh! What’s happening?” because something is different, because I know that some kids need stability and structure during the day.” (Educator) *[Translation]*

Mixed Reactions to the Program

Certain program elements resulted in mixed reactions from educators. The main element that sparked this type of reaction is related to encouraging children’s creativity through games and crafts. Although some educators say they really appreciated this element, other educators said the element was hard to implement.

As pointed out by the trainer, creativity was among the key components of the Daycare Program. Educators had to let the children create free crafts and encourage them to be creative in all their games. Educators at four of the seven sites said they really appreciated the creativity component since it opened their eyes to the children’s abilities.

“That’s what I appreciated most (...) the crafts. (...) It’s taking the time to sit with the children and communicate with them, and at the same time seeing their development, what they do, and their originality. (...) So when I see a child’s craft on the wall, I can see his perspective and how he sees things, not what I’ve asked him to do.”
(Educator) *[Translation]*

Despite the positive reactions, educators at two of the seven sites stated that the free crafting did not give children enough of a framework and that they would have benefited from more instructions on how to cut out, draw, or use the crafting materials. This difficulty with the creativity component was also confirmed by the trainer during her visits to program daycares and by her interview responses.

“The children always made the same thing because (...) it was free crafting. No friends went there. So you could see it was kind of a waste of space or time. I would have preferred to make a new craft or explain to them, “This is how we do it,” so that they’d really know how to go about their crafts.” (Educator) *[Translation]*

Negative Reactions to the Program

The program element that generated the most negative reactions was the requirement to always communicate in French, even with children who understood only English. This aspect of the program was mentioned at three of the seven sites.

“With my Anglophones, that [speaking in French] was a bit of a disadvantage, because things can get solved if you say it in English - the child will understand.”
(Educator) [Translation]

5.8. FACILITATORS AND BARRIERS

The educators and trainer were interviewed in order to collect their impressions on delivery of the Daycare Program. Some of the interview questions concerned elements that facilitated or complicated project implementation. Sometimes information was also gathered through the trainer’s observation notes from training and from visits to daycares in the Program Daycare group. The answers of the educators and trainer soon made it clear that an element’s presence was identified as a facilitator while its absence was perceived as an obstacle to the program’s smooth delivery. For example, educators found that having time to plan their activities was a facilitating factor. Educators who were not given this opportunity stated that one obstacle was not having time to plan their activities. To facilitate reading, we wrote the text by presenting elements in a positive way, that is, by referring to facilitators rather than barriers. In reading this report, the absence of a facilitator must be understood as an obstacle to implementation.

The results in this section are broken down into three parts, each of which consists of elements presented in order of importance based on analyses.

Key facilitators in the opinion of the educators and trainer:

- Management’s support for the program;
- Teamwork by educators;
- The material supplied to the daycares;
- The program’s flexibility;
- Follow-up by the trainer;
- The support from community coordinators.

Other facilitators in the opinion of the trainer:

- The educators’ openness to feedback and new ideas;
- Their experience working with children.

Special barriers in the opinion of the educators and trainer:

- Use of time in relationship to program delivery;
- Themes difficult to implement;
- Staff turnover.

Key facilitators in the Opinion of the Educators and Trainer

First, almost all the educators (at six of the seven sites) mentioned that good support from daycare management was important to properly apply the program. The tested program required an additional investment of time compared to the program the educators had been giving before participation in the Readiness to Learn project. Educators had to plan their activities, some of which required that educators prepare material in advance (example: book-related activities). This meant that management's support was essential because it allowed the educators to free up the time necessary to plan the program's required activities. The program's trainer also mentioned that support from the daycare's management was an important element that made a real difference with respect to implementation.

"She [the director] also took the four-day sessions. Yes, I think that if she had found it was not worth doing, she would not have done it. She encouraged us a lot (...) to open up the boxes; she gave us the time to do our planning, to make sure it was done. If we needed anything at all, she also helped us find information using the computer."
(Educator) [Translation]

Conversely, the lack of support from daycare management was considered an obstacle to applying the program. This element was mentioned at five of the seven sites. In fact, lack of support was a major challenge for educators. This poor support from management could be expressed, among other ways, as educators not being given enough time to plan the program's required activities. This element was also mentioned by the trainer as an obstacle that prevented educators from fulfilling their potential.

A second element that facilitated project implementation was the mutual support among educators. In most classes, two educators worked together to take care of a group ranging from 10 to 16 children. Harmony between the two educators greatly influenced the quality of implementation. It led to better planning of activities, as well as a more harmonious application of the program from day to day. Educators at six of the seven sites, as well as the trainer, reported that good teamwork facilitated program delivery.

"I have a good relationship with [the other educators]: pretty much everyone helps everybody else. If I need help with something, I know I don't have to look far. I know that really all the educators would be there if I asked for help." (Educator)
[Translation]

Some educators reported a lack of support from their peers. In fact, at three of the seven sites, educators indicated that working on the same team with a disrespectful or unpleasant educator made it hard to implement the project. This lack of collaboration among educators reduced the quality of the program's application in some classes. The trainer also reported that conflicts between educators affected their ability to properly deliver the program.

Third, the addition of new resources to the daycare was an element that fostered program implementation. In fact, implementation activities provided every daycare in the Program Daycare group with a good deal of material to ensure that every daycare had the tools necessary to apply all the program components. This contribution of material was a significant asset for the majority of sites (six of the seven sites). The trainer also indicated that at several daycares, these resources replaced older material.

“The program gave us a lot of material. We received material for each theme, new material in fact, and it helped us. (...) Material was always there and we always had it ready for every theme. Voilà! Nothing was missing.” (Educator)

Despite the considerable quantity of material given to the daycares, educators at three of the seven sites said they could have used additional resources to present the themes in greater depth.

Fourth, educators at four of the seven sites stated the importance of the trainer’s follow-up visits throughout the project. Educators particularly liked the trainer’s positive approach and her encouragement. The trainer also brought a lot of material and documentation to the daycares in order to facilitate the program’s application. This was another aspect greatly appreciated by educators.

“I really liked that because she said to us ‘Alright, I see the progress in the room. I see the difference.’ Then she said: ‘You did this, so I am just going to show the other daycares how it can be done.’ (...) because we don’t get much encouragement. What I loved were the observations and the feedback.” (Educator)

Fifth, educators at four of the seven sites pointed out that they appreciated the program’s flexibility. The program’s structure and flexibility inspired the educators’ choice of activities with children, which were adapted to the needs of children in the class. This aspect of the program enabled educators to change their routines in response to the day’s requirements.

“Like different activities, different centres and what to place in the centres by age group and theme. You could see how the program was advancing. I found that it was advancing very well, perfectly.” (Educator)

The sixth facilitator identified by educators at two of the seven sites and by the trainer was the community coordinator’s support. The community coordinator was the person responsible for the Readiness to Learn project within the community. Based on the trainer’s observation notes, the coordinator’s support was present in most communities for both cohorts. The community coordinators showed this support by being attentive to staff needs and had an in-depth understanding of early childhood development.

Other Facilitators in the Opinion of the Trainer

The first element noted by the trainer as having an impact on project implementation was the educators’ openness to feedback and new ideas. This element was very important to a program that included regular follow-up visits. Educators who were open to feedback were able to benefit from each of the trainer’s visits and to improve how they were implementing the program.

“Note that [the educator] took my comments and suggestions seriously and to heart. The next day, she set up her room differently. (...) She admitted that already the children were calmer and she had more control.” (Trainer) [Translation]

Educators who were not open to feedback neither benefited from the trainer’s visit nor changed how they implemented the program. Some educators did not take into account the changes suggested by the trainer while others did not appear to properly understand the program’s specificities.

The second facilitating element identified by the trainer concerned the educators’ level of experience. In fact, according to the trainer, the program seemed easier to implement when

educators already had a certain amount of experience working with children. They were then better able to meet the program's requirements because they were knowledgeable about children's needs and managing them. Some educators also had a natural talent with children, which enhanced both the program's application and quality. Along the same lines, educators who participated in the project for two consecutive years were better able to master program elements in the second year.

"It was much easier for those who already understood because they continued to improve. Those who had some difficulty the first year saw an improvement in the second year." (Trainer) [Translation]

Conversely, a lack of work experience and basic training was a challenge to program implementation. In fact, the conduct of certain educators clearly showed their lack of experience. For example, they lacked confidence in applying the program and interacting with children.

Special Barriers in the Opinion of the Educators and Trainer

The first challenge concerns use of time as it relates to delivery of the tested program. In particular, educators at five of the seven sites said they did not have enough time to plan the program. Moreover, educators at three of the seven sites would have liked to have more time to prepare the class in accordance with program requirements. Finally, educators at three of the seven sites found it difficult to find time for the program in a day that was already very busy. This sometimes made it hard to find time to integrate certain activities specific to the program.

"Sometimes you want to do an activity but you run out of time. You realize, oh, it's already such-and-such a time. It's time that's a big challenge." (Educator) [Translation]

The second challenge of the tested program concerns the themes. The program includes a general theme for each month with subthemes for each week. At three of the seven sites, educators reported that certain themes were harder to address because of the lack of educational material given to the classes (e.g., sugar bush, shapes).

Third, stability of staff was essential to the program's success. The program involved a certain level of complexity, requiring a period of adaptation that was more or less long depending on the educator. As a result, project implementation was more difficult at daycares where staff was rotated often. This challenge was mentioned by educators at two sites and by the trainer.

5.9. PERCEIVED EFFECTS

Interviews with 20 educators tell us what they consider to be the program's positive effects on children. Educators also told us about the changes they had made to their own practices with children. Interviews with the Daycare Program trainer were another source of information in addition to the educators' responses.

Perceived Effects for Children

First, several positive effects on children arose from the routine chart. According to educators (at six of the seven sites), the routine chart encouraged children to be independent and to feel calm and secure, as well as to better understand the routine and to develop more of a space–time reference.

Second, educators (at five of the seven sites) reported that the program helped them properly prepare children for school. This preparation involves several aspects, in particular, children’s ability to follow a routine, meet the expectations of the school environment, and develop their autonomy. Several educators also referred to the children’s ability to recognize the letters of the alphabet as well as their name and that of their friends. Children were also beginning to write their names. Educators were surprised at how easily the children learned to do so. Familiarity with the letters of the alphabet is in fact among the best predictors of literacy for children (Lonigan et al., 2008). It is therefore encouraging that educators perceived this change among children participating in the program. As regards literacy, some educators pointed out that children expressed a greater love of books and reading thanks to their regular exposure to them as part of the program.

“I was surprised by that, because I didn’t think it could be that instantaneous (...) within a month a child was able to recognize all the names in his or her group. (...) I found that really impressive, I did not think it could be so effective.” (Educator)
[Translation]

Third, at over half the sites, educators believed that the program helped children improve their understanding of French and enriched their vocabulary. The effect was particularly marked for young Anglophones. The trainer also indicated that the program’s francization strategies had a considerable impact on the children’s French-language development. Finally, some educators identified other positive effects on children, including an improvement in their behaviour, socialization and well-being.

Perceived Effects for Educators

First, for educators at all sites, the program and follow-up training were an opportunity to update their knowledge about child development. A number of them also indicated that the training had helped them feel more confident about their competence.

“Sometimes we doubt our competence as educators because we never know how parents will react; we also have difficulty situating ourselves in terms of laws and regulations, in terms of lots of thing. But hearing someone tell us that what we’re doing is appropriate, that we’re on the right track, really helps us to evolve, to identify new challenges, and to continue our job, which is not always easy.”
(Educator) [Translation]

Second, educators at all sites stated that their participation in the program had made them more focused on children’s needs. This approach made them more attentive to children and led them to create activities related to their interests.

“Now we definitely focus on the child. Now we even listen to the child, like during conversations, that’s where we’ll be starting from.(...) So instead of creating a program and saying OK, this is the theme, we’ll focus on eliciting a topic in

conversation with the child. This is something that's interesting, so why not try bringing it into the classroom?" (Educator) [Translation]

Third, educators at four of the seven sites chose to adopt the weekly program used in the Daycare Program. This grid allowed educators to inform parents about the activities at the daycare, for example, the month's theme and the planned activities.

Fourth, educators at four of the seven sites indicated that they had learned to encourage children to be more creative in their activities (e.g., crafts). This change of practice among educators was also observed by the trainer. This is a significant impact because, according to the trainer, the creativity component is central to the program.

"There are things I never used to do (...) like the crafts, letting them really do it themselves. (...) Now that I give them free rein, I really like it. It really opened my eyes, like 'Wow! Look what they can do!' It's not all the same. That's what's fun on the chalk boards. All this has really opened up my eyes." (Educator) [Translation]

Fifth, some educators mentioned that the program helped them develop some of their work skills and strategies. For example, they improved the daycare routine, alternating the material in the various learning centres, and gave children more support in terms of their autonomy.

Finally, at over half the sites, the educators indicated that, thanks to the program, they improved their relationships with parents. Nevertheless, a few educators in the first cohort admitted they did not work on this aspect. In actual fact, at most sites, educators said that the program did not affect their relationships with parents because these were already fine.

5.10. SUMMARY

Chapter 5 details the results of the implementation study for the daycare component. Various aspects of the program were assessed, including Structural Quality (general quality) of the childcare settings observed, Structural Fidelity, Fidelity of Educational Content, Educative Quality, Quality of Educator–Child Interactions, Dosage, Educator Reactions to the program, Facilitators and Challenges, and the program's Perceived Effects. This section summarizes the results for each of these dimensions.

First, the observation notes by community coordinators revealed that there was a very slight difference in **structural quality** between daycares in the Program Daycare group and those in the Comparison Daycare group, to the benefit of the Program Daycare group. However, these differences are minor and both groups show a good quality (of 5 or more on a scale of 7) for each of the three dimensions assessed: quality of indoor space, health and safety, and greeting/departing. As such, we can say that there is no differentiation in terms of general quality between daycares in the Program Daycare group and those in the Comparison Daycare group. These results are presented in parallel with those of the study by Japel et al. (2005), which assessed the quality of 296 daycares in the province of Quebec as part of the *Étude longitudinale du développement des enfants du Québec*. The results indicate that for every ECERS–R subscale, the daycares participating in the Readiness to Learn project provide better Structural Quality than the Quebec daycares.

Second, the results for the assessment of the program's **structural fidelity** indicate that on average, in the first year, classes in the Program Daycare group implemented 78% of the

Daycare Program elements, while all these elements were found in less than half of the classes at comparison daycares (average of 40%). In the second year, the level of implementation at daycares was very similar, that is, an average of 73% for daycares in the Program Daycare group, versus 42% for those in the Comparison Daycare group. In the first year, all the program elements were present and practiced more often at daycares in the Program Daycare group than at comparison daycares. In particular, the learning centres were more often identified, cards displaying a picture and a word were more present, and a routine chart was displayed and used during the day. In the program's second year, all the elements were implemented more often by classes in the Program Daycare group, except for the presence of cards displaying a picture and a word, for which the rate of compliance was similar for both treatment groups (70% versus 71%). All these results suggest that most program elements were implemented in the classes of the Program Daycare group in a relatively consistent manner.

Third, the results for *fidelity and quality of educational content* revealed that program content was respected to a degree of 74% by classes in the Program Daycare group in the first year of implementation and to a degree of 54% by classes in the Comparison Daycare group. All the program elements were present as often or more often in the classes of the Program Daycare group than in those of the Comparison Daycare group for the first year. In particular, reading during circle time was more frequent, the interactive reading style was more frequently observed, and there was greater emphasis on early writing. In the second year, the level of application rose for both groups. For the Program Daycare group, the average increased to 83%, while the average was 63% for the Comparison Daycare group. This means that both environments shared certain points in common, but that the Program Daycare group retained certain specificities. In the second year of the program, daycares in the Program Daycare group implemented all the elements as often as or more often than the Comparison Daycare group except for one (i.e., effective transitions). As for the Quality of Educational Content, the results indicate that Educative Quality was higher in the Program Daycare group than in the Comparison Daycare group or in the Quebec daycares.

Observations also allowed us to compare the *quality of educator-child interactions* (i.e., educator sensitivity) at daycares in the Program Daycare group with those in the Comparison Daycare group. To measure this dimension, we used the ECERS-R subscale for "Staff-Child Interactions," which measures the degree to which staff members respond to children, are warm, respect children, and like to be with them. The scale showed that educators in the Program Daycare group and those in the Comparison Daycare group were very sensitive to children's needs. The average for the Program Daycare group was 6.22 out of 7 (SD = 1.63), while that of the Comparison Daycare group was 5.26 out of 7 (SD = 2.31). These observations show that the daycares in both treatment groups offer a good quality of interaction between educators and children.

Moreover, observations for *dosage* show that on average, children in both treatment groups attended daycare for a similar number of hours per week over the two years. Specifically, in the first year, children in the Program Daycare group attended daycare on average 27.8 hours per week, while children in the Comparison Daycare group attended on average 29.8 hours per week. In the second year of the program, children in the Program Daycare group attended daycare on average 18.0 hours per week, while children in the Comparison Daycare group attended on average 19.0 hours per week. Average dosages for the program's second year are lower because many children attended daycare only on a part-time basis during that period.

Next, *educator reactions to the daycare program* were gathered through interviews with 20 educators selected in the Program Daycare group. Two program elements sparked positive reactions among educators. First, educators benefited from the content of circle time, which they believed was effective in meeting the needs of children in minority settings. Second, they took advantage of the routine chart, which showed children the order in which the day's activities were planned. What's more, the creativity aspect led to mixed reactions by educators. Although educators appreciated this element, some said they had trouble implementing it. Finally, the program element least appreciated by educators was the requirement to always speak French, even to Anglophone children.

In order to assess *facilitators and barriers*, the educators and trainer were interviewed in order to gather their impressions on delivery of the Daycare Program. Information was also sometimes gathered through the trainer's observations and training at daycares in the Program Daycare group. The educators and trainer identified several facilitators for implementation of the tested program. The main facilitators they identified include management support for the program, teamwork by educators, the material supplied to the daycares, the program's flexibility, follow-up by the trainer, and the support from community coordinators. Conversely, the absence of these factors was an obstacle to project implementation. A few other facilitators were mentioned by the trainer: the educators' openness to feedback and new ideas, as well as their experience working with children. The educators and trainer also mentioned three challenges: use of time in relation to program delivery, the difficulty of addressing certain themes, and educator turnover in the Program Daycare group.

Finally, the information collected from the educators and trainer allowed us to assess the program's *perceived effects* on children and educators in terms of their classroom practices. The educators and trainer saw that the program affected children in several ways. Among other effects, they mentioned that thanks to certain program elements (e.g., the routine chart), children were behaving better, were calmer and more independent, and felt more secure. The program was also more effective in preparing children for school and fostered the development of French-language and literacy skills. For their part, educators learned more about child development through the training. They also changed some of their practices at the daycares, the main being that they focused more on children's needs, adopted the weekly program, and encouraged children's creativity. Finally, some educators found that the program enhanced their relationships with parents.

6. Implementation Review for the Family Literacy Program

This chapter details the findings of the implementation study for the Family Literacy Workshops component for both cohorts of the study. The Family Literacy Program consisted of a series of 10 workshops, each addressing several topics, some of which were mandatory and others, optional. Appendix C summarizes the topics of the 10 workshops. The practitioner served as the facilitator for each workshop, conveying information and then encouraging discussions among the group of parents. Parents also had an opportunity to share their experience, and in so doing, to learn from one another.

The chapter begins with a section (6.1) on **content fidelity**, or the proportion of workshop topics that was addressed. Section 6.2 concerns **dosage**, or the rate of parents' workshop participation. Section 6.3 details the results for the **quality of the delivery** for workshops, while Section 6.4 presents **parents' reactions** to the workshops. The section that follows (6.5) discusses the **facilitators and barriers** encountered during project implementation. Finally, the last section (6.6) concerns the program's **perceived effects** on parents.

Analyses for each program implementation dimension assessed were conducted in two ways. First, analyses were conducted by series of 10 workshops given to 15 separate groups of participants in the Family Literacy Program, that is, 11 groups of participants distributed across six communities for the first cohort and four groups of participants distributed across two communities for the second cohort. Each "group" consists of parents and their practitioner for the 10 workshops, that is, 11 groups for the first cohort and four groups for the second cohort. This first series of analyses enables us to better understand the experiences of a group of parents. Analyses were then conducted for each workshop, thereby bringing together the information for all 15 parent groups for each of the 10 workshops presented. This breakdown gives us an overview of the experiences tied to the content of every workshop.

The following points should be noted:

1. The assessment of the Family Literacy Workshops mainly focuses on the parent component since it aims to equip parents to better support their child's development and school readiness.
2. The analyses include data collected in the six communities of the first cohort and in the two communities of the second cohort, for a total of eight sites.²⁶
3. Only 6 of the 10 workshops given to the *second cohort* were observed while all 10 workshops given to the *first cohort* were observed. Specifically, for the second cohort, observations were conducted for the workshops deemed most relevant to the program: workshops 1, 2, 5, 6, 9 and 10. To make it easier to gather data for both

²⁶ Note that the implementation review for the Daycare Program is based on data collected in the two communities of the second cohort and five of the six communities in the first cohort. The data for this sixth community were not used in the analyses because the program was never implemented at the daycare.

cohorts, we are reporting only the observation notes gathered for these six workshops in the implementation review for the Family Literacy Workshops.²⁷

4. In addition to the observation notes, the project's implementation study is based on the three following sources: (1) the parents' workshop appraisals filled out at the end of the first nine workshops, a logistical survey filled out at workshop 10, and a post-intervention survey filled out two weeks after the last workshop; (2) the assessment by the practitioners who led the workshops, that is, through interviews or feedback in a journal; and (3) parents' workshop attendance.
5. This report does not provide the results by cohort. Likewise, there is no direct comparison of the degree of implementation for the two cohorts. This decision was made because there were fewer than five practitioners who delivered the Family Literacy Workshops for the second cohort, which means that presenting the results by cohort would jeopardize the anonymity of these practitioners.

6.1. CONTENT FIDELITY FOR THE FAMILY LITERACY WORKSHOPS

Fidelity refers to the integrity of the program applied: in other words, the degree to which the applied intervention matches the planned program. As part of the Family Literacy Workshops, the program included certain topics that practitioners were required to address, in addition to optional topics. The observation notes tell us about coverage of each workshop's mandatory topics. This information is rounded out by an analysis of the flip chart content used by the practitioner during delivery, such that a topic is reported as not covered if it is found in neither the observations nor the flip chart.

To compare the fidelity of one series of workshops to that of another, the percentage of mandatory topics covered in the six workshops observed was calculated by parent group. The results indicate excellent coverage of the topics in the 15 parent groups (average coverage of 98%). As seen in Table 6.1, at least 92% of the mandatory material was addressed with all parent groups.

²⁷ This decision was made because of the very high rate of fidelity obtained in the 11 series of workshops for the first cohort (average coverage of 96% for all series of workshops; *Readiness to Learn in Minority Francophone Communities: Project Implementation Report* (Bérubé et al., 2014)). The expression "series of workshops" in the *Project Implementation Report* refers to the "parent groups" that participated in the workshops in this report.

Table 6.1: Content Coverage by Parent Group

Parent Group	% of Content Coverage
1	92%
2	93%
3	97%
4	97%
5	100%
6	93%
7	100%
8	100%
9	100%
10	96%
11	100%
12	97%
13	97%
14	100%
15	100%
Average for both cohorts	98%

Fidelity by workshop was calculated based on the content covered in each of the six workshops observed, all parent groups combined. Table 6.2 shows that 87% to 100% of the mandatory topics were addressed in each workshop for an average of 98%, which indicates excellent coverage. Workshop 9 stands out from other workshops due to its 87% coverage of mandatory topics. According to the observation notes, this workshop's material was covered completely in 6 of the 15 parent groups, but in 8 other groups, the mandatory activity called "*C'est ma communauté*" [This is my community] was not carried out. It should be noted that all the workshops gave parents several opportunities to discuss the services available in their community (the topic of the omitted activity), and practitioners distributed information and brochures on community services to parents on a number of occasions. This may be why the practitioners may have deemed another discussion of this topic as irrelevant.

In short, the observed workshop content for both cohorts was covered very well and the only topic covered less well was presented in other workshops. Presentation of this topic may therefore be planned differently if the Family Literacy Workshop program is given again.

Table 6.2: Content Coverage by Observed Workshop

Workshop Number	% of Content Coverage for Both Cohorts
1	100%
2	98%
5	100%
6	100%
9	87%
10	100%
Total	98%

6.2. QUALITY OF FAMILY WORKSHOP DELIVERY

“The staff received us very well, meaning that they put us at ease from the moment we arrived and guided us throughout the evening.” (Parent) [Translation]

“Everything was very well explained; very good workshop and very interesting subject.” (Parent) [Translation]

The coordinators’ observations tell us about the quality of delivery for the Family Literacy Workshops given to the 15 parent groups. The observers noted practitioners’ skill in conducting the workshops and the difficulties they encountered during delivery. First, the observers reported that in most cases, practitioners were skilled in leading the workshops. In particular, practitioners were successful in conveying their message and were perceived as being welcoming. They also encouraged parents to participate without making judgments about their comments, questions and answers, which fostered an environment conducive to discussion and learning.

The observers also noted that some practitioners showed certain difficulties or challenges that may have impeded effective transmission of the message to parents. The main difficulties noted involved the practitioners’ style of delivery. For example, some practitioners imposed answers during discussions, cut off participants who were speaking, had trouble conveying the message, and covered the content too quickly. Observations revealed that three practitioners repeatedly had difficulty during the workshops. Two of them consistently covered the material too quickly and in a manner that too closely resembled a lecture. The third tended to be disorganized and had trouble leading in general. This means that for these three groups, parents did not receive an optimal level of quality in terms of the Family Literacy Program. In this regard, the quantitative analyses presented in the impact study of this report provide more information about how practitioners’ quality of delivery for the Family Literacy Workshops affects parents’ knowledge, beliefs and habits.

Another indicator of the quality of delivery noted by the observers was the ability to respect the workshop’s allotted time. In general, practitioners had trouble staying within the time granted to them. Most tended to take longer than the workshop’s allotted time. Delivery of the workshop’s parent component was supposed to take 55 minutes. Several practitioners took from

60 to 80 minutes to convey the content to parents. In this regard, observers noted that the workshops contained too much material to cover within the allotted time for the parent component. It should also be mentioned that parents really enjoyed talking to one another. The workshops should therefore include longer discussion periods, even if this means exposing parents to less material.

The last indicator of quality observed concerned the spatial environment in which the workshops took place. At most sites, seating arrangements encouraged discussion among parents. At one site, two parent groups were combined on two occasions to form a single group led by two practitioners. On both occasions, the parents had to be placed in two rows, which made discussion among participants difficult.

Parents also evaluated the quality of the delivery through appraisals filled out at the end of each workshop. First, the findings indicate that with the exception of one group of parents, the participants found the practitioner excellent. Moreover, results show that several parent groups identified the practitioner as among the workshops' strong points. More specifically, these parents said that their practitioner was dynamic and presented the information in a way that made the topic interesting. Most parent groups reported that they particularly benefited from the concrete examples, tips and anecdotes shared by the practitioner. They also said they benefited from the wealth of information presented at the workshops. What's more, some parents said they appreciated the practitioners' interactive and informal approach to the workshops. Finally, all parent groups mentioned that the workshops would have been more pleasant had they been more spaced out in time, for example, every two weeks instead of every week.

The results of the logistical survey answered at the beginning of the 10th workshop provide more information about practitioners' skills. More specifically, the majority of parents found the practitioners excelled in several regards, that is, their knowledge of the topics, their preparation, their general organization, their style and workshop delivery, their sensitivity to the group's needs, as well as their ability to create an environment conducive to discussion and learning.

In short, both observers and parents reported the quality of the workshop delivery to be good. In fact, most practitioners appeared willing to conduct the Family Literacy Workshops in an appropriate manner and were considered engaging and respectful. However, one practice that was widespread among practitioners was exceeding the allotted time for each workshop, which in turn negatively affected the quality of the delivery. This shortcoming in time management may be attributed to the considerable quantity of content that had to be addressed at each workshop of the parent component. In addition, several practitioners experienced difficulties with workshop delivery. Finally, except for two occasions, the spatial environment of the workshops' parent component encouraged participation.

6.3. DOSAGE

Rate of Participation at Each Site for Both Cohorts

The average rate of participation for participants in both cohorts was 63.5% and ranged from 38% to 81%. For the purpose of clarity, results are shown in Figure 6.1 by site rather than by parent group. For the purposes of this report, the communities of the second cohort are counted

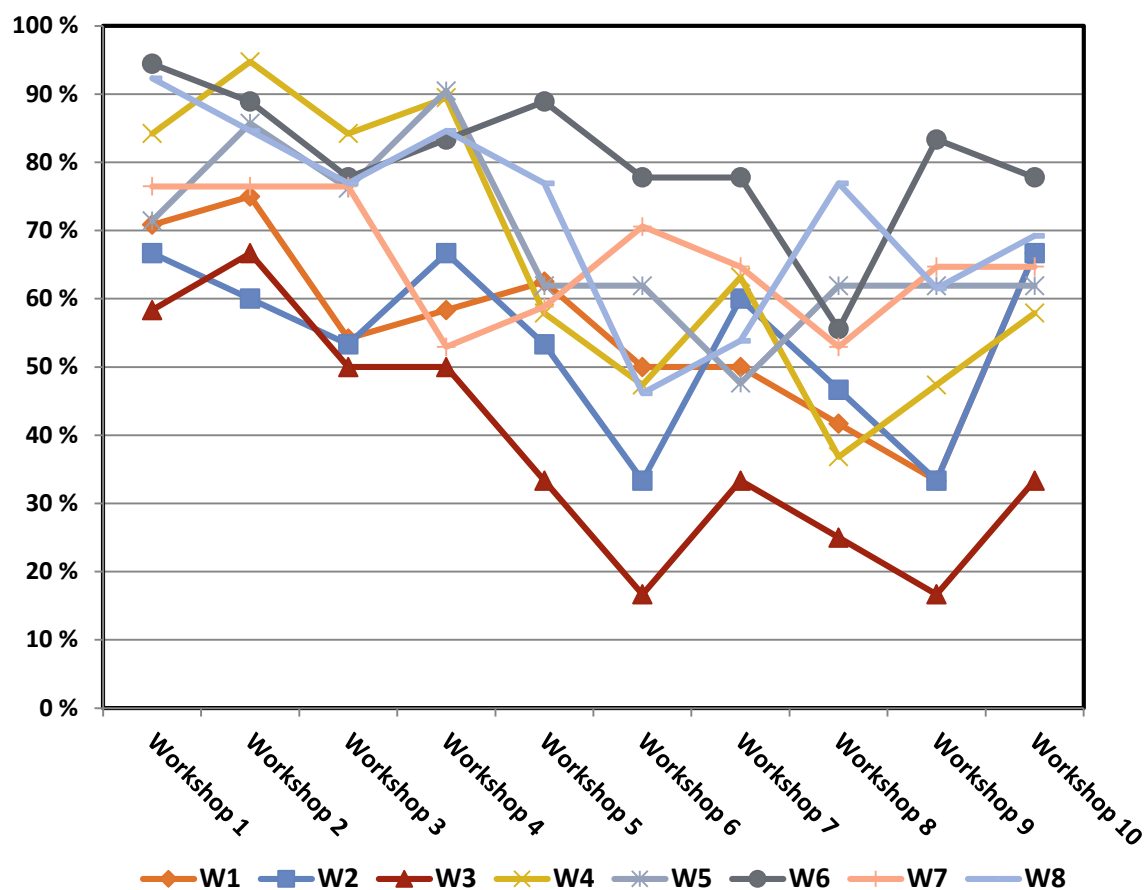
with the communities of the first cohort, which makes for a total of eight participating sites (six sites for the first cohort + two sites for the second cohort).

For the first site (A1), the rate of workshop participation ranged from 33% to 75%, with an average of 56%. It was also observed that 2 of the 23 families (9%) participated in two or fewer workshops. For the second site (A2), the rate of workshop participation ranged from 33% to 67%, with an average of 54%. At this site, 4 of the 15 families (27%) participated in two or fewer workshops, and of these, three families participated in none of the workshops. For the third site (A3), fewer families participated in the workshops than at other sites with a comparable number of parents enrolled in the Program Daycare group. A total of three practitioners working with parents was required to deliver the Family Literacy Workshops at this site. This situation was the result of a significant staff turnover. The rate of workshop participation at this site ranged from 17% to 67%, with an average of 38%, which is considerably lower than at other sites where the workshops were attended by at least half the parents. Moreover, 4 of the 12 families (33%) attended two or fewer workshops.

Many parents at the fourth site (A4) participated in the Family Literacy Workshops. In fact, 19 of the 20 families enrolled in the Program Daycare group participated in the workshops. Only one family (5%) did not attend any of the workshops. The rate of workshop participation ranged from 37% to 95%, with an average of 66%. At the fifth site (A5), the rate of workshop participation ranged from 48% to 90%, with an average of 68%. Only one family (5%) participated in two or fewer workshops. Finally, the sixth site (A6) experienced the highest rate of family participation, with a workshop participation rate ranging from 56% to 94%, and an average of 81%. Only one family (6%) participated in two or fewer workshops.

At the seventh site (A7), the rate of workshop participation ranged from 53% to 76%, with an average of 66%. At this site, only one family (6%) participated in none of the workshops (6%) and three other families attended two or fewer workshops. However, among the 18 families at this site, 10 of them (56%) participated in at least 8 workshops. Finally, the eighth site (A8) had the second best rate of family participation, ranging from 46% to 92%, and an average of 72%. At this site, all families participated in the workshops and only one family participated in two or fewer workshops. What's more, 9 of the 13 families (69%) participated in at least 8 out of 10 workshops.

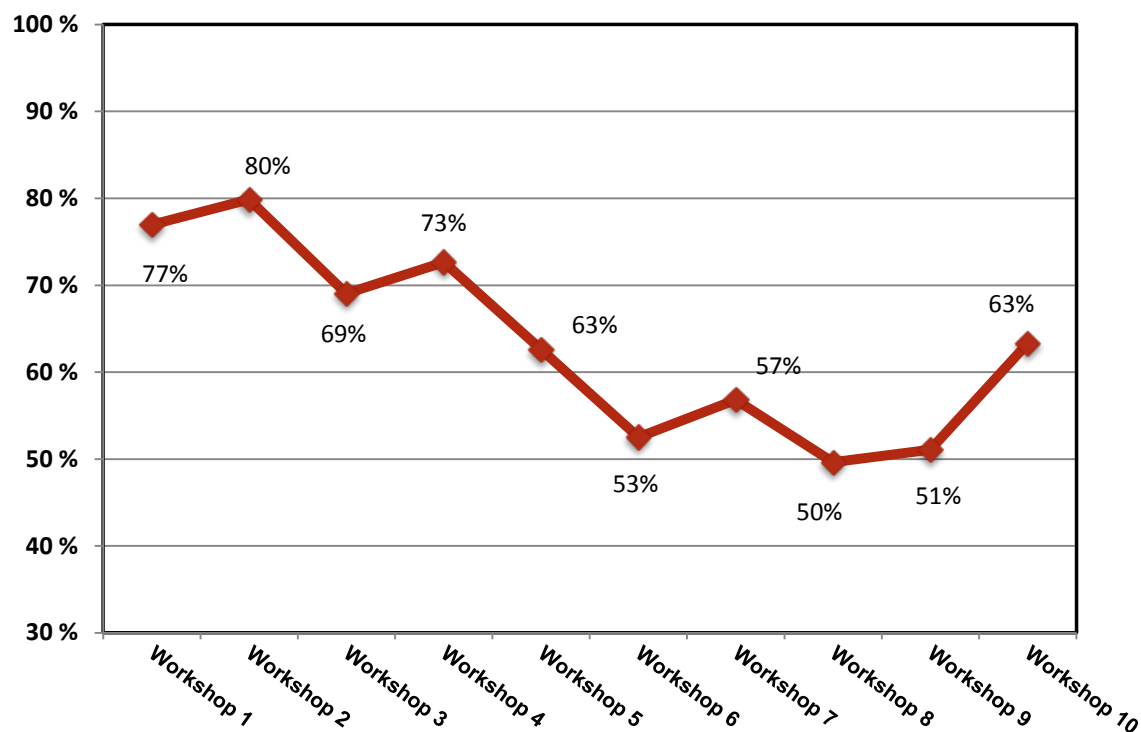
Figure 6.1: Parental Attendance by Workshop and Site



Participation Rate by Workshop

The average participation rate by workshop was 64%. Based on the pattern observed in Figure 6.2, the rate of participation was highest for the first four workshops (average rate of 75%) than for the last six workshops (average rate of 56%). As anticipated, the highest participation rates were for the first two workshops, a time when the participants are weighing the workshops' relevance to them. Workshops 6, 8 and 9 are distinguished by a low participation rate ranging from 50% to 53%.

Figure 6.2: Rate of Parental Participation by Workshop



6.4. PARENTAL REACTIONS TO WORKSHOPS

“My child enjoyed the time we spent together doing an activity and eating with his friends.” (Parent) [Translation]

“As usual, I loved discussing various subjects with the other parents.” (Parent) [Translation]

“It was a really positive experience. We had fantastic parents. With the groups, we were talking at the same time, we got along so well...” (Practitioner) [Translation]

Parents’ reactions to the workshops were first captured through observation notes. Parents reacted positively to most workshop topics. Thus, observers described parents as being participatory, attentive, relaxed or enthusiastic for most workshops. Among the six workshops observed, workshops 1, 2 and 5 sparked the most reactions among parents. The topics covered in these workshops gave rise to both negative reactions from most parents (incomprehension, misgivings, or disagreement) and positive reactions (high level of participation and interest). Note that the first two workshops presented the Readiness to Learn project, the details of the Family Literacy Workshops and the concept of the parent as a child’s first educator, while workshop 5 focused on helping the child develop self-discipline. As regards the first two workshops, parents’ reaction may have been partly attributable to the new situation and to their discomfort. Workshop 5, however, was more likely to produce strong reactions because it covered a topic relevant to participants, but also central to parental identity. Finally, observers reported that participants benefited a great deal from workshop 9, which focused on life in the

Francophone community. The majority of parents had a very positive reaction to this workshop, with very few negative reactions.

Parent reactions were also captured through the appraisal filled out at the end of every workshop. The questions in this appraisal elicited parents' opinions about every workshop they attended. According to these appraisals, parents generally appreciated the workshop content. Three favourite topics can be identified through the comments of over half of the parent groups. First, parents particularly liked the topic of parent-child communication, which covered effective strategies for conveying a message to a child and the importance of active listening when speaking with the child (a topic from workshop 7). Next, parents liked learning about multiple intelligences, an approach that factors in all the dimensions of a child's unique cognitive potential (a topic of workshop 2). Finally, parents appreciated the topic of emergent literacy, which included strategies to develop literacy skills (a topic of workshop 3).

Parents were also asked to answer the logistical survey administered at the beginning of the last workshop. Specifically, parents were asked to answer multiple-choice questions on workshop appreciation and to openly express their opinions on what they appreciated about the workshop, what they had learned and what they would have liked to learn. The results of the closed-ended questions show that most parents favourably assessed workshop content and delivery. They rated workshop goals as often clear (90%) and workshop organization as good or excellent (94%). Concerning the content, the results show that parents found the topics discussed during workshops to be often interesting (84%), sometimes or often useful (92%) and often relevant (70%). Answers to the survey's open-ended questions indicate that all parent groups appreciated workshops 2, 5 and 7. The most popular topics for workshop 2 concerned the concept of the parent as a child's first educator and the concept of multiple intelligences. For workshop 5, parents particularly enjoyed the topic on developing a child's autonomy. Finally, for workshop 7, parents benefited from the content on communication.

The results of the logistical survey tell us not only about the assessment of the workshop content per se, but also give us a more general overview of the meetings. For example, most parents were satisfied or very satisfied with their experience of the Family Literacy Workshops (91%). Most also reported that their child had fairly or really enjoyed participating in the Family Literacy Workshops (95%). This finding is also reflected in the workshop appraisals, in which many parents said that their child looked forward to the workshops. Parents especially enjoyed talking with other parents. This is the element mentioned by the most parents, both in the workshop appraisals and in the logistical survey. The second element mentioned most in both types of survey was the activity with the child as part of the parent-child component. Parents really enjoyed this special time with their child and several of them said they would have liked more time for the parent-child component. Finally, several parents reported that access to French-language resources was a concrete benefit of the Family Literacy Program.

Through practitioner interviews and feedback in their journal, most practitioners indicated that parents seemed generally interested in the workshops and participated actively. They also said that workshop 2 (*"Je suis le premier éducateur de mon enfant"* [I am my child's first educator]) and workshop 6 (*"Le développement langagier, culturel et identitaire de mon enfant"* [My child's linguistic, cultural and identity development]) stand out as those having generated the most parental interest and participation. Further, certain practitioners mentioned that parents appreciated the French-language resources made available to them as part of the workshops.

In short, information from various sources suggests that overall, parents reacted positively to the workshops. A number of favoured topics were identified among those addressed in workshops 2, 3, 5, 6, 7 and 9. Note that workshop 2 was mentioned by the most sources, that is, by observers, parents and practitioners. This workshop addressed the parent as a child's first educator, including the subthemes on scaffolding, parenting styles and multiple intelligences. Workshops 3, 5 and 7 appear to also involve favoured topics since they were mentioned by more than one source. Finally, the three other elements that stand out in the analyses are the appreciation for discussions among parents, the activity with the child as part of the parent-child component, and French-language resources.

6.5. FACILITATORS AND BARRIERS

Programming for the Family Literacy Workshops was designed based on certain best practices identified in earlier studies on family literacy (LeTouzé, 2005, 2006, 2007; Phillips, et al., 2006; Raikes, Summers, and Roggman, 2005). For example, research shows that workshop attendance is greater when programming includes children, the proposed schedule is most convenient to participants, practitioners are qualified, topic presentation is adapted to the clientele, and there is access to learning kits. Effective intervention models for family literacy also favour an adult education approach that builds on the participants' prior knowledge and skills.

A comprehensive review of all data collected within the framework of the Family Literacy Workshops has identified the program's main facilitators and barriers. These findings stem from data collected from the practitioners who gave the workshops (through interviews or their journal), from parents (through workshop appraisals, the logistical survey and the post-intervention survey), as well as from observations by the community coordinators.

Facilitators

"The children have fun and it's a great opportunity to meet other parents." (Parent)
[Translation]

"Parents adore the Resource Centre we set up for them." (Practitioner)
[Translation]

As part of the workshops, a few strategies were successful in attracting and maintaining parental participation in the workshops. Analysis results confirm the considerable support of practices implemented during programming in this regard. The first winning element was to provide parents with a meal. Thus, for evening workshops, parents arrived at the daycare after their day of work and were served a meal. Also, for the Saturday workshops, lunch was served at noon. Meal time was an opportunity for parents to spend time with their child, exchange information with their child's educator, who was also present, and chat with other parents.

The second facilitator was the parent-child component. In fact, the children were very happy to be given time to show their parents what they had accomplished during the workshop's child component and to do a new activity with them. The children's presence prompted some parents to stay at the end of the day, since the children were enthusiastically waiting to share a meal with them.

The third facilitator involved the educators' presence during the workshops. This gave parents and educators a chance to share information about the child, discuss his or her progress at daycare and at home and get to know one another, thereby improving their relationship.

The fourth facilitator was the quality of the workshops. Several parents commented on this aspect of the workshops, as much in terms of the relevant and interesting topics for parents with young children as in terms of the practitioner's skill in effectively leading the workshops. As suggested by LeTouzé (2005), the practitioners' qualifications are vital to the success of family literacy programs. Note that all practitioners were given training. One trainer remarked that the practitioners in training appeared interested, motivated and competent. For their part, a number of the practitioners said they benefited from an overall deepening of their knowledge and from learning delivery strategies through practical activities, concrete examples and discussion. They also reported that they were very satisfied with the training in terms of content, relevance, process and delivery. Practitioner training was thus an opportunity to foster workshop quality. In turn, workshop quality encouraged parents to participate in the discussions and exercises of the parent component. Further, parents noted that they learned a lot from discussions and exchanges with other parents attending the workshops.

Finally, parents really enjoyed the fact that the workshops gave them access to resources through the learning kits. Also, after the workshops, children and parents could borrow a French-language resource for the week (a book, CD, DVD or board game). As noted by the practitioners, these resources were appreciated not only by parents, but also by children, who showed a great deal of enthusiasm about borrowing learning material.

Barriers

"Consider making the lessons every two weeks." (Parent) [Translation]

"Way too much material; I took much more time than was allotted." (Practitioner) [Translation]

The main barrier to parental participation was the demand that participating in a weekly workshop represented in terms of an already busy work-family life. To this effect, several parents suggested that workshops be more spaced out in time. Others said they would have preferred that workshops be given over a two-year period, thus enabling them to participate less intensively over a longer period of time. Also in relation to scheduling, some parents found that their child was too tired during the workshops, especially those held at the end of a day. In fact, certain workshops were drawn out until 8 p.m., a time of day when, according to parents, their child was too tired to benefit fully from the workshops. Finally, a number of parents pointed out that the workshops occasionally covered too much content.

What's more, practitioner interviews and feedback in their journal provide information about the challenges they had to overcome and what they would do differently to improve delivery of the Family Literacy Workshops. Just as parents had expressed, most practitioners giving the parent component said they ran out of time and were unable to cover the entire workshop's content, which sometimes translated to glossing over the material. Some of them also said they would change a few aspects of the workshop content. More specifically, they would change the order of the topics (e.g., bilingualism as one of the first workshops), present less theory and break the topics down differently among the workshops to better connect them with the activities

of the parent–child component. Practitioners who gave the child component mentioned that the workshop routine and delivery worked well. However, some said they would have benefited from more preparation before the workshops, for example, learning the songs ahead of time.

6.6. PERCEIVED EFFECTS

“I used the things we saw in the workshops at home. I intend to use them in the future.” (Parent) [Translation]

“We are role models for our children and they’ll try to become like us over the years. Even though my kids live in an environment where the French language is less and less valued, I will continue to promote this language and give my kids opportunities to experience French-language activities.” (Parent) [Translation]

“We found that now the parents make more of an effort to talk to their children in French, at least, we saw this happening at the workshops.” (Practitioner) [Translation]

The parent surveys (i.e., workshop appraisals, the logistical survey and the post-intervention survey) included questions on what they had learned during the workshops and what they had put into practice after the workshops. Several parents reported having learned many new things at the workshops, in particular, the workshops on communication (workshop 7), French at home (workshop 6), and autonomy and discipline (workshop 5). Parents also said they had applied the strategies learned at the workshops. One common response on the logistical and post-intervention surveys concerned introducing strategies they had learned to communicate with their child. These practices were discussed in the workshop on communication (workshop 7), which included a situation scenario intended to show parents the importance of relating to children at their level so that they properly understand the message being conveyed. Several parents adopted this way of communicating with their child after the workshop.

A second common response in both surveys (logistical and post-intervention) concerns the parents’ reported new awareness of how important it is to expose their child to French. Introduced at workshop 6, a number of parents said this awareness increased their use of French at home as well as the number of activities they do in French (e.g., reading, movies, television shows). Third, parents pointed out that the workshops had made them more aware of parents’ role in their child’s development. In fact, parents reported learning and applying the concepts concerning children’s emotional control, autonomy and discipline (workshop 5). Note that according to observers, parents showed negative reactions during this workshop as well as immediate positive reactions. Although the workshop made parents react on the spot, they appear to have benefited from it in the longer term.

In addition, numerous parents mentioned that the Family Literacy Workshops prompted them to spend more one-on-one time with their child through play and literacy activities. Thus, most parents reported that participation in the Family Literacy Workshops had a positive effect on the parent-child relationship. Finally, several parents found that the workshops became a means of finding concrete solutions to the challenges of dealing with children, work-family balance, and learning French in a minority context. In this regard, parents said they learned a lot by talking

with other parents. The chance to talk with other parents was a key benefit of the Family Literacy Workshops.

The community coordinators' observations confirmed the effects of the Family Literacy Program reported by parents. Observers noted significant changes among the parents who participated in the workshops. They found that parents were more attentive to their child, communicated more often at the child's level, performed more French-language activities with their child and adopted new parenting techniques associated with developing the child's autonomy, self-discipline and self-esteem.

For their part, practitioners noted that the workshops raised parents' awareness regarding the role parents play in their child's development by providing them with new information and by allowing them to review and consolidate concepts they already knew. Practitioners also stated that the workshops were an opportunity for parents to spend quality time with their child and at the same time improve the parent-child relationship. A few practitioners who gave the parent component found that the workshops also led to exchanges among families, which enhanced the sense of belonging to the community. Finally, some practitioners who led the child component said that the contact with educators was beneficial for parents because this clarified the educator's role and the Daycare Program's goals, particularly the goal of school readiness.

All the evidence suggests that the Family Literacy Workshops had a positive effect on parents. According to various sources of information, one effect of the program is to educate parents about their role in their child's development and the importance of exposing children to French in a minority Francophone environment. The results of parent surveys and observations suggest that parents learned new things, applied new parenting techniques taught at the workshops and increased the number of French-language activities they carry out with their child. Finally, more than one source mentioned the workshops' positive effect on the parent-child relationship and the contribution of discussions among families about learning.

6.7. SUMMARY

This chapter reports the implementation results for the Family Literacy Program for both cohorts of the study. This program consists of a series of 10 workshops, each addressing several mandatory and optional topics. There were 15 separate groups of participants in the program, that is, 11 groups of participants spread out over six communities for the first cohort and four groups of participants spread out over two communities for the second cohort.

The integrity of the Family Literacy Program was measured based on six dimensions: Content Fidelity, which is the proportion of the program covered during the workshops; the quality of the practitioners' workshop delivery; Dosage, that is, parents' rate of participation in the workshops; Parents' Reaction to the workshops; the Facilitators and Barriers encountered during project implementation; and the program's Perceived Effects on parents. The paragraphs that follow explain the analysis results for each of these dimensions in turn. Findings for Content Fidelity stem from observation notes, while results for other indicators are from several sources, including observation notes, parent surveys, and practitioner feedback through interviews or journals.

An assessment of **fidelity** of the Family Literacy Workshops shows that 98% of the mandatory content was covered. Both observers and parents generally rated favourably the **quality** of the workshop delivery. Most practitioners had the qualities necessary for effective workshop delivery. The only negative point was that they tended to exceed the time allotted for each workshop. The topics addressed during the workshops also sparked positive **parent reactions**. Among the 10 workshops, the most popular topics were presented at six workshops. Moreover, parents identified three other elements they appreciated: discussions with other parents, the activity with their child as part of the parent–child component, and access to French-language resources.

The average **participation rate** was 63.5%. Participation varies considerably by family. Some families participated in the workshops on a regular basis while others rarely attended. Analyses by site show that one community in the first cohort had a lower rate of participation than other sites.²⁸ The average participation rate for each workshop points to greater participation in the first four workshops than in the last six.

In the view of parents, observers and practitioners, the main **facilitators** for participation by parents were the meals for families, the parent–child workshop in which parents spent time with their child, and the presence of educators during the workshops. Workshop quality and access to resources also encouraged parental participation. However, certain **barriers** impeded parental participation. In particular, parents said that the weekly scheduling of the workshops was too difficult in the context of a busy work-family life. Finally, practitioners reported that they did not have enough time to cover all the content planned for each workshop, which often made the workshops longer.

Two key **perceived effects** of the Family Literacy Workshops are the parents' awareness of their role in their child's development and the importance of French in minority Francophone environments. Furthermore, parents appeared to have learned new things and to have adopted new parenting techniques. Finally, parents used French more often when doing activities with their child.

²⁸ Note that the two communities in the second cohort were added to the six in the first cohort for a total of eight sites.

7. Preliminary Analyses

This chapter describes the steps used to validate quantitative data from both cohorts of the Readiness to Learn project. These preliminary analyses have two primary goals: to optimize the database and to determine the data's limitations with respect to the external²⁹ and internal³⁰ validity of the results.

The following sections discuss respectively: the quality control process (Section 7.1), management of missing values (Section 7.2), identification of confounding variables (Section 7.3), and assessment of the project sample's representativity (Section 7.4). The chapter ends with a summary of the preliminary analysis as well as its effects on the impact assessment for the tested program (Section 7.5).

The following points should be noted:

1. Unless stipulated otherwise in the text, the results reported in this chapter concern both cohorts rather than one particular cohort.
2. The preliminary analyses reported in this chapter concern four communities: Orleans, Cornwall and Durham in Ontario and Edmundston in New Brunswick. The communities of Edmonton, Alberta, and Saint John, New Brunswick, were excluded from the analysis for methodological reasons: the Daycare Program was never implemented in Edmonton and there was no Comparison Daycare group for either Saint John or Edmonton. Excluding these communities allows us to report results that are valid and more coherent.
3. The first cohort of participants consists of families from all four communities, while the second cohort consists exclusively of families from Orleans and Cornwall.

7.1. QUALITY CONTROL PROCESS

SRDC applied a rigorous quality control process for data used in the preliminary analyses and the impact analyses. A short description of these steps is given in Section 4.1 of the *Readiness to Learn in Minority Francophone Communities: First Cohort Findings Report* (Thompson, Legault, Lalonde, & Bérubé, 2014; short title: *First Cohort Findings Report*). These precautions ruled out errors from the outset and optimized the database in accordance with the basic assumptions of each analysis (e.g., no outliers).

²⁹ Concerning the following question: "Is it reasonable to assume that the effect would be obtained with a sample representative of the target population?"

³⁰ Concerning the following question: "To what degree can we state that the reported effects are solely the result of the tested program?"

7.2. MISSING VALUES ANALYSIS

Missing values in a database can threaten the validity of an analysis. This threat stems from two main sources. First, cases with missing values for a variable are usually excluded from an analysis, which may be problematic when they have characteristics that differ from those that remain in the analysis. If the sample composition changes after a subsample of children are excluded, this has implications for the external validity of the results. For example, if the missing values occur disproportionately for girls, the results may not generalize to this population. Similarly, the internal validity of the estimates of the program effect depends on the stability of group composition over time. If missing values disrupt group composition, then this may bias estimates of the program effect. Next we present an analysis of the missing values that assesses the potential for both types of bias. The strategies applied to cancel out these effects are also identified.

7.2.1. Breakdown of the Missing Values

The missing values of a database can be divided into two main categories: planned and unplanned. Under the Readiness to Learn project, planned missing values are generated by the use of start rules for administration of the ÉPE-AD. Unplanned missing values are ubiquitous and occur for two reasons: a) inability to administer a measurement instrument for a given test period (complete missing); or b) the result of a process whereby a participant failed to answer a particular question or was not administered a particular scale (partial missing). The frequency and treatment of both types of missing data are explained below.

Planned Missing Values

Planned missing values were generated by the use of a start rule when administering the ÉPE-AD to children in the first cohort. The frequency and treatment of these missing values are described in detail, respectively, in Sections 4.2.1 and 4.2.3 of the *First Cohort Findings Report* (2014). In the report at hand, it is sufficient to reiterate that the start rule meant that certain children skipped items that they might have failed, requiring the use of an imputation algorithm to estimate missing values.

After the experience with the first cohort, a different strategy was adopted for the second cohort. Thus, no start rule was applied before the fifth assessment when it became clear that most children easily answered the first items of every ÉPE-AD scale. Practically speaking, we chose to use a “customized” start rule³¹ for each child based on his or her performance for the prior assessment. This approach meant that the maximum value of the skipped items could simply be imputed without losing information.

³¹ In principle, the items of every ÉPE-AD scale are ordered by level of difficulty. As a result, a child’s probability of achieving the maximum score for an item is greater at the start of a scale than at the end. Typically, a child over the age of four has no problem with the first x items of a scale, only to fail with the more difficult items. Hence the generic start rule adopted for the first cohort (see Section 4.2.1 of the *First Cohort Findings Report*, Thompson et al., 2014) and the customized rule applied to children in the second cohort at their fifth assessment. Specifically, the start rule for the second cohort stipulated that scale administration began with one item from toward the end of a series of consecutive correct answers observed at the prior assessment [start item = item number $(x - 2)$]. With this rule, a child who achieves the maximum score for the first five items of the fourth assessment would begin the scale at the third item for the fifth assessment provided the language of test administration remained the same. If the language of administration changed from one assessment to the next, scale administration would begin with the first item.

Unplanned Missing Values

Unplanned missing values are presented in the data from assessments and surveys. These missing values may result from a survey or an assessment that was not received. Moreover, data from surveys are sometimes incomplete, which means that valid data are obtained for certain questions but not others, either because the interviewer made a mistake or because the participant refused to answer.³² Of the two types of missing values, the second may be more problematic since the mechanism that resulted in the missing value may be linked to the characteristics of the question itself.³³ The paragraphs that follow concern the sources of missing values as well as their effects on selecting an imputation strategy.

At each assessment period, there is a low percentage of complete missing data. These are due to the non administration of the survey or to participant attrition (see Chapter 3). Data that is missing due to participant attrition are problematic because they limit the imputation options for data in a longitudinal design. For example, in a longitudinal design, the same question can be asked several times in order to increase the chance that the desired information will be obtained from all participants. To illustrate this in the context of the project at hand, we asked about family income three times in the first two years of the study.³⁴ For participants who remained in the project until the eighth survey, such a redundancy can be used fully by imputing the missing values for these questions when they occur.³⁵

Another source of missing values in the database is surveys that contain incomplete data. As noted earlier, this type of missing data can lead to problems with data analysis. The main issue is to determine whether some aspect of the question systematically increases the likelihood that a participant will not answer it. For example, it is well known that respondents at the extreme ends of an income bracket are more likely to not respond to income-related questions. In this example the process that produces the missing values is clearly systematic. If the mechanism producing the missing values is systematic (not random), excluding cases with missing values may significantly change the characteristics of the gross sample.

The rule used to determine the ÉPE-AD's language of administration exemplifies this type of mechanism (detailed in Appendix D). For the French versions of all scales (with the exception the Communication scale), this rule generates missing values for children with the weakest French-language skills since they were tested using the tool's English version. This attrition reduces the range of the variable (and its variability), thereby potentially mitigating the program effects that depend on French proficiency. All similar mechanisms may similarly bias the results. The analysis in Section 7.2.2 elucidates the nature of the mechanism that resulted in the missing data, while Section 7.2.3 describes the imputation strategies used to minimize the effect of the missing values in the impact analyses.

³² This type of partially missing value was not observed for the French version of the ÉPE-AD scales. Since all the analyses reported in this document are based on the French version of the ÉPE-AD, the issue of partially missing values is no longer a factor with respect to these outcome measurements.

³³ The classic example used to illustrate this point is when people are asked about family income.

³⁴ This question was asked in the baseline survey and in parent follow-up surveys administered during the fifth and seventh evaluation period.

³⁵ It must also be assumed that the measured variable is unlikely to change systematically over time or to be affected by the treatment. We consider family income an excellent example of a variable that can evolve over time, but does not really show a systematic relationship with time (over the period of one to two years in question) or with the participants' assignment to groups in the study.

7.2.2. Pattern of Missing Data

Missing values are like any other outcome in that it is possible to model the process that generated them. The conclusions drawn from this modeling exercise determine what steps are taken to preserve the internal validity and external validity of the study. If the process that led to the missing data is random, validity is not threatened. However, if the process is not random, steps must then be taken to avoid introducing bias by excluding cases that have been self-selected. Below we describe the modeling strategy used to elucidate the pattern of missing data in the project database.

According to Little and Rubin (1987), there are three types of unplanned missing values: a) missing completely at random (MCAR); b) missing at random (MAR); and c) not missing at random (non-ignorable MNAR). The most desirable scenario is a situation where a small number of missing values is distributed randomly throughout the data (i.e., MCAR). Conversely, the worst scenario is having a large number of missing data distributed non-randomly. In the first case, the problem of missing values can be solved by applying a “listwise deletion” without risk of biasing the results of an analysis (i.e., removal of cases with missing values). This solution is not advisable for the second scenario. Using this method to deal with missing values that are distributed non-randomly may bias estimates of the treatment’s effects. This bias can be avoided by: a) excluding any irregular variables from the analysis; b) applying a data imputation strategy; or c) accepting bias in the analysis and considering its nature when interpreting results (e.g., missing values were more common in population X, therefore the results based on cases with complete data may not apply to population X).

The first step in this decision process is to determine the prevalence of missing values in the database (i.e., the percentage of missing values across all variables in the database) and the pattern of these missing values. The prevalence of the missing values is simply assessed by means of basic descriptive analyses. However, a more in-depth analysis is required to determine the pattern of the missing data.

The most vital question to answer is whether the missing values are predictable or not. As we have already seen, unpredictable missing values or those missing “at random” are classified as MCAR. One necessary condition for demonstrating MCAR is to show that the relationship between the occurrence of the missing values and the observed values of variables in the database is statistically null. This property can be verified by conducting independent tests of association for each variable in the database. For continuous variables, one may also use Little’s MCAR test. If the MCAR test or another association test reveals that values are systematically missing, then the working hypothesis of an MCAR distribution must be rejected in favour of a less restrictive assumption: the distribution of missing values is actually MAR (i.e., missing at random).

For a pattern of missing data to be considered MAR, two conditions must be respected: a) the missing values are non-randomly distributed and b) it must be possible to predict the value of the missing value. The first condition is met when one or more variables in the database allows us to predict the incidence of missing values; the second condition is met when strong predictors of the measured variable allow us to accurately predict the value of the missing observation. When the first condition is met but not the second, there is an MNAR-type distribution of missing data. In this case, we have no other option but to describe the pattern of missing data in enough detail to properly understand the limitations with respect to the study’s validity.

The issues that are at stake in the assessment of missing values have now been reviewed. In the following section, we report the results of the missing-values analysis conducted in preparation for the impact analysis of the Readiness to Learn project. First we report the results of a quantitative analysis of participant attrition. We then present an analysis of complete missing values due to a failure to assess a child or administer a survey. We end by detailing the result of the missing values analysis attributable to the ÉPE-AD's language of administration. For each analysis, we describe the missing values from two angles: in function of their breakdown by treatment group and in function of their relationship to other covariates in the database.

Values Missing Due to Attrition

The attrition rate is low for both cohorts of participants. In the first cohort, 14 families withdrew from the study by the end of the second year. Specifically, 10 families left the study by the end of the first year, for an attrition rate of 3.9% of the 254 families in the sample recruited in the four communities retained for the impact analyses. The second year, the attrition rate rose to 5.5% of the sample for the four communities. In the second cohort, only four families withdrew from the study in the first year and none in the second year, for an attrition rate of 3.9% of the 102 families in the enrolment sample. In total, both cohorts combined constituted an enrolment sample of 356 children with a 95% retention rate for the first two years of the study.

The attrition frequency was too low to formally test whether participation retention was tied to treatment group assignment in the study. However, a descriptive analysis of frequency does not show a systematic relationship. In total, six families in the Program Daycare group, two families in the Comparison Daycare group and five families in the Informal Care group left the study. Moreover, withdrawal from the study is not significantly associated with any ÉPE-AD scale in the first assessment. The strongest correlation observed is with expressive vocabulary, $r = 0.10$. Note that these analyses by treatment group and by ÉPE-AD scale do not include four participants who left the study before the first assessment.

The attrition rate for the Readiness to Learn project compares favourably to those of other assessment studies (e.g., Rogers, Fernandez, Thurber, and Smitley, 2004). At less than five percent, it is a negligible source of missing data (Tabachnick and Fidell, 2006, p. 63). In addition, the attrition rate does not appear to be linked to treatment group assignment. In conclusion, there is no risk of attrition biasing the estimate of the tested program's effect, either with respect to the generalization of results or the comparability of the groups in the study.

Missing Values Due to Complete Missing (Assessment or Survey)

Table 7.1 shows a low percentage of missing data for each assessment or survey administered in the first year of the study. The incidence of missing data is too low to test its association with treatment group assignment. In addition, the percentage of missing data observed for each treatment group is comparable.

Cases for which one or more parent survey is missing tend to do less well at the initial assessment. A negative correlation is observed between number of missing parent surveys and scores obtained on the Communication ($r = -0.21$), Self-Awareness ($r = -0.13$), Cognition ($r = -0.11$), Physical ($r = -0.11$), Expressive Vocabulary ($r = -0.14$) and Receptive Vocabulary ($r = -0.13$) domains. Similarly, a negative correlation is found between the probability of missing a survey and the score obtained on the Language of Literacy Activities scale ($r = -0.19$) whereby

those who used English most were most likely to having a missing survey. Likewise, cases for which one or more assessments are missing tend to do less well at the initial assessment. Negative correlations are observed between number of missing child assessments and scores obtained in the Communication ($r = -0.16$), Expressive Vocabulary ($r = -0.12$) and Receptive Vocabulary ($r = -0.11$) domains. Yet again, Language of Literacy Activities is also negatively associated with the presence of missing data ($r = -0.20$). The missing values must therefore be imputed to offset the bias to the generalization of results that may result from excluding cases with missing values.

In short, missing values are infrequent and are distributed equally among groups; their real values can also be estimated accurately given the many predictors and the repeated administration of several measurement instruments. Despite the negligible quantity of missing values, the fact that some of the study's outcomes are negatively and significantly associated with the presence of missing values means that simply excluding these cases would artificially increase the sample's competence, which would limit the generalization of results. This is particularly true for this study since program impact is assumed to vary based on French-language skills. The problem posed by the missing values must be managed using a more sophisticated strategy: replacing the missing values by imputation (see Section 7.2.3.).

Table 7.1: Percentage of Cases Missing at Least One Survey or One Assessment in the First Year of the Study (Four Assessment Periods)

	Entire Sample (%)	Program Daycare Group (%)	Comparison Daycare Group (%)	Informal Care Group (%)
Missing at least one survey	4.68	4.90	2.26	7.48
Missing at least one assessment	4.09	2.94	5.26	3.74

Missing Values Due to Language of Administration (Assessment Only)

The incidence of missing values due to English-language administration of the ÉPE-AD (the EYE-DA) is reported for the entire sample and by treatment group in Table 7.2. Based on the χ^2 tests, membership in a treatment group is associated with the rate of missing assessments only at the fourth assessment, for which a higher percentage of English-language administration is found for the Informal Care group. Overall, the results do not suggest that the language of administration may bias the planned comparisons. Nonetheless, it is still necessary to impute the missing values because a great deal of data is missing (over 10% of the observations) systematically (i.e., related to language proficiency). Wherever possible, the impact analyses should represent the full range of language proficiency found in the population. Section 7.2.3 details the success of the imputation strategy used to replace these missing values.

Table 7.2: Children Assessed in English during the First Five Assessments by Group

	Intra-Group				χ^2 test (2)
	Entire Sample	Program Daycare Group	Comparison Daycare Group	Informal Care Group	
	Frequency (%)	Frequency (%)	Frequency (%)	Frequency (%)	
Baseline	43 (12.8)	17 (17.0)	15 (11.5)	11 (10.4)	2.32
4 months	46 (13.5)	14 (13.9)	14 (10.8)	18 (16.4)	1.61
8 months	40 (11.8)	12 (12.5)	11 (8.5)	17 (15.2)	2.66
12 months	26 (7.7)	6 (6.3)	5 (3.9)	15 (13.3)	7.84**
16 months	30 (12.6)	9 (14.1)	11 (9.8)	10 (16.1)	1.61

Note: N = 342. Significance levels set at *** < 1%; ** 5%; * < 10%.

Missing Values Due to Partial Missing Data (Survey Only)

An analysis of the missing values due to partially missing data concerns first the confounding variables to be included in the regression analyses (for a complete list, see Section 7.3). For some of these variables, there is enough redundancy in the database (e.g., the question on family income is asked three times in the first three years) to begin by performing a simple imputation of the missing values.³⁶ After this first imputation, the list of covariates is subjected to a formal missing values analysis (this section) and to an imputation (see Section 7.2.3).

All the covariates in the impact analyses have less than 1% of data missing, except Vitality with a rate of 14.9%. Language Spoken by Mother to Child, Language Spoken by Father to Child and Authoritarian Parenting each have only 0.9% of data missing. Since Little's MCAR test is not significant, $\chi^2(46) = 51.87$, $p = 0.26$, we retain the hypothesis that the missing values for these covariates are distributed randomly. Concerning Vitality, the result of the preliminary analyses reported in Table B.4 indicates that the latter is associated with the treatment groups and must therefore be controlled for in the impact analyses. Since the rate of missing values associated with this variable is greater than 5% (Tabachnick and Fidell, 2006, p. 63), an imputation is performed to include cases with missing data in the impact analyses.³⁷

³⁶ When a given question or scale is administered several times, the various instances can be combined in order to create a single variable for the impact analyses that contains the greatest amount of information and the fewest missing values. In some cases (e.g., income), a difference in measurement scale required a regression imputation to bridge the two measurement instances for this variable, thereby creating a single variable that contains all the information available in the database on the measured concept. See Section 4.2.3 of the *First Cohort Findings Report* (Thompson et al., 2014) for a detailed description of the strategy for this type of imputation. The exact figures are not reported for the combined analysis of both cohorts since they add little value to the interpretation of results. Note that in contrast to the *First Cohort Findings Report*, the mother's characteristics offset the missing data for the father's characteristics for single-parent families.

³⁷ In the *First Cohort Findings Report* (Thompson et al., 2014), imputation was avoided by eliminating Vitality from the list of covariates. This is not an option in the present analysis because Vitality emerges as a significant confounding variable for this sample.

7.2.3. Data Imputation Strategy

The strategy used to treat missing values depends on the variable being treated. In all cases, we rely on the use of repeated measures in the study. As mentioned earlier, by repeatedly measuring the same variables (or very similar variables) with the same people throughout the study we are able to accurately estimate the value that would have been obtained had the variable been observed. Separate sections below describe the imputation strategy for assessment measurements (i.e., child outcomes) and for survey measurements (i.e., covariates and parent outcomes).

Imputation of Child Outcomes

In an earlier report (*First Cohort Findings Report*, Thompson et al., 2014), we described an imputation strategy for ÉPE-AD scales that drew on the study's longitudinal nature (see Section 4.2.3 of the *First Cohort Findings Report*). This strategy led to an imputed database for which impact analyses were conducted in the *First Cohort Findings Report*. The same strategy is used in this report.

Impact analyses for both cohorts combined include the data from the first cohort used in the impact analyses of the *First Cohort Findings Report*. An integral analysis of the data from both cohorts requires the creation of an equivalent database of ÉPE-AD scores for the second cohort. Accordingly, the same imputation strategy must be used for the data of the second cohort.

This report details only the imputation of data for the second cohort since this information was already reported for the first cohort (see Section 4.2.3 of the *First Cohort Findings Report*, 2014). What's more, only the imputations for the first five assessments are reported since these are most affected by the main source of missing values: English-language administration of the ÉPE-AD.

The algorithm for imputing the scores of the missing scales uses participant information drawn only from the French version of the ÉPE-AD.³⁸ First, we estimated a “mean” gain score for each child–scale combination. A child's mean gain score represents the mean change observed for the child from one assessment to the next for a given scale.³⁹ In executing this strategy, the observed values serve as anchor points for the estimation process. For example, the missing values from the second assessment were imputed by the sum of the score obtained at the first assessment and the mean gain score (imputed score = previous score + gain score). If values were missing from the first assessment, the process then estimated “backward” by subtracting the mean gain score from the score for the third assessment (imputed score = subsequent score – mean gain score). This strategy is limited by the fact that it requires at least two valid observations (i.e., two assessments) to complete a child's series of observations.

³⁸ Use of the participant's information means that the participant's unique attributes are correctly represented by the imputation strategy. As such, the strategy does not exaggerate the congruence between the scores of a given child and the rest of the group. However, intra-individual variability may be underestimated. The magnitude of the potential bias is determined by the weighting of the imputed values.

³⁹ Mean gain scores were calculated based on the average difference expected between two consecutive assessment periods. If the child was tested twice in the first and fourth assessment periods and scored respectively 4 and 14, the mean change score would be the difference observed between the assessments ($14 - 4 = 10$), divided by the number of intervals or steps between the assessments, that is, in the case at hand, three (average change score: $10/3 = 3.33$). All differences between the assessments contributed to the calculation, provided they are within three steps of one another. In addition, no imputation was performed if both the scores observed for the child for a given scale were from the first and the fifth assessment (i.e., four steps).

Imputation results are reported by treatment group for the second cohort in Table 7.3. We detail the Communication scale, which was administered in French to every child, separately from the other ÉPE-AD scales. Note that there are very few missing values for the Communication scale, especially in comparison with the rate observed for other scales. In any case, the imputation algorithm was able to estimate the missing value in most cases,⁴⁰ which allows us to retain in the impact analyses several children less proficient in French. The accuracy of this imputation method is proven by the result of an analysis intended to validate the imputation strategy for missing data reported in Appendix E.

Table 7.3: Distribution of Missing Values and ÉPE-AD Imputations by Treatment Group (Second Cohort)

	<i>ÉPE-AD Assessment Periods</i>				
	1 st	2 nd	3 rd	4 th	5 th
Communication Scale	# imputed / n (non-imputed)	# imputed / n (non-imputed)	# imputed / n (non-imputed)	# imputed / n (non-imputed)	# imputed / n (non-imputed)
Program Daycare group	0/31	0/31 (1)	0/31 (1)	1/31 (1)	0/31 (1)
Comparison Daycare group	0/37	0/37	0/37	0/37	0/37
Informal Care group	0/33	0/33	1/33	2/33	2/33
Total	0/101	0/101	1/101 (1)	3/101 (1)	2/101 (1)
Total (% imputed)	0.0%	0.0%	1.0% / 1.0%	3.0% / 1.0%	2.0% / 1.0%
Other ÉPE-AD Scales	# imputed / n (non-imputed)	# imputed / n (non-imputed)	# imputed / n (non-imputed)	# imputed / n (non-imputed)	# imputed / n (non-imputed)
Program Daycare group	5/31 (2)	2/31 (3)	1/31 (3)	1/31 (3)	0/31 (1)
Comparison Daycare group	4/37 (1)	3/37 (1)	2/37 (1)	0/37 (1)	0/37
Informal Care group	3/33 (1)	5/33 (2)	3/33 (2)	3/33 (2)	3/33 (2)
Total	12/101 (4)	10/101 (6)	9/101 (6)	4/101 (6)	3/101 (3)
Total (% imputed / % missing non-imputed)	10.9% / 4.0%	9.9% / 5.9%	8.9% / 5.9%	4.0% / 5.9%	3.0% / 3.0%

Note: The exact figures for the “Other ÉPE-AD Scales” category are from the Self-Awareness scale. The results reflect approximately what is observed for the other scales (with a difference of 1–2).

Imputation of Survey Data

Only four of the covariates retained for the impact analyses have missing data: Vitality, Language Spoken by Mother to Child, Language Spoken by Father to Child, and Authoritarian Parenting. The value for these missing values is estimated through multiple regression, a method that does not artificially reduce variability.⁴¹ The regression model is created using continuous variables from the surveys. This analysis includes all the non-redundant covariates available in the database, not just the reduced list of variables retained for the impact analyses. However, the outcomes of the ÉPE-AD were not used for that purpose.

⁴⁰ Other imputation strategies are based on the hypothesis that the parameters of the gross sample allow for a satisfactory estimate of an individual’s missing value. Since the majority of missing values are for atypical children (i.e., who were unable to complete the French-language test on at least one occasion), this hypothesis would be invalidated in principle. The strategy for these impact analyses uses only intra-individual information.

⁴¹ The residuals for a randomly sampled case are added to the value estimated by the regression model in order to imitate the “random” variability typical of a real observation.

7.3. CONFOUNDING VARIABLES

This section summarizes the relevance of the variables measured as part of the Readiness to Learn project impact analyses. This exercise involves listing the variables to be retained as covariates in the impact analyses. The selection criteria can be summarized as follows: a variable is deemed useful as a covariate if it is significantly associated with the outcomes and if it varies significantly by treatment group. The importance of these two criteria is easy to understand. To affect the results (e.g., increase statistical power, correct a bias in group composition), a variable must be significantly associated with the outcomes. For a variable to bias the estimate of program effect, the variable must be significantly associated with the treatment groups. If one of these conditions is missing, controlling for the variable or not in the impact analyses will not dramatically change the estimate of program impact. In particular, controlling for biases related to group composition is essential to maximizing the internal validity of the comparisons of interest in a quasi-experimental study.⁴²

Accordingly, preference has been given to variables that are associated both with the study's outcomes and with membership in treatment groups. Note that redundant variables and variables measured only within the Program Daycare group (e.g., those from the post-program survey) were excluded from these preliminary analyses. However, we examined certain variables specific to type of daycare (e.g., quality indices, dosage).

We began by checking a variable's association with treatment conditions and with child outcomes for the baseline period in a series of preliminary analyses reported in Section 7.3.1. The variables significantly associated with treatment groups were then examined more closely. The goal of this additional analysis was to identify a subset of variables that predict child development from the first assessment (see Section 7.3.2). These variables will be controlled for as part of the impact analyses.

7.3.1. Associations with Treatment Groups and Child Outcomes

The following strategy was used to identify the variables to be included as covariates in the regression analyses.

- 1) We verified whether the relationship between the variable and treatment group membership was significant at the baseline. The choice of test depended on the type of variable examined: the F-test was used for continuous variables, the Chi-square test for categorical variables.
- 2) We verified whether there was a significant relationship between the variable and at least one outcome observed in the baseline.
- 3) Variables that met both the above criteria are included in the specification for the regression models to correct the potential bias they represent. They were also the object of an additional analysis (see Section 7.3.2). The results of these preliminary

⁴² Note that a measured variable can only correct for differences observed between treatment groups. Because the measured variable is imperfect (it contains the measurement error), there may well be an undetected real difference between the groups. A difference that is real but undetected cannot be corrected by including the measured variable in a regression analysis. This is an example of an unobserved difference. Unobserved differences can be cancelled only through random assignment to the treatment groups. However, as discussed in Chapter 8, the static effect of unobserved differences on resulting variables may be cancelled using the DiD estimator.

analyses are reported in Appendix F. The variables presented in bold are those retained as covariates. For more details, see the appendix.

The covariates retained for the impact analyses are: Community, Cohort, Younger Siblings, Household Size, Frequency of Literacy Activities (baseline), Language of Literacy Activities (baseline), Language Spoken by Mother to Child, Language Spoken by Father to Child, Continuum of French Spoken by Child, Child's Age (in months), Vitality, and Authoritarian Parenting. In the second year, another covariate was added to this list: a child's School Enrolment (full-time, part-time, not enrolled).⁴³

In short, a control process for the observed variables and their associations with treatment groups and child outcomes is rigorously applied. The observed variables that were not retained for the impact analyses are distributed similarly across the treatment groups and therefore do not significantly contribute to the estimate of the tested program's impact (i.e., the groups are matched for these characteristics). The variables that were retained are used to perform a "statistical" matching as part of the impact analyses. There remains to verify one potential source of bias in group composition, that is, time constancy for the effect of confounding variables. This is the subject of the next section.

7.3.2. Unstable Effect of Covariates over Time

In the previous section, we considered only the effect of observed variables on child outcomes measured at the baseline. It remains to be verified whether the size of these effects is constant for all assessments. Effects that vary significantly from one assessment to the next must be dealt with in a special way in the specification of regression analyses.⁴⁴ The analyses reported in this section concern only the first year of the study and only the variables identified in Appendix F that were significantly associated with treatment conditions.⁴⁵

Only the effect of Household Size is constant over time for all outcome variables (see Table F.8 in Appendix F). For example, children's developmental trajectories vary significantly by community when it comes to Communication, Receptive Vocabulary and Frequency of Literacy Activities. Parameters representing this variability will therefore be included in the impact analyses to correct its potential bias to the estimate of program effects. Equivalent parameters are also added for the other variables identified in Tables F1 to F8, found in Appendix F, including Household Size, since this variable approaches the minimum significance level for Frequency of Literacy Activities. We hypothesize that, after statistical matching of the treatment groups, these would develop at the same rate in the absence of the tested program.

⁴³ The language of instruction at school (French versus French immersion or English) does not vary systematically by treatment group.

⁴⁴ Parameters are added to the specification in order to correctly represent changes in the size of the covariate's effect over time. In fact, Abadie (2005) comments on the usefulness of this technique in correcting the potential bias resulting from observed variables with an unstable effect. As with all estimates based on statistical correction, the impact analyses are validated to the extent that the core assumptions are valid (e.g., the regression model is properly specified).

⁴⁵ The child's "cohort" was also considered.

7.4. REPRESENTATIVITY OF THE SAMPLE: READINESS TO LEARN PROJECT VS. SVOLM (SURVEY OF THE VITALITY OF OFFICIAL-LANGUAGE MINORITIES)

The last section mainly concerned group comparability for impact analysis purposes, a criterion affecting the study's internal validity. This section now focuses on the study's external validity. This topic was addressed earlier in Chapter 5 of the *Readiness to Learn in Minority Francophone Communities: Reference Report* (Legault, Mák, Verstraete, & Bérubé, 2014; short title: *Reference Report*), in which a systematic series of comparisons was carried out to examine the sample of the Readiness to Learn project (first cohort only) and that of the Survey on the Vitality of Official-Language Minorities (SVOLM). The same series of analyses was repeated in Chapter 4 of the *First Cohort Findings Report* for a smaller sample: the sample of children from the four communities retained for the impact analyses of the first cohort. This section is an update of this second analysis, this time taking into account the children in both cohorts.

7.4.1. Issue

The SVOLM sample serves as a benchmark for the sample of the Readiness to Learn project. The purpose of the analyses in question is simply to answer the following question: If the Daycare Program and the Family Literacy Workshops were extended to the entire Francophone minority in the project's participating communities, would the observed effects be similar to those of the Readiness to Learn project? In order to answer this question, it is essential that both samples be drawn from the same Francophone minority population. However, this is not the case.

The interpretation of results must be nuanced by the fact that the SVOLM and the Readiness to Learn project used different sampling strategies; their populations are different, thereby limiting the samples' comparability in several regards. These differences stem in part from how the two projects define the minority Francophone population. As explained in the *Reference Report*, the SVOLM's sampling process is less restrictive. There may be substantial demographic differences between the two samples due solely to a difference in the studies' definitions of the minority Francophone population.

Next, the sampling of the Readiness to Learn project and the SVOLM differ in terms of demographic data. To ensure that a large enough sample was extracted from the SVOLM database, the data for children ages three to five were used in the comparative analyses. This contrasts with the average age of three observed for the sample in the study when the baseline survey was administered. Finally, the project participants were selected in a non-probabilistic manner based on daycare attendance, which means that the sample's geographic distribution is located within given communities. In contrast, the distribution of the SVOLM sample is more geographically diffuse because the survey in question uses random sampling.

We conclude by specifying that the analyses reported in the *Reference Report* and in the *First Cohort Findings Report* evaluating the representativity of the study sample are based on different samples, the first including the communities of Edmonton and Saint John, the second excluding them. Certain differences in results for the two series of analyses are noted in the *First Cohort Findings Report*. It remains to be seen whether the same pattern of results is obtained when both cohorts of the Readiness to Learn project sample are included. We expect that the conclusions of

these new analyses will be similar to those in the *First Cohort Findings Report* because adding the second-cohort children (a much smaller sample than first-cohort children) should not considerably disturb sample composition.

Note that wherever possible, we used the available data to re-evaluate SVOLM statistics based on the reduced sample of the four communities so as to maximize the validity of the comparison with the Readiness to Learn project sample used in the impact analyses.⁴⁶ When this strategy was impossible for practical reasons, we made comparisons with the SVOLM sample gleaned from the six geographic regions (*Reference Report*). These are acknowledged in the body of the text.

7.4.2. Immigrant Status and Linguistic Profile

Sampling procedures for the Readiness to Learn project and the SVOLM were carefully detailed and compared in the *Reference Report*. The following quotation summarizes the conclusions of that analysis:

“According to Forgues and Landry (2006), a Francophone population (such as the one used in the Readiness to Learn project) that is defined using the ‘ayant droit’ criterion would result in a much more restrictive pool whereas a Francophone population (such as the one used in the SVOLM) that is defined using several criteria (e.g., mother tongue, knowledge of official languages and languages spoken at home) would result in a greater number of eligible individuals.”

Two predictions were made in light of this analysis: a) the SVOLM sample likely includes a higher proportion of immigrants than the Readiness to Learn project and b) relatively fewer children are likely to report French as their mother tongue in the SVOLM. The next sections present the analysis results for information collected on immigrant status. Comparisons based on the mother tongue of the children and of the parents are also reported.⁴⁷ The pattern of results reported here for the analysis by mother tongue is equivalent to that reported in the *Reference Report* and the *First Cohort Findings Report*.

Respondents Born in Canada

The immigrant status of respondents in the Readiness to Learn project and SVOLM samples (four communities) is reported in Table 7.4. The first row reports the frequency of respondents who say they were born in Canada and the second, that of people born outside Canada. As anticipated, an examination of the distribution of responses for both surveys clearly indicates that the Readiness to Learn project has a higher proportion of respondents who were born in Canada. About 92% of the study sample members were born in Canada, while only 75.7% of respondents in the SVOLM sample were born here. This apparent difference was confirmed by applying a statistical Chi-square test that proved to be significant [$\chi^2(1, N = 1\,120) = 36.90, p < 0.001$].

⁴⁶ SRDC currently has access to frequency data by community, which allows for calculating appropriate estimates for the gross sample including the four communities. However, this calculation was not possible for certain variables where the analysis by community resulted in sample sizes too small to be extracted from Statistics Canada data. The six communities in the SVOLM sample nevertheless represent a worthwhile comparative group for the purposes of establishing the external validity of the Readiness to Learn Project sample for the four communities.

⁴⁷ FOLS (first official language spoken) was not used here to compare the samples because these data were collected in the SVOLM in such a manner as to invalidate all comparisons (see the *Reference Report*).

Table 7.4: Comparison of the Readiness to Learn project and the SVOLM

Immigrant Status	Readiness to Learn Project	SVOLM	Significant Differences Between the Two Samples?
	N (%)	N (%)	Chi-square
Born in Canada	302 (91.5)	598 (75.7)	Yes***
Born outside Canada	28 (8.4)	192 (24.3)	

Note: The SVOLM sample is from four regions. The project sample is based only on families that filled out the eighth survey. Significance levels set at *** $p < 0.1\%$; ** $p < 1\%$; * $p < 5\%$

First Language Learned and Still Understood — Children

Data on mother tongue for children in the Readiness to Learn project were gathered using the consent form filled out by parents. For the SVOLM, a child's mother tongue is deduced from the following question: (Statistics Canada, 2006, p. 35): "What is the language that [child's name] first learned at home in childhood and still understands?"

Table 7.5 shows that the sample for the Readiness to Learn project is more Francophone than that of the SVOLM. In fact, a higher percentage of children in the project report French only as their mother tongue (first row of the table). The percentage of children whose mother tongue is English only or English and another language is greater in the SVOLM sample (third row of the table). The representation of bilingual children in the two samples is practically identical (second row of the table).

Application of the statistical Chi-square test confirms that the distribution of the project's children (four communities) across the mother-tongue categories is not representative of the Francophone minority population in the six geographic regions based on SVOLM data [$\chi^2(2, N = 1\ 103) = 103.73, p < 0.001$]. These results are not surprising, given the above-mentioned differences in the sampling strategies of the two research projects.

Table 7.5: Comparison of the Readiness to Learn project and the SVOLM —Children Categorized by Mother Tongue

Mother Tongue	Readiness to Learn project	SVOLM	Significant Differences Between the Two Samples?
	N (%)	N (%)	Chi-square
French only	247 (72.2)	306 (40.2)	Yes***
English and French equally OR French and another language	33 (9.7)	89 (11.7)	
English only OR English and another language OR other language(s)	62 (18.1)	366 (48.1)	

Note: The SVOLM sample is from six regions. Significance levels set at *** $p < 0.1\%$; ** $p < 1\%$; * $p < 5\%$.

First Language Learned and Still Understood — Mothers

Table 7.6 indicates that most mothers in the project (67.0%) and SVOLM (58.3%) samples report French as their only mother tongue (first row of the table), although the proportion in this regard is slightly higher in the sample of the Readiness to Learn project. Moreover, a smaller

proportion of mothers in the project fall under the “English only OR English and another language OR other language(s)” category (23.1% in the third row). A Chi-square test suggests that the distribution of the project’s mothers (four communities) across the mother-tongue categories is not representative of the Francophone minority population in the six geographic regions based on SVOLM data [$X^2(2, N = 1\,131) = 9.68, p < 0.01$].

Table 7.6: Comparison of Mothers in the Readiness to Learn project and the SVOLM — Mothers Categorized by Mother Tongue

Mother Tongue	Readiness to Learn project	SVOLM	Significant Differences Between the Two Samples?
	N (%)	N (%)	Chi-square
French only	229 (67.0)	460 (58.3)	Yes**
English and French equally OR French and another language	34 (9.9)	75 (9.5)	
English only OR English and another language OR other language(s)	79 (23.1)	254 (32.2)	

Note: The SVOLM sample is from six regions. Significance levels set at *** $p < 0.1\%$; ** $p < 1\%$; * $p < 5\%$.

First Language Learned and Still Understood — Fathers

Table 7.7 shows the linguistic profile of fathers in the Readiness to Learn project and in the SVOLM based on their mother tongue. At first glance the pattern seems similar to that observed among mothers. The main difference is the substantial number of fathers in the SVOLM who reported “English only OR English and another language OR other language(s)” as their mother tongue. The proportion of fathers in the SVOLM at either end of the distribution in Table 7.7 is virtually identical (46.7% versus 47.7%). However, fathers in the study sample are more massively represented in the “French only” category (56.1% in the first row) than in the “English only OR English and another language OR other language(s)” category (35.5% in the third row). This last pattern resembles that observed in both samples for the mother’s mother tongue.

A Chi-square test confirms that the distribution of the project’s fathers (four communities) across the mother-tongue categories is not representative of the Francophone minority population in the six geographic regions based on SVOLM data [$X^2(2, N = 1\,116) = 15.04, p < 0.001$].

Table 7.7: Comparison of Fathers in the Readiness to Learn project and the SVOLM — Fathers Categorized by Mother Tongue

Mother Tongue	Readiness to Learn project	SVOLM	Significant Differences Between the Two Samples?
	N (%)	N (%)	Chi-square
French only	185 (56.1)	367 (46.7)	Yes***
English and French equally OR French and another language	28 (8.5)	44 (5.6)	
English only OR English and another language OR other language(s)	117 (35.4)	375 (47.7)	

Note: The SVOLM sample is from six regions. Significance levels set at *** $p < 0.1\%$; ** $p < 1\%$; * $p < 5\%$.

7.4.3. Sociodemographic Characteristics

In the *Reference Report* and the *First Cohort Findings Report*, the Readiness to Learn project sample was compared with the SVOLM sample based on child gender, family composition (the family's size and structure) and socioeconomic status (parents' education, family income). The reports conclude that the samples of the two surveys are comparable in terms of child gender (verified only for the *Reference Report*), family structure, family size and family income. However, the distribution of responses is found to vary between the samples for parents' education (father and mother) and number of siblings. In all cases, the general pattern of these results was reproduced in the analysis of both cohorts, the results of which are presented below.

Total Family Income

Table 7.8 shows that the families in the Readiness to Learn project and those in the SVOLM are distributed similarly among the income brackets considered here. In both cases, the modal and median category for both samples is \$60,000 or more per year. A Chi-square test confirms that the distribution of the project's parents (four communities) across the income brackets is statistically equivalent to that observed for the Francophone minority population in the six geographic regions based on SVOLM data [$\chi^2 (5, N = 1\ 131) = 5.58, p > 0.05$]. The results suggest that most children in both samples enjoy a good quality and good quantity of physical resources for their development.

Table 7.8: Comparison of the Readiness to Learn project and the SVOLM — Families by Income Bracket

Income Bracket	Readiness to Learn project	SVOLM	Significant Differences Between the Two Samples?
	N (%)	N (%)	Chi-square
\$10,000 or less	20 (5.8)	54 (6.9)	No
From \$20,000 to \$29,999	14 (4.1)	23 (2.9)	
From \$30,000 to \$39,999	19 (5.6)	64 (8.1)	
From \$40,000 to \$49,999	18 (5.3)	57 (7.2)	
From \$50,000 to \$59,999	46 (13.4)	95 (12.0)	
\$60,000 and over	225 (65.8)	496 (62.9)	

Note: The SVOLM sample is from six regions. Significance levels set at *** $p < 0.1\%$; ** $p < 1\%$; * $p < 5\%$.

Mothers' Level of Education

Table 7.9 shows three main points. First, the mothers in the Readiness to Learn project are, on average, more educated than mothers in the SVOLM sample. In fact, almost 80% of the project's mothers have a college diploma (DEC, or diploma of collegial studies, DCS) or a university degree versus about 70% of mothers in the SVOLM. This difference is mainly attributable to the low number of mothers with at least one diploma or one certificate of collegial studies in the SVOLM versus those of the project (second row). Second, there are as many mothers with a college diploma (38.9%) as there are mothers with a university degree (40.9%) in the project. Third, more mothers in the SVOLM went to university (42.7%) than did mothers in the project (40.9%), although this difference is negligible.

A Chi-square test confirms that the mothers' level of education in the Readiness to Learn project is not representative of the Francophone minority population in the four geographic regions based on SVOLM data [$X^2(2, N = 883) = 22.74, p < 0.01$].

Table 7.9: Comparison of Mothers' Level of Education in the Readiness to Learn project and in the SVOLM

Level of Education	Readiness to Learn project	SVOLM	Significant Differences Between the Two Samples?
	N (%)	N (%)	Chi-square
High school diploma or less OR a few post-secondary courses	69 (20.2)	172 (31.8)	Yes***
College diploma/certificate (e.g., trade school)	133 (38.9)	138 (25.5)	
University degree (bachelor's, master's or PhD)	140 (40.9)	231 (42.7)	

Note: The SVOLM sample is from four regions. Significance levels set at *** $p < 0.1\%$; ** $p < 1\%$; * $p < 5\%$.

Fathers' Level of Education

Table 7.10 indicates that more fathers attended university in the SVOLM (35.9%) than in the Readiness to Learn project (34.6%). However, the project's fathers are generally more educated than fathers in the SVOLM. In fact, almost two-thirds of them have a college diploma (DEC, or diploma of collegial studies, DCS) or a university degree, while 60% of fathers in the SVOLM have an equivalent level of education. Finally, application of the Chi-square test suggests that fathers' level of education in the Readiness to Learn project is not representative of the Francophone minority population in the four geographic regions based on SVOLM data [$X^2(2, N = 873) = 7.48, p < .05$].

Table 7.10: Comparison of Fathers' Level of Education in the Readiness to Learn project and the SVOLM

Level of Education	Readiness to Learn project	SVOLM	Significant Differences Between the Two Samples?
	N (%)	N (%)	Chi-square
High school diploma or less OR a few post-secondary courses	110 (33.1)	216 (39.9)	Yes**
College diploma/certificate (e.g., trade school)	107 (32.2)	131 (24.2)	
University degree (bachelor's, master's or PhD)	115 (34.7)	194 (35.9)	

Note: The SVOLM sample is from four regions. Significance levels set at *** $p < 0.1\%$; ** $p < 1\%$; * $p < 5\%$.

Family Size

According to Table 7.11, the modal and median family size is four for both samples (four communities). In both surveys, families of four represent about half of the sample. The remaining families are distributed fairly equally between families of three or less and families of five or more. A Chi-square test suggests that there is no significant difference in the distribution of family size for families participating in the Readiness to Learn project and families from the four geographic regions of the SVOLM [$X^2(2, N = 880) = 2.48, p > 0.05$].

Table 7.11: Comparison of Family Size¹ in the Readiness to Learn project and the SVOLM

Number of Family Members	Readiness to Learn project	SVOLM	Significant Differences Between the Two Samples?
	N (%)	N (%)	Chi-square
Three or less	78 (22.8)	143 (26.6)	No
Four	196 (57.3)	280 (52.0)	
Five or more	68 (19.9)	115 (21.4)	

Note: The SVOLM sample is from four regions. Significance levels set at *** $p < 0.1\%$; ** $p < 1\%$; * $p < 5\%$. ¹ The number of people in a family refers to the number of parents and children only.

Siblings

Table 7.12 indicates that the modal and median number of children per respondent (family) is two for the Readiness to Learn project and for the SVOLM. However, there are slightly more two-children families in the project (60.8%) than in the SVOLM (50.4%). In addition, there are more families with three or more children in the SVOLM (30.1%) than in the project (20.4%). Conversely, the number of families with a single child, about 20%, is roughly the same for both samples. A Chi-square test confirms that the distribution of the number of children per respondent in the Readiness to Learn project is not representative of the Francophone minority population in the four geographic regions based on SVOLM data [$X^2(2, N = 1\,128) = 12.97, p < 0.01$].

Table 7.12: Comparison of the Readiness to Learn project and the SVOLM — Number of Children per Respondent

Number of Children	Readiness to Learn project	SVOLM	Significant Differences Between the Two Samples?
	N (%)	N (%)	Chi-square
One child	64 (18.7)	154 (19.6)	Yes**
Two children	208 (60.8)	396 (50.4)	
Three or more children	70 (20.5)	236 (30.0)	

Note: The SVOLM sample is from four regions. Significance levels set at *** $p < 0.1\%$; ** $p < 1\%$; * $p < 5\%$.

Family Structure

Families in the Readiness to Learn project had to be redefined as either single-parent families or two-parent families in order to compare family structure in the project sample with that in the SVOLM (see Table 7.13). Note that the two-parent category includes intact families and blended

families where both parents (or one parent and his or her spouse) live with the child. The single-parent category consists of families where only one parent lives in the home with the child.

Note that a child's father or mother may be either his or her biological or adoptive parent. Finally, same-sex couples were excluded from the analysis as were children raised by a person other than the child's biological or adoptive mother or father. A Chi-square test confirms that children's distribution in single-parent or two-parent homes in the project is representative of the Francophone minority population in the four geographic regions based on SVOLM data [$\chi^2(1, N = 1\,131) = 1.68, p > 0.05$].

Table 7.13: Comparison of the Readiness to Learn project and the SVOLM — Number of Single-Parent and Two-Parent Families

Family Structure	Readiness to Learn project	SVOLM	Significant Differences Between the Two Samples?
	N (%)	N (%)	Chi-square
Single-parent	29 (8.5)	87 (11.0)	No
Two-parent	313 (91.5)	702 (89.0)	

Note: The SVOLM sample is from the four regions. Significance levels set at *** $p < 0.1\%$; ** $p < 1\%$; * $p < 5\%$.

7.5. SUMMARY OF THE IMPLICATIONS FOR IMPACT ANALYSES

In this chapter, we covered the methodological issues that affect the internal and external validity of the impact analyses to follow. Technical matters related to quality control processes, missing values, imputation, confounding variables and external validity were examined.

The chapter begins by identifying data verification and quality control processes (Section 7.1). These processes were designed to minimize problems (e.g., measurement errors) caused by the data collection process and to identify and correct problems in the electronic databases. These procedures and the use of a mixed-method approach (or “converging operations”) in the research ensure the validity of the results from the impact analyses.

The chapter continues with an analysis of the missing values (Section 7.2), the result of which shows that, overall, the missing values are broken down equally across treatment groups and constitute a problem that is very limited in magnitude. The attrition rate is very low for the first two years of the project, while the response rate is very high. The analyses also indicate that performance in terms of child outcomes is not a predictor of participant attrition. Although the analyses show that participants who withdrew from the study have specific characteristics (e.g., they are less “Francophone,” see the *First Cohort Findings Report*), the generalization of results to the Francophone minority population is not affected by that fact. Statistical testing shows that other missing values associated with survey data are produced randomly. These were imputed to maximize numbers for the purpose of the impact analyses. However, treatment of the data missing not at random was problematic with all ÉPE-AD scales, except for Communication. These data are missing due to English-language administration of the ÉPE-AD rather than French-language administration. The adopted imputation strategy led us to impute certain missing data as MAR (i.e., missing not at random that can reliably be imputed based on other variables in the database), but not all (MNAR, i.e., missing not at random that cannot be imputed). As a result, the impact analyses based on these scales can only be generalized to a

population of minority Francophones whose children are likely to meet the criteria necessary to take the ÉPE-AD in French (i.e., children with a relatively high score on the French Language and Communication scale, Domain C, and the Awareness and Engagement in Francophone Culture scale, Domain E) or, more specifically, to a population of children who could have taken the ÉPE-AD in French twice or more in the course of the five assessments while they were approximately between the ages of three and four.

This chapter then reports the result of the confounding variables analysis (Section 7.3). The goal of the analysis is to identify variables that allow for: a) statistical adjustments to offset changes to group composition over time; and b) statistical adjustments to offset differences in developmental trajectory that would occur among the groups in the absence of treatment. Further to this analysis, a number of covariates were identified for inclusion in the impact analyses (see Appendix F for an exhaustive list of the variables examined and the detailed results). Introducing these covariates in the impact analyses will maximize the internal validity of this quasi-experimental study with non-equivalent control groups.

Finally, we examined the issue of the study's external validity. In earlier research we concluded that there were more Francophones in the sample of the Readiness to Learn project than in that of the SVOLM. This result is reproduced here with the sample of both the project's cohorts. In the *Reference Report*, we hypothesized that this apparent bias would stem from the population of children typically found in French-language daycares. To the extent that this argument is justified, all the results of this study would apply only to a population of Francophone children enrolled in daycare. First, this limit appears to be self-evident in the sense that this population of children would be affected by a daycare intervention. However, if the differences in mother tongue observed between the Readiness to Learn project and the SVOLM reflect a true lack of representativity - which is tantamount to saying that the project sample is not representative of the target population - then the results of the impact analyses in Chapters 9 and 10 of this report may in fact underestimate the magnitude of the program's true effect for a population characterized by greater linguistic diversity. The possibility that linguistic variables may emphasize or mitigate the tested program's effect can be examined by reassessing program effects separately for children who mainly speak French and those who speak it less (see Sections 9.2.5 and 9.3.5). The information resulting from these analyses may help determine whether the intervention would be more effective if it targeted given subpopulations, including that consisting of children who are mainly exposed to languages other than French in their family environment.

The fact that the samples of the Readiness to Learn project and the SVOLM are generally similar with regard to non-linguistic characteristics supports the argument that the project sample is representative of Francophones living in a minority environment. The only difference worth mentioning is the fact that parents in the Readiness to Learn project are slightly more likely to report a level of education beyond high school than parents in the SVOLM sample.

In short, a number of precautions were taken to ensure the internal validity of the program's estimated effects and to estimate its degree of external validity. Other verifications are discussed as they become relevant to the interpretation of results in the following sections of this report.

8. Approach to Analyses

This chapter deals with the logic that underpins the analyses described in Chapters 9 and 10. The econometric approach to the analyses factors in the data's longitudinal and multi-level structure and involves systematic verification of the robustness of results (see Section 8.1). Finally, the method used to identify the program's effect is suitable for a quasi-experimental study with non-equivalent groups (see Section 8.2).

This general approach was applied for the specification of several empirical data models. Many of them are presented in the chapters that follow, each representing, through its specification, a different way to conceive of families' and children's exposure to the tested program. Here, exposure to the tested program is designed as a continuum ranging in intensity from low to high (see Section 8.3). The strongest possible intensity of exposure would be received by a child who attends daycare full-time at a facility that implements the program exactly as it was designed and with the highest degree of quality. The additional models considered in this report are introduced in Sections 8.3 and 8.4. Finally, Section 8.5 explains how the results are organized in Chapters 9 and 10.

8.1. HIERARCHICAL LINEAR MODELING (HLM)

The data were analyzed using linear regression models based on the core assumption that each observation point or data point in the analysis was observed independently. This assumption is not respected when the sample units (e.g., daycares or children) contribute to multiple observations for a set of data. In this case, the observations made by a given sample unit are said to be nested or clustered. In a longitudinal design, observations are grouped by participant (i.e., that each participant contributes to several observations) and sometimes based on another type of analysis unit. The Readiness to Learn project in particular presents data that are nested by daycare and by participant. This multi-level structure, often analyzed using hierarchical linear modeling (HLM), must be reflected in the analyses to avoid overestimating the statistical significance of results (Hox, 2002; Moulton, 1990). Treatment of the study's longitudinal aspect is detailed later in Section 8.2, whereas treatment of the "daycare" effect will be addressed now.

To maximize the robustness of results, we use the Huber-White heterogeneity-consistent estimator (White, 1980) with a modification that makes it robust to clustering (Williams, 2000). According to Woodbridge (2002), the properties of this method are satisfactory for analyzing a database like that of the Readiness to Learn project (i.e., the observed ratio of the number of groups relative to the number of observations per group) when the number of groups determines the degrees of freedom for the significance tests. The accuracy of the reported effects is then robust to heterogeneity and to clustering by childcare environment.⁴⁸

⁴⁸ Since this method is sometimes criticized (Donald and Lang, 2007), the robustness of results was verified by repeating the analyses several times. At each iteration, the "daycare" effect was treated in a different way (e.g., unspecified; as a random effect; as a fixed effect). This verification shows that the method used does not dramatically affect the results and that this method is typically more conservative for this database. Our decision was to be conservative in our methodological choices, preferring "consistency" over "efficiency" (Hayes & Cai, 2007).

8.2. THE DIFFERENCES-IN-DIFFERENCES METHOD

A standard strategy for estimating program effects in econometrics literature is to use a differences-in-differences (DinD) estimator (Abadie, 2005; Bertrand, Duflos, and Mulliaianathan, 2004). The DinD estimator is a panel data technique that is appropriate for non-experimental repeated-measures research designs with both a pre-test “baseline” measurement and a comparison group. Readers who are less familiar with the econometric literature may be more familiar with the concept of an interaction or moderated effect, of which the DinD estimator is a special case. The term *DinD estimator* refers to an interaction involving two dummy variables, one representing the pre- and post-treatment test periods (Pretest vs. Posttest) and the other representing the two groups being compared (Intervention group vs. Comparison group).

This estimator can be generalized to more complex scenarios by specifying several interaction terms of this kind in the same regression model (e.g., to compare the pre-test to multiple post-test measures in the same regression model) or to compare multiple groups. The impact analyses described in this report use multiple DinD estimators to represent comparisons of the three groups in the study based on several post-intervention assessments. All three groups were included in the same analysis in order to maximize the stability of statistical testing, thereby increasing the chances of discovering the program’s true effects.

As its name suggests, the DinD estimator has two main components. The first is an estimate of the change (Δ) from the pre-test assessment to a post-test assessment for each group ($\Delta_{\text{intervention group}} = \text{post-test} - \text{pre-test}$; $\Delta_{\text{comparison group}} = \text{post-test} - \text{pre-test}$). Here, the pre-test measure is the standard against which all subsequent assessments are compared. In other words, the post-test measures are always compared to the pre-test measure when estimating the program’s effects. However, these change scores (i.e., differences) are not sufficient to isolate the treatment effect, since the change being estimated may have occurred as the result of natural developmental processes (e.g., maturation).

The second component of the DinD estimator eliminates this ambiguity by recording the difference between the change scores for the intervention group and those for the comparison group ($\text{DinD} = \Delta_{\text{intervention}} - \Delta_{\text{comparison}}$). Here, the change estimate provided by the comparison group is used to adjust the change estimates for the intervention group. The validity of the DinD estimator stems from the premise that once you eliminate the counterfactual change estimate from the comparison group, all that remains is the change resulting from the treatment effect. As with most statistics, the assumptions underlying this premise are likely to be violated in practice. This is particularly true for a non-experimental study conducted in the field, of which the Readiness to Learn project is an example.

The validity of the DinD estimator depends on at least two assumptions. With respect to change scores, it is assumed that group composition remains constant for all assessment periods subjected to comparison, which may not be the case if participants migrate from one group to the other during the study. As a result, changes in group are problematic for the Readiness to Learn project, which requires a control in the analyses. With respect to the comparison of change scores, it is assumed that in the absence of the tested program, the groups being compared would have evolved in the same way. Of course, there is no way to verify this assumption directly, but the initial differences between the groups in terms of the variables associated with the dependent measure raise the possibility of non-parallel slopes.

As with the *First Cohort Findings Report*, we minimized these threats to the validity of the DiD estimator by including, in our analyses, covariates that allowed for adjustments to reflect differences between the groups in their initial state. The static effect and dynamic effect of these covariates are represented in the empirical models respectively by the direct effect of these variables and their interactions with the longitudinal factor. The details concerning this selection process were presented earlier in Chapter 7 and in Appendix F. A more in-depth discussion of the measures taken to ensure valid results is presented in Section 8.4.

8.3. CONCEPTUALIZATION OF TREATMENT AND DOSAGE

The simplest model to capture a treatment effect involves two groups of participants: the “treated” group and the “untreated” group. This type of model is most valid when the distinction between a treated group and an untreated group is absolute, that is, the untreated group receives absolutely no treatment, while all members in the treated group receive the same treatment (e.g., the same dosage or “exposure”). For a study carried out in the field, such clean-cut distinctions are rather rare, and the Readiness to Learn project is no exception to this rule. The intensity of treatment received by participants varied in at least two regards: degree of treatment exposure (dosage) and quality of treatment. These two sources of variations in treatment intensity are detailed in the sections that follow.

8.3.1. Heterogeneous Exposure to Treatment

When they enrolled in the project, participants were part of one of three groups: the Program Daycare group, the Comparison Daycare group, or the Informal Care group. It may come as no surprise that group membership sometimes varied over time depending on the parents’ choice of childcare. For example, a relatively low percentage of parents (see Table 3.2) decided to change daycares between assessments. As such, a child who was part of the Program Daycare group for the first two assessments could wind up in the Comparison Daycare group or in the Informal Care group for subsequent assessments.

Even if participants remained in the same group, the time they spent at daycare may have varied over time. The results reported in Appendix F show that there is no difference in average exposure to the childcare environment for the two daycare groups. One practical effect of this equivalence is that this variable is not an obvious threat to the validity of our estimates of the tested program’s effects. Nevertheless, it is worth wondering whether the degree of exposure to a treatment condition is important and whether the program’s effects, if any, interact synergistically with the degree of exposure. In other words, to ask the question: Does spending an hour at a program daycare yield better results than spending an hour at a comparison daycare? We treated this question by adding variables for the degree of treatment exposure to the impact analyses.

8.3.2. Evaluating the Effect of Dosage/Exposure

Two strategies were adopted for managing the migration of participants over time. The first of these strategies is coarse, but its simplicity makes it easier to compare groups (i.e., treated vs. untreated). The second strategy more specifically defines the treatment by indicating the hours of exposure to a particular daycare setting (i.e., the average hours spent at daycare per week). This last strategy is somewhat more complex, but has the advantage of representing the degree of

exposure to the two daycare treatment conditions. These two techniques for representing the heterogeneity of treatment exposure are the object of two independent series of analyses, which are explained in greater detail below.

The first series of analyses is based on the coarse definition of treatment exposure, whereby dummy codes were used to represent group membership. For any given period of time, a participant coded as belonging to a certain treatment group was deemed to have received the entire treatment associated with that group. Group membership was allowed to vary over time, but the integrity of the treatment groups was maximized by requiring that participants be exposed to the new type of childcare for a minimum period of time before recognizing this change in the analyses. More specifically, the two following conditions were set. First, a child was considered to have changed groups only if the change occurred over one month before the assessment. Second, changes in daycare arrangements over the summer months were deemed valid only if maintained for several months into the school year. A participant who reported withdrawing from a program daycare and subsequently enrolling at a comparison daycare over the summer, for example, was considered a member of the Program Daycare group for the fall assessment. For analysis purposes, the change in group came into force only for the winter assessment, since by then the child had been exposed to the daycare environment for a few months during which the regular program was in place. The two standardized vocabulary measures are exceptions since they are administered each only once. The participant breakdown in force when the measurements were administered was used for the analyses. Regardless of the outcomes measured, we evaluated in all cases the potential bias in the estimate of the program effect that may result from the changes in group.

In the second series of analyses for first-year data, treatment was defined by using hours of exposure to a childcare environment, along with the characteristics of the daycare setting (i.e., the Program Daycare group or the Comparison Daycare group). Hours of exposure was defined as the average number of hours per week a child spent at daycare during the four-month period prior to a given assessment. The baseline period, for which only two months were used (September and October 2007 and 2008, respectively, for the first two cohorts), was an exception. Including this variable in the analyses enabled us to more specifically define exposure to our Daycare Program. By crossing the hours of exposure variable with our grouping variable (i.e., by specifying an interaction term), we were able to: a) estimate the average treatment effect associated with a given number of hours per week of exposure to daycare; and b) test whether the effect of the degree of exposure to daycare varies as a function of program type. The latter test is simply an extension of our basic research hypothesis, whereby given an equivalent degree of exposure, participants in the Program Daycare group will have an advantage in the three domains of school readiness. This more specific conceptualization of dosage enhances the statistical power of our analyses to detect program effects. It also gives us an opportunity to determine the degree of program exposure required to produce the desired effect.

For the follow-up analysis of second-year outcomes, we used the data on hours of exposure slightly differently. At the end of the first year, most children in the study (i.e., children living in the three Ontario communities) were enrolled in kindergarten on a part-time or full-time basis. That is why the data on daycare attendance gathered in the project's second year are difficult to interpret as predictors of outcomes since they do not adequately reflect the time spent in a quality childcare environment. Children who are enrolled in school on a full-time or part-time basis are exposed to a quality environment. Thus, a child who spends a few hours a week at a program

daycare may in fact spend many hours outside the home and experience a treatment of comparable intensity.

“Hours of exposure to daycare” is certainly not a meaningful variable in the project’s second year. It would be technically possible to conduct this type of analysis using only the community of Edmundston (i.e., no child in Edmundston was enrolled in school), but there would be too few children in such an analysis to provide useful information.⁴⁹ Instead, we treated the second-year assessments as follow-up tests of the effect of the hours of exposure to daycare in the first year. Nevertheless, the investment necessary to provide childcare services in the second year of the project was important in order to ensure that program effects were maintained for children not attending school full-time. Although the performance of these children did not undergo a targeted analysis, it did contribute to the program effect reported in Chapters 9 and 10. In other words, the persistent effects of the hours of treatment exposure were tested in this analysis.

To represent this aspect of children’s exposure to the childcare environment, we calculated a variable representing the average number of hours per week that a child spent either at a program daycare or at a comparison daycare during the first eight months of program delivery.⁵⁰ This variable was crossed with the treatment group in a regression model predicting child outcomes at the end of the second year. Otherwise stated, the treatment group for the first year was crossed with the daycare dosage for the first year in an attempt to predict child outcomes for the second year. If a year spent at a program daycare in the first year translates, on average, to better outcomes than a year spent at a comparison daycare, then one can expect the dosage to have a much more significant effect for the group attending a program daycare than for the group attending a comparison daycare in the second year, whether the program’s effect is persistent or delayed.

8.3.3. Evaluating the Effect of Daycare Program Fidelity/Quality

The mechanism by which the intervention was supposed to affect the developmental outcomes of children was the quality of the program and the fidelity of its implementation. Thus, treatment group membership can be seen as approximation proxy for quality and fidelity. The study’s internal validity is fundamentally based on the truth of this statement, hence the examination of the differentiation presented in Section 9.2.4. It follows that one condition required to observe a program effect on the targeted outcomes is that children enrolled at the daycares where the intervention was implemented (the Program Daycare group) have experiences that compare favourably with those of children in the comparison group (the Comparison Daycare group), who are also exposed to a childcare program. Likewise, within each group, the nature of the program offered from one daycare to the next must be as similar as possible, that is, it must be consistent at all daycares in the same group. In other words, the daycares must be grouped in a coherent manner. Both these conditions were verified by means of the qualitative analyses presented in Chapter 5. The results of formal quantitative tests on the differentiation of daycare programs are also reported in Chapter 9.

⁴⁹ If the study were repeated with 100 different samples, a real effect of 0.30 standard deviations would be detected for less than 50% of the repetitions, assuming a sample of 90, covariates that can explain 50% of the residual variance, and a level of significance of 0.05. Moreover, the stability of the estimated parameters would be doubtful in such an analysis.

⁵⁰ We excluded the data for the four months prior to the fourth assessment since they concerned variations attributable to the disturbances of summertime. The purest measurement of the average effect of treatment exposure is the information collected during the school year.

The use of treatment groups to estimate the magnitude of treatment effects is a useful simplification where the potentially continuous dimensions (e.g., fidelity and quality) are reduced to categories (i.e., the Program Daycare group and the Comparison Daycare group) for comparative purposes. This simplification is the most appropriate method when members of every group are very similar with regard to continuous dimensions. Of course, the nature of a daycare program always varies somewhat from daycare to daycare, which means that using treatment groups results in a loss of information (i.e., intra-group variability). In fact, the distribution of fidelity/quality scores for both groups may, in principle, overlap.

We verified whether this loss of information, resulting from the use of treatment groups, was significant. To do so, we conducted a series of analyses where the presence of an intermediary (the treatment group) was eliminated and replaced by continuous fidelity and quality indicators as child outcome predictors. This method optimizes the use of the available information and offers a better chance of identifying the program's true effects.

We also wished to determine whether certain effects observed in the treatment group were attributable to the program itself or to other characteristics of the program daycares. We tested this idea by estimating the effects of the treatment group after controlling for fidelity and quality. If the tested program is specifically responsible for the effects observed, then the adjusted effects of the treatment group resulting from this analysis should not be statistically significant.⁵¹ The logic of this analysis is based on a mediation test (Cohen, Cohen, West, and Aiken, 2003, p. 457) where it is assumed that membership in a treatment group has an indirect effect on outcomes through the quality or fidelity of the service.⁵² Any residual effect linked to the treatment group would necessarily be attributable to another factor (e.g., uncontrolled differences between the groups when the project begins, another source of bias, an aspect of the program not captured by fidelity or quality indicators). Although including the analyses in this report makes it substantially longer, their inclusion improves our understanding of the reported results and enhances our confidence in them.

In the analyses reported in Sections 9.2.4 and 9.3.4, each type of index has been operationalized at a global level and at a more detailed level. Two detailed estimates of fidelity were computed based on adherence to the elements specific to program structure and to program content, respectively. The global estimate of fidelity was computed simply by averaging these two indices. As regards quality per se, we calculated the indices representing the Structural Quality, Educative Quality, and Educator Sensitivity dimensions. Yet again, a general quality index was calculated by averaging these three quality indices. A fourth index captured the quality of literacy activities in the classroom (Reading Quality). This index was dealt with separately from the others due to its theoretical and empirical importance. A more detailed description of these indices is given in Section 9.1.2.

⁵¹ This statement must, however, be nuanced. Fidelity and quality may explain most of the program effect without completely eliminating it. In fact, the size of the coefficient representing the program effect may even increase in the presence of a suppression effect (Cohen, et al., 2003, pp. 457–458). The importance of fidelity and quality can nevertheless be established with certainty through a formal mediation test (Frazier, Tix, and Barron, 2004).

⁵² An appropriate mediation test usually involves a series of regression analyses (Baron & Kenny, 1986; Frazier, et al., 2004). One condition essential to demonstrating complete mediation is that the direct effect of variable X (in this case, the treatment) is eliminated when the M mediator is controlled for (in this case, quality/fidelity). Partial mediation is observed if the direct effect is reduced but not completely eliminated. Note that a mediation effect can exist without reducing the direct effect due to a phenomenon called “suppression” (Cohen, et al., 2003, pp. 457–458). What is important is to show that the indirect effect is significant.

Fidelity and quality indices were inserted in the analyses as follows. We began by conducting a series of analyses for which the fidelity and quality indices, rather than membership in a treatment group, were used as indicators for treatment exposure. The goal of this analysis was to verify whether: a) the results obtained by using the simple definition of the treatment would be reproduced; and b) the more sensitive analysis would show more marked and more persistent effects, or both. In a second series of analyses, we reintroduced membership in a treatment group as an indicator of treatment exposure, while controlling for the fidelity and quality indices. The goal of this second test was to verify whether the simplest way of defining treatment is redundant with fidelity/quality in explaining the fluctuation in child outcomes. If the fidelity and quality indices adequately capture the route by which treatment group membership exerts its effect on child outcomes, then we would expect the magnitude of treatment effect estimates to drop substantially when daycare quality and fidelity are controlled for in analyses or a significant mediator effect (indirect effect) for quality and fidelity in the presence of a suppression effect. Such a result would be further evidence that the estimated treatment effects reported here are not simply some methodological artefact, but instead indicate a true impact of the tested daycare.

8.4. BIAS CONTROL IN A LONGITUDINAL STUDY

The DinD estimator effectively neutralizes the static effect of observed and unobserved differences in the treatment groups at the baseline period.⁵³ By construction, such differences are eliminated by way of the initial differencing of Posttest and Pretest. Nevertheless, as discussed in the previous section, the DinD estimator may be biased when group composition changes over time and when pre-test differences are suspected in the developmental trajectories of the compared groups. We will address in turn each strategy used to manage both types of bias.

8.4.1. Changes in Group Composition

Changes in group composition can bias the DinD estimate. Such bias may occur when some characteristic that is related to an outcome varies over time for a group on average. For example, we know that gender is associated with a number of outcomes in the development literature. If the proportion of girls in a group increases suddenly before a post-test measure, the difference calculations required to calculate the DinD estimator will be biased.

This threat to statistical validity was controlled for in two ways. First, we monitored changes in group composition over time. For example, we verified whether participants changing groups or withdrawing from the study was associated with dependent variables or with treatment condition. We did not find any such association (see Appendix F). In the end, our preliminary analyses did not identify any significant problems with changes in group composition. We nonetheless resorted to a second strategy that consisted of including as covariates all the non-redundant baseline variables significantly associated with an outcome. The purpose of including covariates in this way was to maximize the validity of the DinD estimator.⁵⁴ It is assumed that

⁵³ The static effect of a baseline variable refers to an association with outcomes that is stable over time. One example would be the fact that the average effect of gender is comparable in scope to all first-year assessments. If the effect of gender was dynamic, its strength of association with an outcome would vary considerably depending on the assessment in question.

⁵⁴ In this case, the DinD estimator represents the estimated average treatment effect for our intervention, subject to the covariates included in the model. It is assumed that the treatment condition does not temper the effect of the covariates.

the set of covariates included in our model has adequately compensated for any bias arising from the selection of participants to treatment groups.

8.4.2. Non-parallel Developmental Slopes

Although group composition remains constant over time, initial differences among the treatment groups may be associated with non-parallel developmental trends. When present, non-parallel developmental trajectories (in the absence of treatment) jeopardize the validity of the DinD estimator. If the cause of this non-parallelism is observed (i.e., we collected the data for an appropriate covariate), it is possible to make statistical adjustments to correct that bias (Abadie, 2005). One example of this adjustment would be to include, as a covariate, interaction between time and the relevant variable. The variables we might use to make this adjustment include basic exogenous baseline covariates that are: a) related to the outcome measure; and b) distributed differently across the groups to be compared. Ideally, all covariates identified in this way would be treated this way to maximize control of the bias. However, the practical limitations imposed by our sample size require that we be selective in our choice of covariates for this adjustment in order to avoid over-fitting the data. We therefore only included Time by Covariate interaction terms if they were statistically significant predictors of an outcome (any outcome) for our sample (see the stability tests reported in Appendix F, Table F.8). As mentioned in Chapter 7, non-significant predictors do not necessarily require additional control.

8.4.3. Validity of Effects Based on a Variable Other than Group

In the discussion above, treatment exposure or “dosage” as well as the fidelity/quality of program delivery were presented as alternatives to treatment group membership alone for classifying the participants along a continuum of untreated to treated. Estimates of effects based on these more specific definitions may be more effective, but are subject to the same limitations as estimates based solely on treatment group membership. We cite here the assumptions that support the validity of our estimates of the treatment effect: that the treatment conditions, in the absence of any new investment by other levels of government, would have been comparable.

The main goal of the preliminary analysis was to establish the comparability of the treatment groups. If children with varying degrees of dosage are compared within groups and between groups, (e.g., to estimate the effect of hours spent at daycare), then we must assume that these children are otherwise comparable. For example, it is assumed that children who spend an average of 40 hours a week at daycare are equivalent in every other regard to children who spend only an average of 10 hours a week at daycare. This assumption has little chance of being met, since a range of sociodemographic variables are potentially associated both with time spent at daycare and with outcomes. We also assumed that program daycares and comparison daycares would have been equal in terms of fidelity and quality in the absence of intervention. This assumption was not verified empirically (i.e., no true pre-test measures were taken of these dimensions) and is unlikely to be (perfectly) met given the few daycares and the impossibility of matching on these variables.

We used two strategies to address these two potential sources of bias in the estimated effect of dosage and fidelity/quality. First, the longitudinal design of this study allows for use of the DinD estimator, which neutralizes the static effect of the baseline variables (both observed and unobserved). Second, the final specification used to estimate treatment effects contained all the

non-redundant covariates associated to at least one of the outcome measures. Assuming that all variables associated with varying degrees of daycare exposure (or with fidelity/quality) have been adequately accounted for using this strategy (i.e., that the model has been correctly specified), then the resulting conditional estimates of the dosage effects are unbiased. The same strategies and assumptions were enlisted to estimate the effect of program fidelity and quality.⁵⁵

8.4.4. Family Literacy Workshops: A Special Case

Data analysis for the Family Literacy Workshops must be treated separately because it constitutes a special case in the broader context of this report's analyses. Only two parental scales were used in each follow-up survey: Frequency of Literacy Activities and Language of Literacy Activities. An analysis comparable to that of the ÉPE-AD (i.e., using the DinD estimator) was conducted for these scales. For other outcomes (i.e., the knowledge, self-efficacy and modeling scales), the data were available only for parents whose children were enrolled in a program daycare. From this point on, this group of scales will be referred to as the Parental Workshop Scales. A somewhat different strategy, described in the sections below, was used to analyze these scales.

Parental Workshop Scales: Choice of Sample

The first distinguishing feature of the analysis for these parental subscales is that it is based on the total sample of participants recruited for the project and whose children were enrolled in a program daycare at the time of workshops delivery (N = 116, from six communities and two cohorts). Unlike the analysis of child outcomes, the communities of Edmonton and Saint John were not excluded in the analysis. There were three reasons for this decision. First, in contrast to the daycare intervention, the Family Literacy Workshops were correctly implemented in all the communities (see the *Project Implementation Report*). Next, since the analysis targets only parents whose children were enrolled in a program daycare, the availability of a Comparison Daycare group in all communities does not pose a problem. Finally, excluding the communities of Saint John and Edmonton would have reduced the sample size at a level where a fair test of the workshop impacts would be impossible. Characteristics for the total sample of families in the Program Daycare group are not reported here (see Chapter 3 for the response rates), but are detailed in the *Reference Report*.

Parental Workshop Scales: Analytical Strategy

The second feature that distinguishes the analysis of the Parental Workshop Scales is the absence of a planned comparison group. Measures were administered only to parents whose children were enrolled at a daycare delivering the tested program and there are no data for certain measures for families who participated in fewer than three workshops (n = 14). A balanced design would have a complete set of pre-test and post-test measurements on all variables for all participants, but this is not the case here. For reasons of efficiency, an unbalanced design was used, which required a selective method of data collection.

Specifically, the families classified as “non-participants” were only asked to provide post-test estimates of their attitudes, which means that retrospective estimates (Lamb & Tschillard, 2005;

⁵⁵ All the covariates are at the “participant” level rather than at the daycare level, except for the variable Community. There were too few daycares to include a broad set of characteristics at this level in the analysis.

Rockwell & Kohn, 1989) are not available for this group. A complete set of post-test data and retrospective pre-test data were available for the Knowledge and Self-Efficacy scales for those families classified as “participants.” True pre-test and post-test estimates were collected for the Modeling scale of all the families regardless of participation.. The analytical approach used in the analyses was adapted to the complications that this design entails.

Given the structure of the dataset, there are at least two ways to estimate the impact of the workshops. The first method estimates the workshops’ impact based on the difference between the pre-test and post-test measures. The second method estimates the workshops’ impact by comparing “participating” parents (n=116) with “non-participating” parents (n=14). Each type of comparison on its own lacks validity, but the global pattern produced by a series of tests can be informative.

The validity of the change scores (difference between the pre-test and post-test measures for participants only) is doubtful, not because they are retrospective (for a discussion on the advantages of retrospective estimates versus conventional pre-testing, see Lamb and Tschillard, 2003; 2005) but because no comparison group is used to control for maturation and history effects. This validity problem is exacerbated by the fact that participants, who are conscious of being treated, may estimate their pre-test and post-test states in a manner consistent with their expectations of a positive effect from the workshops.⁵⁶ Along the same lines, the comparison of participants and non-participants is invalidated by the obvious possibility of a biased selection that cannot be controlled for, either statistically or through matching because there are too few participants in the non-participant group.

Individually, the two treatment effect estimates are biased, but together they can produce informative results without completely ruling out all sources of bias. For example, we maintain that the following pattern of results is a true treatment effect: a) the treatment effect based on the participants’ change score is statistically significant and positive; b) the retrospective pre-test estimate for participants is equivalent to the “post-test” estimate provided by non-participants; and c) participants report higher post-workshop estimates than non-participants on average. It is assumed that in the absence of participation in the program, the “post-test” estimate by non-participants is as valid as an estimate of their pre-test state. In other words, we postulate that their true condition, on average, did not change during the period when the workshops were offered. If the expected pattern of results is obtained, this leads us to assume that the workshops had a positive effect. If only result (b) is obtained, then it can be assumed that the treatment effect is nil (or negative). If results (a) and (c) are obtained without result (b), then it can be strongly assumed that the estimates are biased and that interpretation of the program’s positive effects must be nuanced accordingly. Expressed otherwise, there is a specific pattern of results that strongly supports the real positive impact of the Family Literacy Workshops, and many other patterns that do not support this. If our “risky” prediction is confirmed, it lends credibility to the validity of our interpretation with respect to the treatment effects.

⁵⁶ This type of bias may take a number of forms and is not specific to retrospective pre-test estimates. Parents may exaggerate the program effect by under-estimating their pre-test state and overestimating their post-test state, or by producing estimates contaminated in both ways. Experience has shown that an underestimate of the pre-test state is characteristic of retrospective measures (Taylor, Russ-Eft & Taylor, 2009). Conventional prospective pre-test measures are also biased in other ways (e.g., improved response bias; Howard, 1980), which sometimes results in an overestimate of the pre-test state (Moore & Tananis, 2009).

The limitations imposed by the available data mean that an estimator like the DinD estimator must be ruled out for variables without a pre-test measure for both groups. The three hypotheses formulated above may instead be tested using a series of independent tests, one for each hypothesis. The covariates could not be used to test for differences between the groups because the sample size for the non-participant group was too small ($n = 15$). Moreover, covariates were not used to estimate the participants' change scores because: a) the static effect of participant characteristics is neutralized ("differenced out") in the process of this estimation; and b) the changes in group composition are not an issue.

Estimates for program fidelity, quality and dosage (proportion of workshops attended) were available for the sample of parents who attended the workshops ($n = 114$). The impact of these factors was assessed using a DinD estimator as in the other analyses reported here. The resulting estimates represent the workshops' effects in terms of the intensity of treatment received. Again, since changes in group composition were not an issue for this analysis, we relied on the fact that the DinD estimator is unbiased by either the observed or the unobserved baseline characteristics.

8.4.5. Moderation and Mediation Tests

The analysis strategy for this report is based on several secondary analyses. We test whether the program had an impact, but also whether that impact is tied to quality/fidelity, dosage and linguistic profile. The contribution of each analysis is limited by the sample's modest size, but together they can rule out several alternative explanations for the results, provided they produce a coherent pattern of results. They can prove that the mechanism that generated the results is indeed well understood, which enables us to anticipate certain contingencies in achieving the program effect.

8.5. PLAN FOR THE NEXT CHAPTERS

The following chapters are organized as follows. Chapter 9 presents an assessment of the tested program's impact on children. It also presents a comparison of the treatment groups as well as a series of more in-depth analyses. We describe analyses examining how the effect of the tested program depends on factors like dosage, the fidelity and quality of the Daycare Program's implementation, and the children's linguistic profile. In Chapter 10, the impact of the family literacy component on parental behaviour and attitudes is assessed. We report the results of analyses examining how the Family Literacy Workshops' effect on parents depends on factors like dosage and the fidelity/quality of program implementation. This chapter also presents the result of an analysis examining the workshops' indirect effect on children. This last analysis connects the changes observed among parents to child development.

9. The Tested Program's Impact on Children

The two core components of the intervention were the new Daycare Program and the Family Literacy Workshops. The main goal of the first component was to *directly* influence child outcomes, while that of the second component was to *indirectly* influence child outcomes by modifying parents' attitudes and behaviours. The analyses reported in this chapter cannot clearly distinguish the effect of one component from that of the other. A more complex experimental design would have been necessary to allow for that distinction. As such, the main analyses comparing the treatment groups test the effect of the dual-component program, that is, the *combined effect* of the program's two components on child development. This is the primary topic of this chapter.

Nevertheless, analyses can be performed to give an idea of the relative importance of the two components. For the daycare component, analyses of dosage and quality/fidelity of program delivery quantify the magnitude of its contribution to the program effects. The greater the effect of these variables is on child development, the greater the contribution of the Daycare Program component to children's outcomes during the comparison of treatment groups. This chapter presents the results of these secondary analyses. For the family literacy component, child development is identified through the changes reported by parents. This second series of analyses is examined in Chapter 10 after the results of analyses on the Family Literacy Workshops' impact on parents.

In this chapter, data from children in the first and second year of project implementation were analyzed separately. We conducted the analyses in this manner because there were too few degrees of freedom to include all the assessments in a single regression model.⁵⁷ What's more, because most children entered kindergarten at the start of the program's second year, that was a natural conceptual breaking point. See Section 9.3 for a graphic presentation of all results for both years.

Section 9.1 provides a description of the variables used for the impact analyses. The analyses for the first year are presented in Section 9.2 and the analyses for the second year, in Section 9.3. The technical details specific to each analysis are found in Sections 9.2.1 and 9.3.1. We report a series of additional analyses for both years of the project.

The impact of the tested program is evaluated: (a) by comparing the three treatment groups (see Sections 9.2.2 and 9.3.2); (b) by determining whether the program effect varies based on dosage (see Sections 9.2.3 and 9.3.3); (c) by testing the direct effect of program fidelity and quality on the development of children attending daycare (see Sections 9.2.4 and 9.3.4); (d) by testing whether program fidelity and quality explain all or part of the observed differences between the Program Daycare group and the Comparison Daycare group (see Sections 9.2.4 and 9.3.4); and (e) by re-evaluating the differences among the treatment groups, on one hand, for children whose exposure to French is high and, on the other, for those whose exposure to French

⁵⁷ The number of daycares determines the degrees of freedom (about 20) for tests of significance. The inclusion of all the measures repeated in the same analysis would drop the degrees of freedom below the bar of ten degrees for certain tests of significance, which means that a fair test of the program would not be possible.

is low (see Sections 9.2.5 and 9.3.5). The purpose of the latter analyses is to determine the subpopulations for which the tested Daycare Program seems particularly effective.

The following points should be noted:

- The impact analyses reported in this chapter concern four communities: Orleans, Cornwall and Durham in Ontario and Edmundston in New Brunswick. The communities of Edmonton, Alberta, and Saint John, New Brunswick, were excluded from the analysis for methodological reasons: in Edmonton, the Daycare Program was never really implemented, and in Saint John, there was no Comparison Daycare group. Excluding these communities allows us to report a set of valid and coherent results.
- The first cohort of participants consists of families from all four communities, while the second cohort consists only of families from Orleans and Cornwall.

9.1. TREATMENT OF VARIABLES

Among the numerous variables measured during the study, the following were retained for the impact analyses of the first and second year. These variables can be divided into two categories: outcome measures and predictors. We now describe how both types of variables were treated in the impact analyses.

9.1.1. Outcome Measures

First Year

This impact report concerns French-language school readiness. That is why we measured the following school-readiness domains using only the French version of the ÉPE-AD scales: Language and Communication (Communication), Awareness of Self and the Environment (Self-Awareness), Cognitive Skills (Cognition) and Physical/Motor Development (Physical). Complete data were only available for the Communication scale.⁵⁸ We also created more specific vocabulary subscales by choosing and combining relevant items found in all four scales. By creating these scales, we made a well documented distinction between receptive vocabulary (the Receptive Vocabulary scale) and expressive vocabulary (the Expressive Vocabulary scale). In total, six child outcomes measured by the ÉPE-AD were the object of the following analyses. Note that children who withdrew from the project before the end of the first year are excluded from these analyses.

Second Year

In total, six child outcomes were taken into account for the impact analyses of the second year: Communication, Self-Awareness, Cognition and Expressive Vocabulary measured by the ÉPE-AD. In addition, the analyses concerned the outcomes commonly used in studies on school

⁵⁸ For the three other scales, data were available only for participants who had taken the ÉPE-AD in French at least twice in the first five assessments. For obvious reasons, this meant that the analyses conducted using the scores for the Awareness of Self and the Environment, Cognitive Skills, Physical/Motor Development, Receptive Vocabulary and Expressive Vocabulary scales are effective subsample analyses based on a subgroup that excludes children with weaker French-language skills. Because this exclusion was not strongly tied to groups, the internal validity for estimates of program effects is not jeopardized. In a way, this issue affects external validity by limiting the general application of results to children who met the criteria for taking the test in French.

readiness, including receptive vocabulary (measured using the *Échelle de vocabulaire en images Peabody — Révisé*, or ÉVIP-R, the validated translation of the Peabody Picture Vocabulary Test – Revised, or PPVT-R) and expressive vocabulary (measured using the *Épreuve de dénomination de Gardner*, the validated and standardized French translation for children of the Expressive One-Word Picture Vocabulary Test, or EOWPVT). The last two measures were administered respectively at the sixth and seventh assessments. Note that the Physical and Receptive Vocabulary scales were not part of the second-year analyses⁵⁹ and that children who withdrew from the project before the end of the second year are excluded from the analyses.

Standardizing Outcomes

Before conducting the analyses, outcomes were rescaled so that they were standardized within time period. As a result, each outcome variable had a mean of 0 and a standard deviation of 1 for the total sample of participants. One practical effect of this transformation is that the scores for each time period can be interpreted using a common scale. The decision to standardize the scores was based on the fact that the nature of the scale's raw scores changed qualitatively in the second year of the study. More specifically, the composition of the scales was altered (i.e., items were dropped), thereby changing the total number of items. The strong correlation between the original scales and the revised scales supports our theory that very little information was lost further to these modifications. Nevertheless, the differences in scaling (e.g., maximum score) in our measures across time presented a technical obstacle to statistical analyses in that the meaning of the absolute values of scores varies over time, invalidating any comparisons across time period.. We overcame this minor technical challenge by using the standardization procedure discussed above, which ensures that the outcome scores retain the same significance over time.

Interpreting Standardized Scores

The standardized scores used in the impact analyses are interpreted as follows. Each participant's standardized score represents the difference between the participant's score and the average score of the sample. Critically, this difference is expressed in standard deviation units. For example, a score of 1.11 means that the participants achieved a score that was 1.11 standard deviations above the mean score achieved by participants in the Readiness to Learn project sample for the corresponding assessment. DinD estimates of program effects are similarly interpreted as differences between groups in standard deviation units.

A “standardized” difference between groups is the most common method of expressing an effect size. Following Cohen (1988), we call this statistic *d*. Cohen provides conventional benchmarks for interpreting the magnitude of the effects expressed in a standardized scale. A standardized difference between groups of $d = 0.20$ is considered small, a difference of $d = 0.50$ is considered medium, and a difference of $d = 0.80$ is considered large. These benchmarks match the findings of a meta-analysis examining the distribution of standardized size effects of various intervention studies (Lipsey & Wilson, 1993). However, they should only be used as a general guide in determining the importance of an effect. An effect may be deemed more or less important based on the research context (Kane, 2004; Hill, Bloom, Black, & Lipsey, 2008).

⁵⁹ Data for the Physical scales and the Receptive Vocabulary subscale of the ÉPE-AD were available for the fourth and fifth assessments, but the analyses based on these measures are reported only for analyses of data collected in the first year (see Section 9.2).

According to Kane (2004) and to Hill and colleagues (2008), one useful way to understand the importance of the effect of an early childhood intervention is to compare it with the effect of normal development. In other words, we must ask: How does the size of the effect compare with gains normally observed over one year of development? According to Hill and colleagues (2008; Table 1), the expected average gain in literacy and numeracy development for the period spanning from kindergarten to Grade 1 is approximately $d = 1.33$. Stated otherwise, a program effect of $d = 1.33$ represents one year of development and an effect of 0.67 represents a developmental gain of roughly six months.

Weighing the Costs of Standardization

The cost of standardization is that we lose the ability to: a) compare the raw scores of our sample to a normative population; and b) evaluate the sample's developmental trajectory over time.⁶⁰ We argue that these costs are trivial because: a) no population norms exist for the ÉPE-AD version used in the Readiness to Learn project; and b) the children in the study were assessed several times, which no doubt induced a massive practice effect (i.e., better scores through repetition of a task).⁶¹ These practice effects are necessarily confounded with any attempt to estimate the development of our sample over time. In other words, the intrinsic meaning of the absolute value of observed raw scores is limited, while changes in scores over time (i.e., the developmental trajectory of our sample) cannot be disentangled in any case from practice effects. One important fact is that the practice effect is the same for all three treatment groups because they were assessed the same number of times. This allows us to analyze the standardized data using the DinD method, comparing the developmental trajectories for our three treatment groups.^{62, 63}

9.1.2. Substantive Predictors and Covariates

a number of variables were entered into the regression equations to model the development of children's school readiness outcomes. The broadest distinction that can be made here is that between the substantive predictors and the variables employed only as covariates. The substantive predictors are of particular interest in research terms, while the covariates serve only to improve the internal validity of more important tests. In what follows, we present an overview of substantive predictors included in the impact analyses as well as the covariates used for adjusted models.

⁶⁰ Standardization of scores gives a mean of zero. Standardization within each period of time gives an average of 0 for each period of time. If the mean score is 0 for each period of time, then the average is constant across time and the effect of "Time" in a regression model will necessarily be null. When the overall developmental gains for the sample of participants are not of direct interest, as is the case in this study, this represents a trivial loss of information.

⁶¹ Although the sample's overall trajectory cannot be evaluated, the trajectories of the groups can still be compared through a DinD analysis.

⁶² Even if the sample mean is zero, groups' means are not. They vary freely from one assessment to the next, which means they can be compared through statistical analyses.

⁶³ Standardization obscures certain characteristics of the distribution of raw scores. Note that the preliminary analyses showed that all outcome measures for children were sensitive to change over time. Also, none of the measures had reached a ceiling at the fifth assessment.

Substantive predictors

The list of substantive predictors includes time (assessment period), the child's group, the average number of hours the child spends at daycare every week (dosage), the quality/fidelity of childcare services, and the household's linguistic composition. Each of these variables identifies one or another dimension of the tested program's effect, that is, separately or in combination. The exact specification used to model the effect of these variables may vary. These details are specific to each analysis and are specified in the text.

Time

Time is a categorical variable representing the assessment period at which an observation was collected. This variable serves to identify the dynamic effect of the covariates as well as the effect of the tested program through the DiD estimator.

Group

Group is a categorical variable representing the treatment group to which a child and his or her family belong. They belong to one of three groups: the Program Daycare group, the Comparison Daycare group, or the Informal Care group. In the tables, these groups are sometimes referred to respectively as G1, G2 and G3. The rules used to determine membership in a group from one assessment period to the next are described in Section 8.3.2.

This variable is crucial to identifying the program's effects. It is often combined with other variables to enhance the validity of the estimate (e.g., in the case of the DiD estimator) or to test the effect of a potential moderator (e.g., dosage).

Dosage

Dosage is a continuous scale variable representing the average number of hours per week that a child spent at daycare. The value taken by this variable may vary over time (the sample's averages for various periods are reported in Appendix F, Table F.6), as may the method of calculation. In the first year, dosage is based on children's daycare attendance in the months prior to the assessment. In the second year, dosage represented the average dosage for the project's first year. More details in this regard are provided in Section 8.3.2. The interest of including dosage in the analyses is to verify whether this variable exerts a control over the size of the program effect.

Fidelity and Quality

The implementation study provided information on the daycare programs when children were finishing the first year of the project. This information was used to calculate the indices for program fidelity and quality for daycares in the Program Daycare group and in the Comparison Daycare group. The program delivered at each daycare was scored for a number of fidelity and quality indices. Two separate series of scores were assigned to daycares that delivered the program to two cohorts of children. Each series represents the program as offered to children in their first year of participation.

Note that for the purposes of the impact analyses, no distinction is made among the educators at a given daycare. Daycares with more than one educator were given a score reflecting the mean

scores of its educators. This data grouping simplifies the structure of the database for impact analyses.

Two fidelity indices were retained as daycare descriptors: one for the **Structural Fidelity** of implementation and the other for **Content Fidelity**. Both express the proportion of the tested program's elements that were implemented at each daycare. Structural Fidelity reflects the presence of elements in the environment, such as picture/word cards or routine charts. Content Fidelity, on the other hand, indicates the extent to which program elements have been added to the childcare programming. For example, circle time may include reading or reasoning activities. Section 8.3.3 explains how these two indices were calculated.

Four quality indices were also retained as descriptors of the program and the environment in which it is delivered. These indices express the level of program quality at the daycares on a seven-point scale where "1" means care is well below fundamental childcare requirements and "7" means personalized, high-quality care (Harms, et al., 1998). These quality measures are from the Early Childhood Environment Rating Scale, Revised (ECERS-R). For the purposes of the impact analyses, certain ECERS-R subscales were grouped together to form four quality indices of interest. The first index, **Structural Quality**, measures the overall quality of the environment in childcare settings. It includes the Indoor Space, Health Practices and Greeting/Departing subscales. The second index, **Educative Quality**, gives special attention to activities that encourage children to communicate and broaden their vocabulary. This index combines the ECERS-R subscales Informal Use of Language, Encouraging Children to Communicate, and Using Language to Develop Reasoning Skills. The third index, the Staff-Child Interactions subscale, was used as an index of the **Educator Sensitivity** in order to factor in this particularly important factor to child development. Finally, **Reading Quality** isolates the Books and Pictures subscale, given the special importance of reading with respect to a child's vocabulary acquisition.

In total, six indices were used to estimate the role of the Daycare Program in child development. Depending on the objectives of a given analysis, these indices were treated separately or were combined to form global fidelity and quality indices. Global indices are calculated using a simple average of the indices in order to assign an equal weighting to each index.

In all cases, these variables were used to verify the hypothesis that the quality and fidelity indices capture the mechanism through which the Daycare Program exerts its effect on child development. The effect of these indices is estimated in the same way as for the group's effect: through a DiD estimator.

Household Type

Section 5.1.3 of the *First Cohort Findings Report* (2014) explains in detail how children's linguistic profile was defined. The strategy adopted in the *First Cohort Findings Report* was to report the results for several indicators of children's linguistic profile. These indicators varied based on the measurement unit (continuous or categorical) and their definition of exposure to French (active versus passive).

This report provides the results for a single linguistic profile indicator: household type. This dichotomous variable contrasts families with high and low exposure to French. We define

exposure to French on the basis of the Language Spoken by Mother to Child crossed with the Language Spoken by Father to Child.⁶⁴ This choice is justified by the fact that a categorical variable generates results that are easier to understand in the context of a moderation analysis than does a continuous variable.⁶⁵ Families with low and high exposure can be seen as two points along a continuum.

The analyses detailed in the *First Cohort Findings Report* (2014) tended to confirm the hypothesis that the benefits of the tested program would be seen mainly in terms of linguistic outcomes (e.g., vocabulary) for children characterized by a lower exposure to French while children characterized by a higher exposure to French would benefit mainly through accelerated development of their Cognitive Skills (the Cognition scale). This hypothesis is based on a development model that recognizes that proficiency in the language of instruction facilitates the acquisition the academic skills targeted by the Cognition scale. According to this conceptualization, the tested program positively affects the development of all children, but this impact varies based on their individual readiness.

This analysis is repeated in this report in order to verify whether including the second cohort affects the earlier findings in any way. The hypothesis of a distinct program effect based on child readiness is tested by juxtaposing the factors of time, group and household type to determine whether their interaction is significant (through the DinDinD estimator). This approach allows us to estimate the program effect independently for children from high-exposure and low-exposure families.

Covariates for the First Year

For the first-year analyses, the static and dynamic effects of 12 covariates are represented in the specification of adjusted models. The static effect is captured by the direct effect of the variables while the dynamic effect is captured by their interaction with the time factor (i.e., the dummy variables representing each assessment period). For more details, see Section 8.4 of this report.

Covariates for the Second Year

The same covariates are found in the adjusted models for the second year. A 13th variable is added to the list to represent school enrolment. Recall that at the beginning of the project's second year, certain children were enrolled in school full-time, others part-time, and still others were not enrolled in school. A categorical variable representing these degrees of exposure to school was inserted as a covariate in our regression analyses. This variable is coded as 0 for the baseline period and as 0 or 1 as of the fourth assessment for children enrolled in school; the values assigned to each child remain the same for the project's entire second year, that is, from the fourth to the seventh assessment period.

⁶⁴ The *First Cohort Findings Report* (2014) also considers an alternative definition for household type based on the parents' first official language spoken. We argued that this definition did not properly capture the child's language experience at home.

⁶⁵ It is true that continuous variables like Continuum of French Spoken by the Child are likely more effective in revealing the program's true effects. Thus, the moderating effect of the linguistic profile is sometimes significant with this continuous variable and not with the household type (unreported analyses). That said, the body of results is not clarified when this continuous variable is used. The simpler analysis based on household type conveys essentially the same information in a clearer manner.

9.1.3. Non-Parametric Cross-Validation

The findings detailed in this report are from parametric analyses designed to test average differences among the groups. Parametric analyses are very powerful in detecting effects - a highly attractive attribute when analyzing data from of a small sample. However, they are criticized because their validity is based on several more or less plausible assumptions that are difficult to establish empirically. In response to these potential criticisms, we verified the robustness of the results by estimating logistic ordinal regression models (using the “logit” function) with a robust standard error for all outcomes considered in this chapter. To shorten the presentation, we present the results of non-parametric analyses only for the adjusted models (see Section 9.2.2).

Non-parametric analyses are used to confirm whether the program effect is significant for a given DinD effect. This interpretation of the analyses is given in the text. The following technical details should be noted. The logistic regression analysis concerns a simplified version of the outcomes created by converting them into ordinal variables (i.e., scores were broken down into five intervals, or quintiles). The program estimates four parameters representing the likelihood of being in a higher quintile: i) higher than the first quintile, ii) higher than the second quintile, iii) higher than the third quintile, iv) higher than the fourth quintile.⁶⁶ These parameters are combined to form a single global index representing the likelihood of being in a higher quintile. This global index (an odd) is estimated for the three treatment groups at each assessment.

A DinD analysis produces “odds ratio” statistics used to evaluate the program impact on the global index representing the likelihood of being in a higher quintile. If there is no difference between the groups, the odds ratio takes the value of 1. If the confidence interval for the odds ratio does not straddle the value of 1, we can conclude that the effect is significant. Confidence intervals of 90% are reported, which corresponds to a 10% significance level. In other words, we postulate that the estimated interval will include the true population value of the odds ratio in most cases (9 out of 10 independent samples).

9.2. IN THE FIRST YEAR

This section presents the results of impact analyses for the first year. Treatment effect estimates are based on the relative developmental trajectory of the study’s three groups over the first four assessments. The first assessment is considered a pre-intervention (baseline) assessment and the three subsequent assessments, post-intervention assessments. As mentioned, the treatment effect is measured by the DinD estimator, thus capturing the differences in the developmental trajectories of the treatment groups.

In total, six child outcomes were analyzed based on the data collected through the French version of the ÉPE-AD. The tested program is expected to have a positive effect for all measures except the Physical scale. Below we describe how the outcome measures were treated in the impact analyses.

⁶⁶ It would be redundant to include a fifth parameter.

9.2.1. Technical Details

It is important to determine the extent to which the results depend on the specification used to model the data. The general strategy adopted here is to present two alternative specifications for each analysis. The first, the initial model, considers only the key predictors used to identify the program effect. The second specification is more complex since it also includes the covariates. Both specifications produce results that, when compared, allow us to identify technical problems or nuance the interpretation of data. For this comparison exercise, the initial model serves mainly to enhance the interpretation of the final results from the adjusted model. We also present, for the adjusted model only, the results of an additional analysis using a non-parametric approach.

Details of Model Specification

Initial Model

The initial model of the analysis by treatment group includes a series of dummy variables representing the assessment period, group membership, and a term representing interactions among the indicators for these dummy variables (i.e., the DinD estimators). The time factor or “assessment period” is represented by three dummy variables comparing the follow-up assessments to the first assessment (i.e., the pre-intervention period). In other words, the pre-intervention period is used as the benchmark. The “group” factor is represented by two dummy variables used to compare the comparison groups to the Program Daycare group. In other words, the Program Daycare group is used as the benchmark. *It should be noted that the decision to use the Program Daycare group as the benchmark means that **the negative values of DinD estimates for all group comparisons are a positive treatment effect** (i.e., an advantage for the Program Daycare group).* The initial model is modified to test the various hypotheses (e.g., adding dosage or quality to the model). Changes made to the initial model are noted in the text at the beginning of each subsection.

Adjusted Model

The initial model is nested in the adjusted model, which itself also includes all the covariates identified in Chapter 7 as well as their interaction with the time factor. A complete list of the effects included in the adjusted model is provided in Chapter 7 (Section 7.3).

Analysis Samples

Baseline Samples

The sample used in the analyses varies based on the outcome variable analyzed. The **entire** sample of participants contributes to analyses for the Communication scale. For the other scales, the analyses concern a **subsample** that excludes cases with missing data due to English-language administration of the ÉPE-AD and for which values could not be imputed.

Data Screening

Children who were withdrawn from the study before the 12-month assessment are excluded from the analyses to maximize the internal validity of the DinD estimators. This estimator is also sensitive to bias resulting from the presence of outliers. This type of score can excessively

influence the values of the regression parameters. A data screening rule is typically imposed in order to eliminate the most extreme observations. The rule adopted in the context of the reported analyses affects less than 0.01% of the observations. Based on this rule, an observation is valid if it ranges from -3.5 to +3.5 standard deviations from the group average for a given assessment. The observations affected by the rule vary from one analysis to the next because the rule refers to parameters that are specific to the variable being analyzed.

Certain cases may also be inconsistent due to an atypical combination of characteristics measured by the covariates used in the regression model. These cases can negatively affect analysis results because they excessively influence the estimate of the regression parameters. Preliminary analyses (unreported here) did not reveal any problematic cases with multivariate outliers on covariate scores,⁶⁷ which suggest that no particularly influential case is detrimental to the generalization of results from the adjusted model. This is a particularly important consideration for the analysis of such a modest sample.

9.2.2. Analyses by Group

This subsection presents analyses by treatment group. The results for the initial model are presented in Table 9.1, while those of the adjusted model are presented in Table 9.2. The results for each scale are described first for the initial model, and then for the adjusted model. A non-parametric cross-validation of the results is then presented for the adjusted model. The section ends with a summary of results for the analyses by group.

Communication Scale

Initial Model

The baseline effect reported in Table 9.1 represents the relative position of the groups for the pre-intervention period (first row). A positive value indicates a deficit for the Program Daycare group while a negative value indicates an advantage. For Communication, a non-significant deficit of 0.45 and of 0.23 respectively is observed for the Comparison Daycare group and the Informal Care group. These non-significant differences are cancelled in the estimate of the DiD effects based on which we observe the program's constant effect over time on Communication for the two comparison groups. Moreover, the gains of children in the Program Daycare group for each post-intervention assessment are significantly greater than those observed for children in the comparison groups. The size of this effect is of the order of 0.30 to 0.40 standard deviations. Note that a simple comparison of the values indicated in the table is sufficient to show that the program's positive effects fall short of offsetting the initial deficit of the Program Daycare group in this domain.

Adjusted Model

The baseline effect reported in Table 9.2 represents the relative position of the groups for the pre-intervention period after adjusting for the covariates. As with the initial model, a positive value indicates a deficit for the Program Daycare group while a negative value indicates an advantage. The regression model with controls for confounding variables partially corrects the

⁶⁷ We evaluated the similarity of cases in terms of their combination of characteristics by estimating their distance from the multivariate average, i.e., their Mahalanobis distance value. None differed from this statistic's empirical distribution.

non-significant disadvantage observed in the baseline for the Program Daycare group versus the Comparison Daycare group. The size of the deficit for the Comparison Daycare group is only 0.33 standard deviations in this analysis. The non-significant deficit for the Program Daycare group versus the Informal Care group is cancelled.

The DinD estimates for program effect generally rise for the Communication scale due to inclusion of the covariates. This is the case in particular for all DinD tests for comparison with the daycare groups and for comparison with the Informal Care group. The DinD effects for the comparison of daycare groups are approximately 0.40 standard deviations throughout the first year of the study. The corresponding effect for comparison with the Informal Care group seems to emerge later since an effect of 0.40 is observed only at the 12-month assessment.

Non-Parametric Cross-Validation

In this non-parametric analysis, the children's observed performance for the Communication scale is expressed as the probability of being classified in a higher quintile: the success category. The analysis results reproduce the degree of significance for the DinD effects for comparison of the daycare groups. The odds ratio is 0.31 (90% CI from 0.20 to 0.47), 0.44 (90% CI from 0.25 to 0.80), and 0.41 (90% CI from 0.22 to 0.78) in favour of the Program Daycare group respectively for the DinD effects at 4, 8 and 12 months. In other words, children in the Comparison Daycare group are less likely to be in the upper quintile (i.e., the success category) by 69%, 56% and 59% respectively at 4, 8 and 12 months compared to children in the Program Daycare group.

The non-parametric analysis also shows that the program impact for the Informal Care group is less robust. A significant effect for the DinD estimator is observed only at 4 months, while examination of the 90% CI shows that the 12-month effect is close to the 10% significance level. The odds ratio is 0.69 (90% CI from 0.50 to 0.95), 0.83 (90% CI from 0.51 to 1.35) and 0.64 (90% CI from 0.40 to 1.02) in favour of the Program Daycare group respectively for the DinD effects at 4, 8 and 12 months. In other words, children in the Informal Care group are 31% less likely to be in the upper quintile at 4 months (i.e., the success category) compared to children in the Program Daycare group for the four-month assessment.

Self-Awareness Scale

Initial Model

The baseline effect reported in Table 9.1 represents the relative position of the groups for the pre-intervention period. A positive value indicates a deficit for the Program Daycare group while a negative value indicates an advantage. We observe a non-significant deficit of 0.36 and 0.25 for the comparison daycare groups and Informal Care group respectively. As regards the DinD estimates, we observe a program effect that emerges for the Comparison Daycare group at 4 months and is maintained until the end of the first year. The size of the effect declines from about 0.32 standard deviations at 4 months to 0.21 standard deviations at 12 months. As with the Communication scale, the effect of the tested program, although significant, does not manage to fully offset the deficit observed for the Program Daycare group at the baseline. In comparison with children in the Informal Care group, a significant effect is detected only at eight months for

the Program Daycare group. However, the observed effect lacks credibility due to its instability over time.

Adjusted Model

The baseline effect reported in Table 9.2 represents the relative position of the groups for the pre-intervention period after adjusting for the covariates. As with the initial model, a positive value indicates a deficit for the Program Daycare group while a negative value indicates an advantage. The regression model with controls for confounding variables shows a smaller disadvantage for the Program Daycare group than for the comparison groups. In comparison with the Comparison Daycare group, the observed deficit is reduced to 0.25 standard deviations whereas it is almost entirely cancelled with the Informal Care group.

In terms of the DinD effects, we observe a program effect that is constant over time, indicating a significant advantage in favour of the Program Daycare group versus the Comparison Daycare group even after controlling for confounding variables. In this model, the size of the effect is relatively stable over time, set at roughly 0.30 standard deviations. However, the DinD estimate in favour of the Program Daycare group versus the Informal Care group is significant only at eight months. As with the initial model, the latter may be deemed as lacking credibility due to its instability over time.

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance with the Self-Awareness scale is expressed as the probability of being classified in a higher quintile: the success category. The analysis results reproduce the significance of the DinD effects for comparison of the Program Daycare group with the Comparison Daycare group. The odds ratio amounts to 0.49 (90% CI from 0.36 to 0.66), 0.58 (90% CI from 0.37 to 0.90) and 0.55 (90% CI from 0.39 to 0.79) in favour of the Program Daycare group respectively for the DinD effects at 4, 8 and 12 months. In other words, children in the Comparison Daycare group are less likely to be in the upper quintile by 51%, 42% and 45% at 4, 8 and 12 months compared to children in the Program Daycare group.

The non-parametric analysis also reproduces the observed program effects for the Informal Care group; the only significant DinD effect is found for the eight-month assessment. The odds ratio for this effect is 0.55 (90% CI from 0.36 to 0.83) in favour of the Program Daycare group, indicating a 45% higher likelihood for Program Daycare group children to be in the upper quintile (i.e., success category).

Cognition Scale

Initial Model

The baseline effect reported in Table 9.1 represents the relative position of the groups for the pre-intervention period. A positive value indicates a deficit for the Program Daycare group while a negative value indicates an advantage. We find a significant deficit for the Program Daycare group versus the comparison groups at the baseline. What's more, the DinD analysis shows no significant program effect in terms of the children's subsequent developmental gains for this

dimension of school readiness: the overall effect tested by the Wald statistic and the DinD estimates are not significant.

Adjusted Model

The baseline effect reported in Table 9.2 represents the relative position of the groups for the pre-intervention period after adjusting for the covariates. As with the initial model, a positive value indicates a deficit for the Program Daycare group while a negative value indicates an advantage. The regression model with controls for confounding variables did not correct the deficit of 0.20 standard deviations for the Program Daycare group in comparison with the Comparison Daycare group at the baseline. However, the baseline deficit for the Informal Care group is completely cancelled.

The adjusted DinD analysis reveals significant positive program effects of 0.20 standard deviations observed at 12 months for the two comparison groups. One possible interpretation is that this result is a valid effect that emerges late then persists in the project's second year. It remains to be seen whether this hypothesis will be confirmed with the impact analyses of the second year.

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance with the Cognition scale was expressed as the probability of being classified in a higher quintile: the success category. The analysis results partially reproduce the results of the adjusted model. Thus, the effect at 12 months for comparison of the daycare groups is not significant with an odds ratio of 0.76 (90% CI from 0.48 to 1.19) while that for comparison with the Informal Care group is significant with an odds ratio in favour of the Program Daycare group of 0.57 (90% CI from 0.39 to 0.82).

Table 9.1: Unadjusted Program Impact on School Readiness in the First Year — Standardized Scores

Type of Difference	School Readiness Indicator											
	Communication		Self-Awareness		Cognition		Physical		Receptive Vocabulary		Expressive Vocabulary	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
G1 vs. G2												
Baseline	0.45	0.32	0.36	0.25	0.27**	0.12	0.30	0.19	0.32	0.33	0.43	0.26
4 months (DinD)	-0.43***	0.10	-0.32***	0.04	-0.13	0.13	-0.10	0.13	-0.20	0.30	-0.44***	0.12
8 months (DinD)	-0.27**	0.10	-0.24**	0.11	-0.01	0.16	-0.32**	0.12	-0.41	0.25	-0.20	0.13
12 months (DinD)	-0.30***	0.10	-0.21*	0.11	-0.13	0.15	-0.22	0.15	-0.34	0.21	-0.22**	0.10
G1 vs. G3												
Baseline	0.23	0.33	0.25	0.28	0.13	0.09	-0.09	0.16	0.13	0.35	0.30	0.25
4 months (DinD)	-0.24**	0.09	-0.01	0.10	0.00	0.10	0.13	0.11	-0.11	0.29	-0.01	0.16
8 months (DinD)	-0.24*	0.13	-0.26*	0.13	-0.01	0.15	0.01	0.13	-0.04	0.23	-0.10	0.18
12 months (DinD)	-0.33**	0.14	-0.10	0.08	-0.10	0.14	0.14	0.13	-0.23	0.21	-0.05	-0.05
Group x Time Wald F	4.18***		9.34***		1.11		2.11		3.00**		3.59**	

Note: The degrees of freedom for the Wald F-test are 6 and 17. The Huber-White robust standard errors are reported with error terms clustered by daycare. *The negative values of DinD estimates for all group comparisons are a positive treatment effect (i.e., an advantage for the Program Daycare group).* Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Table 9.2: Program Impact (after Adjusted for Covariates) on School Readiness in the First Year — Standardized Scores

Type of Difference	School Readiness Indicator											
	Communication		Self-Awareness		Cognition		Physical		Receptive Vocabulary		Expressive Vocabulary	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
G1 vs. G2												
Baseline	0.33***	0.08	0.26***	0.07	0.17**	0.07	0.28**	0.10	0.19*	0.10	0.32***	0.08
4 months (DinD)	-0.46***	0.10	-0.36***	0.05	-0.15	0.10	-0.13	0.14	-0.16	0.22	-0.43***	0.12
8 months (DinD)	-0.39***	0.12	-0.32**	0.11	0.01	0.12	-0.35***	0.11	-0.33**	0.15	-0.25*	0.13
12 months (DinD)	-0.43***	0.13	-0.24**	0.09	-0.16*	0.08	-0.23**	0.10	-0.27**	0.12	-0.26**	0.12
G1 vs. G3												
Baseline	0.01	0.06	0.03	0.07	0.04	0.08	-0.04	0.11	-0.07	0.09	0.05	0.07
4 months (DinD)	0.10**	0.09	-0.06	0.06	-0.03	0.12	0.03	0.12	-0.03	0.20	-0.01	0.11
8 months (DinD)	-0.27**	0.11	-0.37***	0.11	-0.08	0.12	0.01	0.07	0.05	0.17	-0.17	0.14
12 months (DinD)	-0.40***	0.11	-0.12	0.08	-0.21***	0.06	0.08	0.14	-0.20*	0.11	-0.10	0.09
Group x Time Wald F	4.74***		11.20***		4.89***		3.33**		2.07		4.56***	

Note: The degrees of freedom for the Wald F-test are 6 and 17. The Huber-White robust standard errors are reported with error terms clustered by daycare. *The negative values of DinD estimates for all group comparisons are a positive treatment effect (i.e., an advantage for the Program Daycare group).* Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Physical Scale

Initial Model

The baseline effect reported in Table 9.1 represents the relative position of the groups for the pre-intervention period. A positive value indicates a deficit for the Program Daycare group while a negative value indicates an advantage. The Wald statistic testing the overall effect of the interaction between group and assessment period is not significant, which indicates that the treatment groups follow trajectories that are generally parallel in the first year of the study. The only disturbance observed among the DinD estimates is the significant effect at eight months for the comparison of the daycare groups. We give little credibility to this effect since it is not predicted in principle and is unstable over time.

Adjusted Model

The baseline effect reported in Table 9.2 represents the relative position of the groups for the pre-intervention period after adjusting for the covariates. As with the initial model, a positive value indicates a deficit for the Program Daycare group while a negative value indicates an advantage. The regression model with controls for confounding variables did not affect the size of the deficit for the Program Daycare group in relation to the comparison groups. However, the Wald statistic is significant after the statistical adjustment, which indicates that the treatment groups follow developmental trajectories that are overall non-parallel in the first year of the study. Note that two DinD effects are significant, namely those at 8 and 12 months, for comparison of the daycare groups. This effect was not anticipated since the tested program did not target the Physical domain.

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance with the Physical scale is expressed as the probability of being classified in a higher quintile: the success category. The analysis results partially reproduce the DinD effects observed in the adjusted model for the comparison of the daycare groups. The odds ratio is 0.76 (90% CI from 0.49 to 1.20), 0.53 (90% CI from 0.33 to 0.84), and 0.52 (90% CI from 0.31 to 0.87) in favour of the Program Daycare group respectively for the DinD effects at 4, 8 and 12 months.

Receptive Vocabulary Scale

Initial Model

The baseline effect reported in Table 9.1 represents the relative position of the groups for the pre-intervention period. Based on the Wald statistic, the trajectory of the three treatment groups diverges significantly in the first year of the study. However, this divergence is not observed at the baseline (see the non-significant DinD effects). The lack of significance found for all assessment periods suggests that the overall effect tested by the Wald statistic is not attributable to an effect of the tested program. Note that Receptive Vocabulary is the scale with the lowest level of internal consistency among all six scales (see Table 3.8).

Adjusted Model

The baseline effect reported in Table 9.2 represents the relative position of the groups for the pre-intervention period after adjusting for the covariates. Based on the Wald statistic, the trajectory of the three treatment groups is generally parallel after statistical adjustment. The DinD estimates calculated based on the two comparison groups indicate a significant effect of 0.20 and of 0.33 standard deviations at the end of the study's first year. This effect may be late in emerging, as observed with the Cognition scale. This observation must be verified in the analyses for the second year of the project.

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance with the Receptive Vocabulary scale is expressed as the probability of being classified in a higher quintile: the success category. The analysis results reproduce the significance of the DinD effects for comparison of the Program Daycare group to the Informal Care group at the 12-month assessment with an odds ratio of 0.65 (90% CI from 0.44 to 0.95). For comparison of the daycare groups, the odds ratio is 0.66 (90% CI from 0.27 to 1.62), 0.40 (90% CI from 0.21 to 0.76), and 0.43 (90% CI from 0.28 to 0.66) respectively for the DinD effects at 4, 8 and 12 months. As with the parametric analysis, only the effects at 8 and 12 months are significant.

Expressive Vocabulary Scale

Initial Model

The baseline effect reported in Table 9.1 represents the relative position of the groups for the pre-intervention period. Based on the Wald statistic, the treatment groups diverge significantly over time. An examination of the DinD estimates shows a significant program effect in relation to the Comparison Daycare group for the assessment periods at 4 and 12 months. The effect appears credible since its size is relatively constant from one assessment to the next, ranging from 0.20 to 0.45 standard deviations. The DinD effect at 4 months is greater than the initial deficit, which indicates a reversal of the groups' relative position with respect to their absolute score on this scale. No significant program effect is observed for the Informal Care group.

Adjusted Model

The baseline effect reported in Table 9.2 represents the relative position of the groups for the pre-intervention period after adjusting for the covariates. As with the initial model, a positive value indicates a deficit for the Program Daycare group while a negative value indicates an advantage. Further to the adjustment made by the inclusion of the covariates, the Wald statistic testing the non-equivalence of the treatment groups' trajectory remains significant. The adjusted model produces significant DinD estimates for all first-year assessments, that is, at 4, 8 and 12 months for the Comparison Daycare group. Inclusion of the covariates appears to have heightened the accuracy of the DinD estimates. However, the size of the effects is not affected, as they still range from about 0.25 to 0.43 standard deviations. No significant program effect is observed in relation to the Informal Care group.

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance with the Expressive Vocabulary scale is expressed as the probability of being classified in a higher quintile: the success category. The analysis results partially reproduce the significance of the DinD effects for comparison of the daycare groups. The odds ratio is 0.43 (90% CI from 0.28 to 0.68), 0.76 (90% CI from 0.41 to 1.42) and 0.60 (90% CI from 0.35 to 1.03) in favour of the Program Daycare group respectively at 4, 8 and 12 months. In other words, the effect at 4 months is confirmed, but that at 12 months is not, although it is just below the significance level. Finally, as with the parametric analysis, no effect is found in comparison with the Informal Care group.

Summary: Analyses by Group

The results of the analyses by treatment group indicate that the tested program had a positive effect on the development of school readiness for children in the Program Daycare group. This conclusion stems from findings of a parametric analysis of the data and the reproduction of these findings using a non-parametric analysis.

In comparison with the baseline, children enrolled in the Program Daycare group experience faster growth on several dimensions of school readiness compared to their peers in the two comparison groups. This fact is supported by the DinD estimates reported in Tables 9.1 and 9.2. The program effect is more marked for the Comparison Daycare group, perhaps due to the relatively homogeneous composition of this comparison group.

Generally, it should be noted that adding the second cohort makes the results more stable over time in comparison with the results in the *First Cohort Findings Report* (2014). For Communication, Self-Awareness and Expressive Vocabulary, the effect of the tested program for the Comparison Daycare group is significant at 4 months and remains significant thereafter, with effects of the order of 0.25 to 0.43 standard deviations. Based on the developmental standards reported by Hill and colleagues (2008), this represents a developmental gain of about 2.3 to 3.9 months for children in the Program Daycare group versus those in the Comparison Daycare group.

Furthermore, the impact analyses in the *First Cohort Findings Report* (2014) did not identify any effects of the tested program for the Informal Care group. This is not the case in these analyses. When children began junior kindergarten, we find a program effect of 0.30, 0.20 and 0.20 standard deviations for Communication, Cognition, and Receptive Vocabulary respectively. In other words, this is a developmental gain of about 1.8 to 2.7 months for children in the Program Daycare group versus those in the Informal Care group.

9.2.3. Analyses by Dosage

This section presents the analyses based on dosage. Note that these analyses concern only the treatment groups attending daycare. The results for the initial model are presented in Table 9.3 and those of the adjusted model, in Table 9.4. The dosage variable (average number of hours spent at daycare per week) was standardized for analysis purposes. Thus, a value of zero

represents the sample mean (i.e., mean-centred). The reported models include only the linear effect of dosage.⁶⁸

The results for each scale are described first for the initial model, and then for the adjusted model. The non-parametric reproduction of the adjusted model's results is then presented. Finally, the section ends with a summary of the results for the analyses by dosage. Note that 98.3% of the daycare groups sample attended daycare for over 10 hours per week.

Communication Scale

Initial Model

For the entire sample, the number of hours spent at daycare is positively associated with performance on the Communication scale (see Table 9.3). The estimated effect of dosage at the baseline indicates that, for this sample, an increase of 10 hours per week corresponds to an average increase of about 0.20 standard deviations for Communication.

We estimated the instability of this effect over time (see the dosage effects by time) and found that the baseline value for the dosage effect applies at both 8 and 12 months. Exceptionally, we found that the dosage effect is stronger at 4 months for this sample (see the dosage effects by time).

The model also enables us to estimate the program impact based on level of dosage. The results reported in Table 9.3 indicate that dosage affects the program impact only at 4 months. For this assessment, the program impact would be about 0.10 standard deviations (non-significant) for a child who spends an average of 17 hours per week at daycare, about 0.30 standard deviations for a dosage of 27 hours per week (significant), and about 0.50 standard deviations when dosage reaches 37 hours per week (significant).^{69,70} Due to the small number of children participating in the study, this estimate of the critical dosage threshold for a significant program impact is not very accurate.

Adjusted Model

The pattern of results for the adjusted model differs from that of the initial model in that the dosage effect lacks significance at the baseline (see Table 9.4). The dosage effect for the Program Daycare group emerges at 4 months, remaining stable for the assessment periods at 8 and 12 months (see the dosage effects by time in Table 9.4). Based on the model, a child who

⁶⁸ The relationship between dosage and development can be linear or non-linear. A linear relationship holds if the dosage effect is constant regardless of the intensity of exposure to the childcare environment (e.g., 10 hours a week versus 40 hours a week). Preliminary analyses (e.g., using quadratic terms; inspection of residuals) did not reveal a non-linear relationship in the adjusted model. The conclusion to be drawn from this result is that the linear dosage effect suffices for the purposes of the analysis in this report, which is coherent with the literature indicating that dosage effects are linear (National Institute of Child Health and Human Development Early Child Care Research Network and Duncan, 2003).

⁶⁹ The sample mean is about 27 hours per week. A dosage of 10 hours less than this weekly average would be 17 hours per week. A dosage of 10 hours more than this weekly average would be 37 hours per week.

⁷⁰ Although the effect of the group is not reported in Table 9.3, it was nevertheless estimated in the dosage model. The value of 0.30 is drawn from the initial regression model for dosage and represents the program impact for the Comparison Daycare group at 4 months for a child with average dosage (27 hours per week). The estimates reported in the text are calculated by adding the effect of the significant three-way interaction among time, group and dosage reported in Table 9.3 to the DinD effect of the corresponding program (not reported). Adding dosage to the model may affect the estimate of the program effect, which means the estimated values may well not agree with estimates from a model without dosage.

spends 10 hours more at daycare experiences an increase of 0.20 standard deviations or a gain of 1.8 months in his or her development. The DinDinD effects at 4 and 12 months indicate that this positive DinD effect for dosage is observed only among children in the Program Daycare group.

Otherwise stated, the model indicates that the size of the program effect for Communication varies by dosage. When a child begins junior kindergarten (at 12 months), the model predicts a program impact of 0.20 standard deviations for a child whose dosage is 20 hours per week at daycare, 0.40 standard deviations for a child whose dosage is 30 hours per week, and 0.60 standard deviations for a child whose dosage is 40 hours per week (approximate values). It is important to note that these estimates properly reflect what is observed on average for this sample, but their accuracy is too low to conclude that they accurately reflect the parameters of the population of minority Francophone children attending French-language daycare.

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance for the Communication scale is expressed as the probability of being classified in a higher quintile: the success category. The effects to be reproduced are the DinDinD estimates testing the differentiation of the dosage effect by treatment group.

For the Comparison Daycare group, the odds ratio for the DinDinD effect is 0.96 (90% CI from 0.91 to 1.01), 1.00 (90% CI from 0.94 to 1.06) and 0.94 (90% CI from 0.89 to 0.99) in favour of the Program Daycare group respectively for the 4, 8 and 12 months assessment periods. Specifically, the dosage effect for children in the Program Daycare group is significantly more pronounced than for children in the Comparison Daycare group for the 12 months assessment. The effect at 4 months is just below the 10% significance level. In short, the results of the non-parametric analyses reproduce the differentiation effect from the dosage effect when children begin junior kindergarten (i.e., the 12-month assessment). Based on the model, the odds ratio for the program effect increases 6% for every hour in excess of the average spent at daycare.

Table 9.3: Unadjusted Differentiation of the Effect of the Number of Hours Spent at Daycare (Dosage) by Program Received (First Year) — Standardized Scores

Type of Difference	School Readiness Indicator											
	Communication		Self-Awareness		Cognition		Physical		Receptive Vocabulary		Expressive Vocabulary	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Dosage												
Baseline	0.019*	0.010	0.009	0.011	-0.010	0.013	-0.001	0.017	0.016*	0.008	0.004	0.009
Dosage x group												
G1 vs. G2 (DinD)	0.000	0.015	0.004	0.016	0.020	0.017	0.002	0.018	-0.008	0.011	0.008	0.014
Dosage x time												
4 months (DinD)	0.020***	0.006	0.012	0.002	0.020***	0.004	0.002	0.013	0.010	0.006	0.018***	0.004
8 months (DinD)	0.004	0.018	0.002	0.011	0.003	0.012	-0.014	0.019	0.009	0.008	-0.003	0.013
12 months (DinD)	0.006	0.009	0.017	0.009	0.027*	0.009	-0.007	0.015	0.006	0.010	0.017**	0.008
Dosage x time x group												
4 months (DinDinD)	-0.021*	0.011	-0.015	0.010	-0.026*	0.011	-0.010	0.016	-0.003	0.012	-0.023**	0.010
8 months (DinDinD)	-0.007	0.022	-0.014	0.018	-0.015	0.018	0.001	0.023	-0.011	0.014	-0.003	0.021
12 months (DinDinD)	-0.013	0.018	-0.031	0.019	-0.054***	0.017	-0.026	0.020	-0.014	0.015	-0.029	0.019
Dosage x Time Wald F	1.38		0.57		0.56		2.39		1.43		0.78	
Dosage x Time x Group Wald F	1.19		1.36		5.14*		2.97*		0.33		2.07	

Note: The degrees of freedom for the Wald F-tests are 6 and 19. The Huber-White robust standard errors are reported with error terms clustered by daycare. *The negative values of DinD estimates for all group comparisons are a positive treatment effect (i.e., an advantage for the Program Daycare group).* Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Table 9.4: Adjusted Differentiation of the Effect of the Number of Hours Spent at Daycare (Dosage) by Program Received (First Year) — Standardized Scores

Type of Difference	School Readiness Indicator											
	Communication		Self-Awareness		Cognition		Physical		Receptive Vocabulary		Expressive Vocabulary	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Dosage												
Baseline	-0.001	0.008	-0.007	0.008	-0.006	0.008	0.004	0.010	0.005	0.007	-0.011	0.007
Dosage x group												
Baseline	0.001	0.009	0.005	0.009	0.012	0.012	-0.002	0.010	-0.012	0.008	0.009	0.009
Dosage x time												
4 months (DinD)	0.019***	0.006	0.015***	0.004	0.011**	0.004	0.005	0.015	0.018***	0.003	0.017*	0.008
8 months (DinD)	0.019**	0.009	0.015**	0.007	-0.002	0.007	-0.009	0.019	0.021***	0.006	0.002	0.008
12 months (DinD)	0.024***	0.008	0.029***	0.009	0.036***	0.010	0.003	0.011	0.017**	0.007	0.031***	0.010
Dosage x time x group												
4 months (DinDinD)	-0.025***	0.009	-0.022**	0.010	-0.025**	0.012	-0.005	0.017	-0.018**	0.008	-0.025**	0.010
8 months (DinDinD)	-0.013	0.013	-0.018	0.013	-0.009	0.016	0.003	0.020	-0.016	0.011	0.000	0.014
12 months (DinDinD)	-0.019*	0.011	-0.033**	0.013	-0.054***	0.015	-0.030**	0.014	-0.009	0.011	-0.033**	0.013
Hours by Time Wald F (3, 16)	1.73		0.99		2.06		1.05		2.59*		1.49	
Hours by Group by Time Wald F (3, 16)	2.83*		2.96*		19.93***		6.24***		2.03		4.06**	

Note: The Huber-White robust standard errors are reported with error terms clustered by daycare. *The negative values of DinD estimates for all group comparisons are a positive treatment effect (i.e., an advantage for the Program Daycare group).* Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Self-Awareness Scale

Initial Model

The dosage effect is significant neither at the baseline nor at later assessments and does not vary significantly by treatment group.

Adjusted Model

The adjusted model does not generate the same pattern of results as the initial model. The dosage effect, which is null at baseline, increases significantly for all children over time based on the DinD estimates (see dosage by time). Note that based on the DinDinD estimates, the dosage effect is significantly stronger for the Program Daycare group at 4 months and at 12 months for the Program Daycare group. At 12 months, the effect is of the order of 0.30 standard deviations per 10 hours at daycare, after controlling for covariates.

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance for the Self-Awareness scale is expressed as the probability of being classified in a higher quintile: the success category. The effects to reproduce are the DinDinD estimates testing the differentiation of the dosage effect by group.

For the Comparison Daycare group, the odds ratio for the DinDinD effect is 0.93 (90% CI from 0.88 to 0.98), 0.95 (90% CI from 0.90 to 1.01) and 0.93 (90% CI from 0.86 to 0.99) in favour of the Program Daycare group respectively for the assessment periods at 4, 8 and 12 months. The analysis results partially reproduce the significance of the DinDinD effects at 4 and 12 months. When children begin junior kindergarten, the model estimates that the odds ratio for program impact increases for the Program Daycare group by 7% for each hour in excess of the weekly average spent at daycare. Conversely, the program impact diminishes by 7% for each hour less than the weekly average spent at daycare.

Cognition Scale

Initial Model

The direct effect of hours spent at daycare is not significant at the baseline but increases significantly and positively for all participants at 4 months and 12 months based on the DinD estimator. The DinDinD estimator serves to specify that at 4 and 12 months, the positive dosage effect is observed only for the Program Daycare group.

Adjusted Model

The adjusted model generates essentially the same pattern of results as the initial model. The direct effect of hours spent at daycare is not significant at the baseline but increases significantly and positively for the Program Daycare group at 4 and 12 months. At 12 months, the effect is very pronounced, representing an advantage of 0.50 standard deviations (about 4.5 months of development) per additional 10 hours at a program daycare.

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance for the Cognition scale is expressed as the probability of being classified in a higher quintile: the success category. The effects to reproduce are the DinDinD estimates testing the differentiation of the dosage effect by group.

For the Comparison Daycare group, the odds ratio for the DinDinD effect is 0.96 (90% CI from 0.92 to 1.01), 1.00 (90% CI from 0.94 to 1.06) and 0.90 (90% CI from 0.86 to 0.96) respectively for the assessments at 4, 8 and 12 months. The analysis results partially reproduce the significance of the observed DinDinD effects, but at 12 months only. The model estimates that the odds ratio for program impact increases by 10% for each hour in excess of the average spent at daycare.

Physical Scale

Initial Model

No significant dosage effect is observed, either directly or through interaction with group and/or time.

Adjusted Model

The adjusted model does not generate the same pattern of results as the initial model. The only observed effect is the emergence of a negative dosage effect at 12 months for the Comparison Daycare group (see the DinDinD effects in Table 9.4). We give little credibility to this result because the effect is low, was not anticipated, and appears specific to the 12-month assessment period.

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance for the Physical scale is expressed as the probability of being classified in a higher quintile: the success category. The effects to reproduce are the DinDinD estimates testing the differentiation of the dosage effect by treatment group.

For the Comparison Daycare group, the odds ratio for the DinDinD effect is 0.99 (90% CI from 0.94 to 1.05), 0.99 (90% CI from 0.91 to 1.08) and 0.90 (90% CI from 0.84 to 0.96) respectively for the assessments at 4, 8 and 12 months. The analysis results reproduce the significance of the DinDinD effects observed at 12 months. The model estimates that the odds ratio associated with the program impact increases by 10% for each hour in excess of the weekly average spent at daycare.

Receptive Vocabulary Scale

Initial Model

According to the DinD estimator, a positive dosage effect is observed at the baseline. Moreover, this effect does not vary significantly based on time, treatment group, or the combination of time and treatment group.

Adjusted Model

The adjusted model produces a null dosage effect at the baseline, emerging as significant, based on the DinD estimator, at 4 months and then remaining stable. Based on the DinDinD estimator, the positive dosage effect is significant in favour of the Program Daycare group only for the 4-month assessment period. There is no significant difference in the dosage effect at 8 and 12 months. Other than the significant effect at 4 months, the dosage effect for Receptive Vocabulary does not vary significantly by type of program received.

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance for the Receptive Vocabulary scale is expressed as the probability of being classified in a higher quintile: the success category. The effects to reproduce are the DinDinD estimates testing the differentiation of the dosage effect by group.

For the Comparison Daycare group, the odds ratio for the DinDinD effect is 0.91 (90% CI from 0.86 to 0.96), 0.93 (90% CI from 0.87 to 0.99) and 0.95 (90% CI from 0.90 to 0.99) respectively for the assessment periods at 4, 8 and 12 months. These results partially reproduce the significance of the DinDinD effects observed at 4 months with a new significant effect detected at the 12-month assessment period. This pattern of results is an indication that the parametric analysis is less sensitive for this assessment period. The model estimates that the odds ratio for program impact increases 5% for each hour spent at daycare in excess of the average for the daycare group in the period prior to the 12-month assessment. The opposite is also estimated using the odds for program impact. Program effects diminish by 5% for each hour less spent at daycare than the weekly average for children in daycare in the period prior to the 12-month assessment.

Expressive Vocabulary Scale

Initial Model

The observed effects based on the DinD estimator indicate a positive dosage effect emerging at 4 and 12 months. However, the DinDinD effects are significant only at 4 months in favour of children in the Program Daycare group. Together, these results suggest that the dosage effect is statistically equal for both groups at 12 months.

Adjusted Model

The adjusted model reproduces the results of the initial model. According to the DinDinD effects, we observe significant positive dosage effects emerging at the 4- and 12-month assessments in favour of the Program Daycare group. In other words, the size of the program effect depends on dosage. Based on the model, 10 additional hours per week at a program daycare would result in an increase of about 0.30 standard deviations or a gain of 2.7 months in the development of expressive vocabulary (using the data of Hill and colleagues, 2008, as a benchmark). This result is not obtained for children in the Comparison Daycare group, for whom dosage has no significant distinguishable effect.

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance for the Expressive Vocabulary scale is expressed as the probability of being classified in a higher quintile: the success category. The effects to reproduce are the DinDinD estimates testing the differentiation of the dosage effect by group.

For the Comparison Daycare group, the odds ratio for the DinDinD effect is 0.92 (90% CI from 0.88 to 0.96), 1.00 (90% CI from 0.94 to 1.07) and 0.92 (90% CI from 0.86 to 0.99) in favour of the Program Daycare group respectively for the assessments at 4, 8 and 12 months. These results partially reproduce the significance of the DinDinD effects observed with the adjusted model for the assessment periods at 4 and 12 months. The model indicates that for these periods, the odds ratio for program effect increases by 8% for each hour spent at daycare in excess of the average for children in daycare. The opposite is also estimated using the odds for program impact. The program effects diminish by 8% for each hour spent at daycare less than the average for children attending daycare at the 4- and 12-month assessment periods.

Summary: Analyses by Dosage

The data analysis for the first cohort described in the *First Cohort Findings Report* (2014) shows a differentiated dosage effect only for the Communication domain and only for the 4-month assessment ("1st post-test" in *First Cohort Findings Report* terminology). Adding the second cohort to the sample significantly increases the numbers for both daycare groups (now at over 100 per group), which increases the fidelity of dosage effect estimates for each group. Adding the second cohort created stability, which in turn allows us to now detect dosage effects (as observed in literature on early childhood development) and, more interesting still, effects differentiated by group.

For the Communication, Self-Awareness, Cognition and Expressive Vocabulary domains, an enhanced program effect is observed when children begin junior kindergarten (i.e., the 12-month assessment period) of about 0.30 standard deviations for every 10 hours in excess of the weekly average spent at daycare ($27 + 10 = 37$ hours per week). In other words, this effect is added to the effect of membership in the Program Daycare group observed in the analysis by group (see the effects at 12 months reported in Table 9.2). For Communication, the program effect for children receiving a high dosage would be about 0.70 standard deviations for Self-Awareness and 0.50 standard deviations for Expressive Vocabulary. Note that the program effect is symmetric. Thus, the program effect at 12 months diminishes by about 0.30 standard deviations for every 10 hours less than the weekly average spent at daycare ($27 - 10 = 17$). Otherwise expressed, the effect of the tested program is nil for children who receive a relatively low dosage. The pattern of these results suggests that the program impact stems mainly from the Daycare Program.

In short, the results suggest that about 25 to 30 hours of daycare per week are necessary to obtain a significant program impact. Further, a greater impact is observed for children who spend even more time at daycare (about 40 hours per week). According to the developmental standards of Hill and colleagues (2008), the accentuation of the program effect associated with another 10 hours per week above the weekly average would be roughly equivalent to 2.7 months of development.

9.2.4. Analyses by Daycare Program Quality and Fidelity

This section presents the analyses examining how the quality and fidelity of the Daycare Program affect child development. The sections below present the results for three series of analyses: (a) a quantitative verification of treatment group differentiation based on their daycare program, (b) a verification of the predictive value of key dimensions of the tested program on the fidelity and quality of implementation, and (c) a mediation analysis verifying whether implementation of the Daycare Program is responsible for all or part of the program effects detected by the analyses by treatment group (reported in Table 9.2). The objective of these analyses is to estimate the relative importance of the Daycare Program in determining the tested program's observed effects in the first year of implementation. Note that in these analyses, the fidelity/quality data describing the program delivered to second-cohort children are treated as though they were from different daycares.

Differentiation of the Daycare Program

This section presents the results of analyses examining the differentiation of daycare programs at program daycares and comparison daycares. Note that the statistical power of these tests is not low.⁷¹ It is therefore important to consider the size of the effects when interpreting the results. Table 9.5 shows the quality analysis, while Table 9.6 shows the fidelity analysis. A parametric (t-test) and non-parametric (Wilcoxon rank-sum z-test) test is reported for each index. To determine the unique contribution of these indices to upcoming analyses, we also report their inter-correlations in Table 9.7.

Quality

The analyses show that program daycares differ from comparison daycares for two dimensions of quality: Structural Quality and Reading Quality. The size of these effects is large, exceeding 0.75 standard deviations, and is reproduced by the non-parametric analysis. Educative Quality ranks third in order of importance with a differentiation of 0.68 standard deviations. Finally, differentiation between program daycares and comparison daycares is negligible in terms of Educator Sensitivity, at 0.26 standard deviations.

Fidelity

The analysis of differentiation for program fidelity indicates that daycares in the Program Daycare group distinguish themselves mainly with respect to Structural Fidelity ($ds > 2$). Differentiation for Content Fidelity is nevertheless considerable, ranging from 0.61 to 1.20 standard deviations. The differentiation at 8 months was most robust for this index.

⁷¹ The probability of detecting an effect of 0.50 standard deviations is less than two out of five times for this analysis.

Table 9.5: Differentiation of the Tested Program's Quality

Index	Program Daycares (N=6)	Comparison Daycares (N=17)	Differentiation (d^P effect size)	Results	
	M (SE)	M (SE)		Robust t-test (d.f.)	Rank z-test
Structural Quality	6.51 (0.48)	5.71 (1.02)	+0.80 (0.78)	2.56** (18)	1.83*
Educative Quality	5.33 (0.88)	4.49 (1.24)	+0.84 (0.68)	1.81* (12)	1.63
Educator Sensitivity	5.92 (1.86)	5.29 (2.41)	+0.62 (0.26)	0.65 (11)	0.62
Reading Quality	5.50 (1.34)	2.15 (1.46)	+3.35 (2.29)	5.15*** (9)	3.36***

Note: Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$.^a The d value, the effect's standardized size, is calculated using the standard deviation of the Comparison Daycare group. M = mean, SE = standard deviation, d.f. = degrees of freedom.

Table 9.6: Differentiation of the Tested Program's Fidelity

Structural Fidelity	Program Daycares (N=6)	Comparison Daycares (N=17)	Differentiation (d^P effect size)	Results	
	M (SE)	M (SE)		Robust t-test (d.f.)	Rank z-test
Baseline	0.77 (0.13)	0.44 (0.14)	+0.33 (2.36)	5.29*** (9)	3.35***
4 months	0.87 (0.14)	0.44 (0.14)	+0.43 (3.07)	6.47*** (9)	3.46***
8 months	0.85 (0.11)	0.45 (0.19)	+0.41 (2.16)	6.56*** (15)	3.37***
Content Fidelity	M (SE)	M (SE)	Differentiation (d^I effect size)	Robust t-test (d.f.)	Rank z-test
Baseline	0.75 (0.07)	0.63 (0.18)	+0.11 (0.61)	2.17** (20)	0.84
4 months	0.79 (0.10)	0.64 (0.18)	+0.15 (0.83)	2.45** (15)	1.54
8 months	0.83 (0.08)	0.65 (0.15)	+0.18 (1.20)	3.61** (15)	2.42**

Note: Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$.^a The d value, the effect's standardized size, is calculated here based on the standard deviation of the Comparison Daycare group. M = mean, SE = standard deviation, d.f. = degrees of freedom.

Correlations Among Quality and Fidelity Indices

Table 9.7 lists the descriptive statistics for the entire sample as well as Pearson correlations for each pair of quality/fidelity indices considered in the impact analyses. Note that there is only a 0.23 probability of detecting a real correlation of 0.20 in this analysis, that is, once out of five times. However, a real correlation of 0.50 would be detected four out of five times, corresponding to a probability of 0.80.

The sample of 23 daycares allows us to present a descriptive analysis of the observed correlations. First, it is clear that the information captured by these indices is not summarized by a single factor. Indices appear to be broken down between two factors at least, one centred on Educative Quality and the other, on Reading Quality. Educative Quality is significantly correlated with Structural Quality, Structural Fidelity, Content Fidelity and Educator Sensitivity. In other words, each of these indices shares variance with Educative Quality, which suggests that

they have a common source. Reading Quality is significantly correlated with Structural Fidelity and Content Fidelity, which suggests yet again a common source of variance.

Table 9.7: Inter-correlations of the Daycare Program’s Quality and Fidelity Indices

Index	Descriptive Statistics		Pearson Correlations				
	M	SE	1	2	3	4	5
1. Structural Quality	5.92	0.97					
2. Educative Quality	4.71	1.19	0.51***				
3. Educator Sensitivity	5.46	2.26	0.17	0.53***			
4. Reading Quality	3.02	2.05	0.18	0.16	0.17		
5. Structural Fidelity ^a	0.55	0.25	0.23	0.45**	0.34	0.68***	
6. Content Fidelity ^a	0.70	0.16	0.23	0.50***	0.28	0.76***	0.62***

Note: N = 23 daycares. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. ^a Index from observations made in the period leading up to the 8-month assessment period. M = mean, SE = standard deviation.

Impact of Fidelity and Quality

The results of the analyses by quality and by fidelity are presented respectively in Tables 9.8 and 9.9. We describe the analysis results for each index separately in the paragraphs that follow.

Structural Quality

At the baseline, the relationship is significantly negative between Structural Quality and two outcomes, Cognition and Receptive Vocabulary. We then find, when children begin junior kindergarten (i.e., the 12-month assessment period), higher gains (DinD) for daycares with better Structural Quality for Communication, Cognition, Receptive Vocabulary and Expressive Vocabulary. The size of this DinD effect is approximately 0.20 standard deviations for each quality “point”, which is more than enough to reverse the negative relationship observed at the baseline for certain outcomes.

Educative Quality

We observe a negative relationship at the baseline between Educative Quality and the mean score for the Communication, Self-Awareness, and Expressive Vocabulary scales. This indicates that high-quality daycares tend to care for children who experience difficulties with these outcomes. The DinD effects indicate that this association corrects itself since the gains in relation to the baseline for Communication and Expressive Vocabulary are positively associated with Educative Quality at the 12-month assessment period. Effects are observed earlier for Self-Awareness (the 4-month assessment period) and Cognition (the 4- and 8-month assessment periods). Note that the DinD effects are non-significant for the Physical and Receptive Vocabulary scales. According to the reported model, a one-point difference in Educative Quality in preschool is associated with a gain of 0.12 standard deviations for Communication and Expressive Vocabulary when children begin junior kindergarten (12-month assessment period).

Educator Sensitivity

Educator Sensitivity is not significantly associated with any outcome at the baseline. The DinD effects indicate that the developmental gains observed at the following assessments are associated with Educator Sensitivity only for Expressive Vocabulary and only at the 12-month assessment period. According to the reported model, a one-point increase in sensitivity in preschool is associated with a gain of 0.17 standard deviations in Expressive Vocabulary when children begin junior kindergarten.

Reading Quality

At the baseline, we observe a significant negative relationship between Reading Quality and the mean score obtained for the Communication, Physical, and Expressive Vocabulary scales. When children begin junior kindergarten (12-month assessment period), we note a positive effect only for the Physical domain. For the Communication, Self-Awareness, Cognition, and Expressive Vocabulary outcomes, a positive and significant effect on developmental gains is found for the 4-month assessment period, although the significance of this effect disappears for the other assessment periods.

Overall Quality

At the baseline, we observe a negative relationship between Overall Quality and all school readiness indicators except Receptive Vocabulary. The DinD estimates indicate that when children begin junior kindergarten (12-month assessment period), the overall quality of the program is positively associated with the developmental gains of children who attended daycare for Communication and Expressive Vocabulary. According to the presented model, a one-point increase on the Overall Quality scale represents an increase of about 0.13 standard deviations for these two outcomes.

Structural Fidelity

At the baseline, we observe a negative association between Structural Fidelity and the mean score for the Communication, Self-Awareness, Cognition, and Expressive Vocabulary scales. Significant gains are associated with the program's Structural Fidelity for Communication at each assessment period in the first year (see the DinD effects). When children begin junior kindergarten (12-month assessment period) and according to proposed model, we estimate a difference of 0.30 standard deviations for the Communication score between a daycare with a 40% fidelity score and a daycare with an 80% fidelity score.⁷² No significant DinD effects associated with other school readiness indicators are observed at any assessment period.

Content Fidelity

At the baseline, we observe a significant negative association between Content Fidelity and the mean score for the Communication scale. The DinD estimates show a single significant positive program effect: at the 4-month assessment period for Expressive Vocabulary. Overall, the impact analyses for Content Fidelity tend to invalidate the hypothesis that program content is

⁷² The values reported in the tables are proportions. Here, the same variable is expressed as a percentage. Not that a fidelity score of 0.80 is equivalent to a fidelity score of 80%.

an important factor in the development of Francophone preschoolers. If the program has a significant effect on development, it occurs through another element of the program.

Overall Fidelity

At the baseline, we observe significant negative associations between Overall Fidelity and the mean score for the Communication, Self-Awareness, and Expressive Vocabulary scales. The DiD estimators show that Overall Fidelity has a significant and stable effect on developmental gains in Communication. According to the proposed model, a 40% to 80% increase in Overall Fidelity is associated with an increase of about 0.32 standard deviations for Communication.

Non-Parametric Cross-Validation

Since analyses of quality and fidelity effects are secondary in importance, this report does not present a detailed account of the cross-validation for these analyses. In general, the effects associated with fidelity indices are very robust, as they are confirmed by the non-parametric analysis. However, the effects associated with quality indices are less robust. Among the latter, the Reading Quality index was the most robust predictor while Receptive Vocabulary was the outcome most often associated with the quality indices.

Table 9.8: Effect of Daycare Program Quality with Adjustment (First Year) — Standardized Scores

Type of Difference	School Readiness Indicator											
	Communication		Self-Awareness		Cognition		Physical		Receptive Vocabulary		Expressive Vocabulary	
Structural Quality	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.16	0.09	-0.07	0.08	-0.13*	0.06	-0.10	0.08	-0.19***	0.05	-0.11	0.10
4 months (DinD)	0.18	0.11	0.16***	0.05	0.12	0.09	0.17**	0.06	0.16*	0.08	0.24**	0.10
8 months (DinD)	0.13*	0.07	0.06	0.07	0.06	0.10	0.06	0.06	0.17***	0.05	0.10*	0.05
12 months (DinD)	0.24*	0.12	0.12	0.07	0.15**	0.06	0.11	0.07	0.23***	0.07	0.22*	0.11
<i>Group x Time Wald F</i>	1.33		2.98*		3.86**		2.73*		3.63**		1.92	
Educative Quality	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.11**	0.04	-0.09*	0.05	-0.13***	0.04	-0.02	0.05	-0.07	0.05	-0.12**	0.04
4 months (DinD)	0.09*	0.04	0.11***	0.03	0.08*	0.04	0.05	0.07	-0.05	0.06	0.22***	0.04
8 months (DinD)	0.06	0.06	0.01	0.06	0.11*	0.06	0.05	0.08	0.02	0.06	0.02	0.07
12 months (DinD)	0.12*	0.06	0.06	0.06	0.06	0.04	0.05	0.06	0.07	0.07	0.12*	0.06
<i>Group x Time Wald F</i>	1.38		6.32***		1.14		0.38		1.47		9.71***	
Educator Sensitivity	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.03	0.03	-0.01	0.02	-0.03	0.02	-0.01	0.03	-0.01	0.02	-0.04	0.03
4 months (DinD)	0.05*	0.02	0.04**	0.01	0.00	0.02	0.02	0.04	0.01	0.04	0.09**	0.03
8 months (DinD)	0.02	0.05	-0.02	0.02	-0.03	0.02	0.01	0.03	-0.02	0.03	0.01	0.03
12 months (DinD)	0.05	0.05	0.01	0.02	-0.03	0.02	-0.03	0.03	0.01	0.03	0.07*	0.03
<i>Group x Time Wald F</i>	5.00**		2.91*		0.61		0.76		0.50		3.82**	
Reading Quality	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.03*	0.01	-0.02	0.01	-0.02	0.02	-0.05*	0.02	0.01	0.02	-0.04**	0.01
4 months (DinD)	0.04**	0.02	0.06***	0.01	0.03***	0.01	-0.01	0.04	-0.03	0.04	0.09***	0.02
8 months (DinD)	0.01	0.02	0.03	0.02	-0.01	0.02	0.05	0.03	0.02	0.02	0.01	0.03
12 months (DinD)	0.01	0.02	0.01	0.02	0.01	0.01	0.07***	0.01	0.01	0.02	0.01	0.02
<i>Group x Time Wald F</i>	2.34		11.12***		7.75***		7.80***		1.18		5.68***	
Overall Quality	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.13***	0.04	-0.10*	0.05	-0.12**	0.04	-0.09*	0.04	-0.04	0.04	-0.15**	0.06
4 months (DinD)	0.17***	0.04	0.20***	0.03	0.10**	0.04	0.04	0.09	-0.03	0.07	0.32***	0.05
8 months (DinD)	0.07	0.07	0.06	0.06	0.01	0.05	0.10*	0.05	0.04	0.05	0.04	0.08
12 months (DinD)	0.14*	0.07	0.09	0.05	0.02	0.04	0.10	0.06	0.08	0.05	0.13*	0.07
<i>Group x Time Wald F</i>	6.45***		12.51***		2.99*		1.56		2.14		11.00***	

Note: Contrary to the analyses by group, positive values for DinD effects indicate a positive relationship between the quality element and the school readiness indicator. The baseline serves as a benchmark in the analyses. The degrees of freedom for the Wald F-test are 3 and 15. Overall Quality is the average of the four quality indices. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$; ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Table 9.9: Effect of Daycare Program Fidelity with Adjustment (First Year) – Standardized Scores

Type of Difference	School Readiness Indicator											
	Communication		Self-Awareness		Cognition		Physical		Receptive Vocabulary		Expressive Vocabulary	
Structural Fidelity	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.64***	0.18	-0.50***	0.16	-0.40*	0.21	-0.40	0.30	-0.12	0.22	-0.71***	0.18
4 months (DinD)	0.77***	0.22	0.76***	0.13	0.35*	0.18	0.01	0.39	-0.26	-0.26	1.13***	0.21
8 months (DinD)	0.74**	0.27	0.60**	0.28	0.19	0.23	0.66**	0.25	0.39	0.26	0.57	0.33
12 months (DinD)	0.74**	0.30	0.32	0.20	0.24	0.19	0.49*	0.24	0.29	0.25	0.40	0.29
<i>Group x Time Wald F</i>	3.62**		14.15***		1.41		3.52**		3.02*		11.62***	
Content Fidelity	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.53**	0.23	-0.28	0.29	-0.17	0.30	0.15	0.26	0.04	0.32	-0.58	0.40
4 months (DinD)	0.40	0.27	0.59	0.37	0.16	0.38	-0.36	0.79	-10.05*	0.52	10.63***	0.52
8 months (DinD)	0.03	0.27	-0.06	0.33	-0.40	0.51	0.01	0.29	-0.05	0.25	-0.29	0.45
12 months (DinD)	0.09	0.28	-0.23	0.35	-0.29	0.45	0.35	0.55	-0.01	0.38	-0.24	0.36
<i>Group x Time Wald F</i>	0.78		2.34		0.68		0.24		1.52		8.07***	
Overall Fidelity	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.83***	0.27	-0.63**	0.24	-0.46	0.30	-0.43	0.34	-0.11	0.30	-0.91***	0.29
4 months (DinD)	0.92**	0.33	0.99***	0.21	0.41	0.28	-0.03	0.59	-0.56	0.49	1.65***	0.31
8 months (DinD)	0.74*	0.38	0.60	0.40	0.08	0.38	0.70**	0.31	0.34	0.28	0.49	0.49
12 months (DinD)	0.77*	0.37	0.30	0.29	0.17	0.29	0.68**	0.31	0.27	0.30	0.36	0.37
<i>Group x Time Wald F</i>	2.37		9.39***		1.15		2.75*		3.02*		8.81***	

Note: Contrary to the analyses by group, positive values for DinD effects indicate a positive relationship between the quality element and the school readiness indicator. The baseline serves as a benchmark in the analyses. The degrees of freedom for the Wald F-test are 3 and 15. Overall Fidelity is the average of the Structural Fidelity and Content Fidelity scores. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$, ** $p < 5\%$, * $p < 10\%$. Diff = difference, SE = standard error.

Mediator Effect of Fidelity and Quality

This section presents the results of a mediation analysis intended to determine the extent to which the fidelity and quality of the Daycare Program are responsible for the tested program's effects. The analyses conducted examine whether all or part of the program effect can be attributed to the quality/fidelity indices. If not, it must be concluded that other aspects of the program are responsible for the observed effects on school readiness indicators (e.g., the Family Workshop Program).

The group variable serves as the exogenous explanatory factor (X), quality and fidelity as endogenous mediating variables (M), and school readiness variables as outcomes (Y). A mediation analysis is based on three parameters and their standard errors: coefficient a captures the relationship between X and M, coefficient b captures the relationship between M and Y, and coefficient c' captures the residual effect of X on Y after eliminating the variance associated with the M variable. Figure 9.1 below shows a diagram illustrating the logic of this analysis (see also Krull and MacKinnon, 1999).

Figure 9.1: Diagram Illustrating the Two Aspects of the Program Effect: The Effect Attributable to Quality and Fidelity of the Daycare Program' (a , b) and That not Attributable to Quality and Fidelity (c').

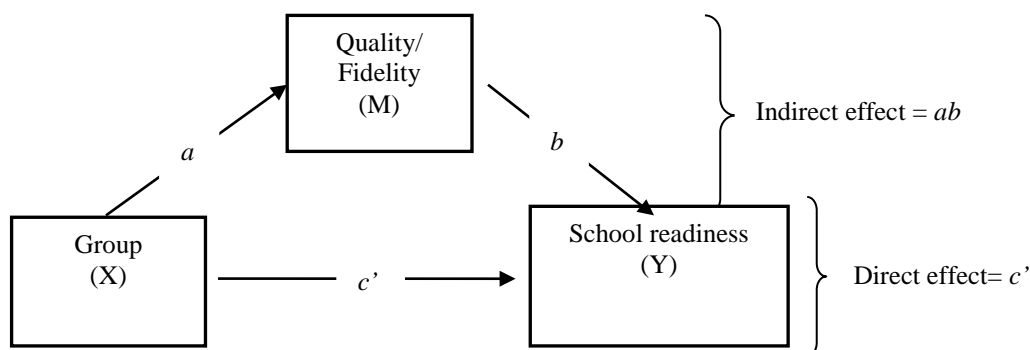


Figure 9.1 depicts two effects that are of interest. The program's *indirect effect* captures that part of the program effect attributable to Daycare Program quality/fidelity. This effect is estimated by introducing the "fidelity" variable into the regression model. The part of the group effect eliminated in the adjusted model by including fidelity is attributed to this element of the tested program. The group's residual effect is the program's *direct effect*, i.e., that part of the effect that can be attributed to other program aspects. If the latter effect is non-significant, then it is assumed that the entire program effect is due to program quality/fidelity. However, if the effect is significant, then it is assumed that a real part of the program effect has yet to be explained (e.g., by the Family Workshop Program). To simplify the presentation, we report only the mediation analysis at 12 months when the children begin junior kindergarten and only for the

quality/fidelity indices that most strongly predict child development. The results are reported in Tables 9.10 and 9.11 for quality and fidelity respectively.⁷³

Quality

The Structural Quality index was most powerful in predicting child development (see Table 9.8). This index is used as the mediator in the analyses reported in Table 9.10. Significant indirect effects are observed for all outcomes significantly influenced by the tested program based on the analyses by treatment group. In fact, mediation is complete in all cases except for Communication (see Direct Effects). This means that Structural Quality is sufficient to explain the tested program's effect on all outcomes except Communication. For Communication only, part of the program effect remains to be explained.

Table 9.10: Mediation Test Using Quality as the Mediator of Program Effect at 12 Months

Outcome	Direct Effect			Indirect Effect	
	DinD effect at 12 months	SE	% reduction	T-test	Sobel SE
Communication ^a	-0.30**	0.12	-30.2%	1.57*	0.11
Self-Awareness ^b	N/A	N/A	N/A	N/A	N/A
Cognition ^a	-0.03	0.08	-81.3%	1.77**	0.07
Receptive Vocabulary ^a	-0.36	0.43	Supp.	2.00**	0.08
Expressive Vocabulary ^a	-0.17	0.11	-34.6%	1.56*	0.11

Note: The negative values of DinD estimates for all group comparisons are a positive treatment effect (i.e., an advantage for the Program Daycare group). The DinD effect is for the Comparison Daycare group. Supp. = Suppression effect according to which the program effect is stronger after controlling for quality. This means there is a low negative correlation between quality and another useful element of the program (e.g., the impact of the Family Workshop Program on parents). The quality effect masked the positive effect of this unknown aspect of the tested program (at 12 months). ^a The mediator is Structural Quality. ^b No quality indicator predicted the development of Self-Awareness at 12 months. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error, N/A = not applicable.

Fidelity

Structural Fidelity was the most powerful fidelity index in predicting child development (see Table 9.9). This index is therefore used as mediating variable for the analyses reported in Table 9.11. Note that only the results for Communication are presented since there is no significant association between fidelity and other outcomes at 12 months. The results in Table 9.11 indicate that Structural Fidelity accounts for a significant share of the program effect (see the indirect effect), although a non-negligible share of the program impact has yet to be explained (see the direct effect). This is essentially the same pattern of results as for quality.

⁷³ An analysis factoring in both dimensions of implementation was not conducted due to the relatively small number of daycares. In accordance with the recommendations by Krull and MacKinnon (1999), we use the Sobel formula to calculate the standard error for the indirect effect.

Table 9.11: Mediation Test Using Fidelity as the Mediator of Program Effect at 12 Months

Outcome	Direct Effect			Indirect Effect	
	DinD effect at 12 months	SE	% reduction	T-test	Sobel SE
Communication ^a	-0.57**	0.26	Supp.	2.43***	0.66

Note: The negative values of DinD estimates for all group comparisons are a positive treatment effect (i.e., an advantage for the Program Daycare group). The DinD effect is for the Comparison Daycare group. Supp. = Suppression effect according to which the program effect is stronger after controlling for fidelity. This means there is a low negative correlation between fidelity and another useful element of the program (e.g., the impact of the Family Workshop Program on parents). The quality effect masked the positive effect of this unknown aspect of the tested program (at 12 months).^a The mediator is Structural Fidelity. The other outcomes are not presented since they were not predicted by the fidelity indices. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Summary: Mediator Effect of Fidelity and Quality

Implementation of the Daycare Program explains all the program effects observed at 12 months, except that of Communication. For this outcome, part of the tested program's effect has yet to be explained. It is interesting to note that Structural Quality and Structural Fidelity were the most powerful indices of implementation. This is also the reason why these indices were considered mediators in this series of analyses. However, the dimension that most distinguished the tested program (i.e., Reading Quality) did not predict performance in Communication or in Expressive Vocabulary at 12 months. As a result, this quality index could not be used as the mediating variable.

9.2.5. Analyses by Linguistic Profile

A child's linguistic profile can be defined in several ways. Two dimensions that must be considered are the ability to properly capture the child's active participation in linguistic exchanges (see the *First Cohort Findings Report*, 2014) and the measurement accuracy for these exchanges (a dichotomy versus a continuous variable). For this report, the preliminary analyses established that the pattern of results is similar regardless of the linguistic variable used (available upon request). Accordingly, we chose to report only those analyses based on a simple indicator (dichotomous) calculated by crossing the language spoken by the child to the mother and that spoken by the child to the father. This indicator, called Household Type, has two categories: high exposure and low exposure. The latter category consists of all linguistic configurations other than two Francophone parents.

The results are reported in two stages: a) the results of the DinDinD estimates testing the moderating effect of Household Type on the tested program's impact (see Table 9.12) and b) the DinD estimates of program effect calculated separately for both Household Types (see Table 9.13).

Table 9.12: The Moderating Effect of Household Type on School Readiness Indicators

Moderation test	School Readiness Indicator											
	Communication		Self-Awareness		Cognition		Physical		Receptive Vocabulary		Expressive Vocabulary	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
G1 vs. G2												
4 months (DinDinD)	-0.19	0.13	0.11	0.12	0.23	0.18	-0.28	0.19	-0.47*	0.24	0.28	0.21
8 months (DinDinD)	0.01	0.13	-0.01	0.17	0.52**	0.19	-0.05	0.13	-0.11	0.21	0.20	0.18
12 months (DinDinD)	-0.25	0.14	-0.04	0.20	0.48**	0.19	0.30	0.27	-0.28	0.18	-0.08	0.24
G1 vs. G3												
4 months (DinDinD)	-0.29	0.20	0.05	0.09	0.59***	0.20	-0.29	0.27	-0.47*	0.27	0.11	0.20
8 months (DinDinD)	-0.39**	0.18	0.10	0.17	0.51**	0.21	-0.07	0.25	-0.74***	0.22	-0.01	0.27
12 months (DinDinD)	-0.59***	0.17	0.28**	0.12	0.54***	0.19	0.02	0.32	-0.63***	0.18	0.04	0.24

Note: The DinDinD effects represent the difference in program impact for children with low and high exposure. The benchmark category is “low exposure,” which means that the DinDinD effect must be added to the program impact for this subgroup in order to obtain the program effect for children with high exposure. *In other words, the DinDinD effects are interpreted as follows: (a) positive effects indicate that the tested program’s effect diminishes with high exposure to French (b) negative effects indicate that the tested program’s effect is enhanced with higher exposure to French. As with the other analyses, the negative values of DinD estimates for all group comparisons are a positive treatment effect (i.e., an advantage for the Program Daycare group).* The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Table 9.13: Impact of the Tested Program by Household Type: With High or Low Exposure to French — (First Year) — Adjusted Scores

School Readiness Indicator												
Type of Difference	Low-Exposure Families											
	Communication		Self-Awareness		Cognition		Physical		Receptive Vocabulary		Expressive Vocabulary	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
G1 vs. G2												
Baseline	0.48***	0.11	0.38**	0.14	0.21*	0.11	0.32**	0.12	0.34**	0.15	0.43***	0.13
4 months (DinD)	-0.55***	0.08	-0.30***	0.07	-0.03	0.11	-0.27*	0.14	-0.40**	0.17	-0.28*	0.15
8 months (DinD)	-0.37**	0.15	-0.33**	0.14	0.29**	0.13	-0.37**	0.15	-0.42*	0.21	-0.14	0.16
12 months (DinD)	-0.54***	0.12	-0.26*	0.14	0.09	0.12	-0.08	0.19	-0.42**	0.15	-0.29**	0.13
G1 vs. G3												
Baseline	-0.08	0.14	-0.08	0.18	-0.18	0.12	-0.11	0.15	0.03	0.20	0.03	0.24
4 months (DinD)	-0.39***	0.13	-0.06	0.06	0.36**	0.16	-0.13	0.17	-0.31**	0.13	0.03	0.15
8 months (DinD)	-0.49***	0.16	-0.30**	0.14	0.22	0.17	-0.01	0.15	-0.48*	0.23	-0.17	0.18
12 months (DinD)	-0.71***	0.12	0.05	0.11	0.09	0.12	0.13	0.25	-0.59***	0.18	-0.04	0.12
Type of Difference	High-Exposure Families											
	Communication		Self-Awareness		Cognition		Physical		Receptive Vocabulary		Expressive Vocabulary	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
G1 vs. G2												
Baseline	0.16	0.12	0.14	0.12	0.13	0.12	0.24*	0.14	0.05	0.12	0.21	0.14
4 months (DinD)	-0.36**	0.15	-0.41***	0.09	-0.26*	0.15	0.01	0.18	0.07	0.27	-0.57***	0.16
8 months (DinD)	-0.38***	0.13	-0.32**	0.16	-0.23	0.14	-0.31**	0.12	-0.30**	0.14	-0.34*	0.18
12 months (DinD)	-0.29*	0.15	-0.22	0.15	-0.40***	0.12	-0.39**	0.14	-0.13	0.13	-0.21	0.20
G1 vs. G3												
Baseline	-0.02	0.10	0.03	0.13	0.14	0.11	-0.08	0.13	-0.20*	0.11	-0.01	0.16
4 months (DinD)	-0.10	0.13	-0.11	0.09	-0.24	0.15	0.16	0.19	0.17	0.27	-0.08	0.13
8 months (DinD)	-0.09	0.10	-0.40**	0.16	-0.30**	0.13	0.07	0.15	0.27*	0.15	-0.16	0.21
12 months (DinD)	-0.12	0.12	-0.24*	0.11	-0.46***	0.09	0.11	0.15	0.05	0.07	-0.09	0.18

Note: The negative values of DinD estimates for all group comparisons are a positive treatment effect (i.e., an advantage for the Program Daycare group). Exposure to the French language is defined based on Household Type, an index that crosses the language spoken by the father to the child with that spoken by the mother to the child. There are two categories: high and low exposure to French. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Communication

For Communication, the DinDinD estimates are significant in favour of the Program Daycare group when compared with the Informal Care group (see Table 9.12). A review of program effects by Household Type (see Table 9.13) shows that the tested program has a marked effect on the development of Communication, but only for children in low-exposure families. For the comparison of daycare groups, the DinD estimates (Table 9.13) indicate a positive program effect, whatever the Household Type (see also the non-significant DinDinD effects in Table 9.12). A deeper examination of Table 9.13 reveals an enhanced program effect for children with low exposure. Overall, the results tend to confirm our hypotheses: the effect of the tested program is particularly pronounced among children with low exposure to French at home.

Self-Awareness

The expected association between Household Type and the tested program's effect on Self-Awareness is not very clear in the results reported in Tables 9.12 and 9.13. The absence of association may be attributable to the fact that Self-Awareness involves measures of language and cognitive skills. The predicted interactions for these two dimensions are incompatible, as suggested by the result of independent analyses on the Communication and Cognition dimensions. It follows that program effects would tend to cancel each other out for a scale combining these two dimensions.

Cognition

For the Cognition domain, the DinDinD estimates reported in Table 9.12 clearly indicate that the tested program's effect varies significantly by Household Type. As predicted, children from families with high exposure experience cognitive benefits. The size of the effect is of the order of 0.40 standard deviations when children begin junior kindergarten (12-month assessment period) no matter what comparison group is used as the benchmark.

However, the tested program seems to impede cognitive development for children from low-exposure families. A few DinD effects reported in Table 9.13 indicate a significantly negative program effect for this subsample. However, we wish to point out that these effects appear transient and, therefore, lack credibility. For example, the program effect for these children when they begin junior kindergarten is near zero, indicating a null effect for the tested program. This observation must be confirmed with the impact analyses for the second year.

Receptive Vocabulary

According to the DinDinD estimates reported in Table 9.12, the program effect for Receptive Vocabulary varies significantly by Household Type. This contingency is particularly obvious when comparing the Program Daycare group with the Informal Care group. The DinD estimates reported in Table 9.13 confirm that children from low-exposure families are those who benefit from the tested program as regards vocabulary development.

Expressive Vocabulary

An examination of the DinDinD estimates in Table 9.12 reveals a lack of association between Household Type and the tested program's impact on Expressive Vocabulary. The DinD

estimates reported in Table 9.13 are similar for both Household Types, which tends to confirm the lack of moderating effect for Household Type. It remains to be seen whether the anticipated effect will emerge in the data analyses for the second year of the project.

Summary: Analyses by Linguistic Profile

The analyses by children's linguistic profile confirm that the program impact tends to vary according to the family's language environment. The program impact on certain linguistic outcomes like Communication and Receptive Vocabulary was stronger among children from low-exposure families. The moderating effect of Household Type was significant for both these variables toward the end of the year (at 8 and 12 months) for the Informal Care group only. However, as expected, the program impact on Cognition was stronger among children from high-exposure families. Yet again, the moderating effect of Household Type was significant toward the end of the year (at 8 and 12 months), but this time for the two comparison groups. In short, all children benefit from the program, although the type of benefit appears to depend on the children's linguistic profile.

9.3. IN THE SECOND YEAR

This section presents the results of impact analyses for the second year. Estimates of the treatment effect are based on the relative developmental trajectory of the study's three groups over the four second-year assessments. The baseline assessment is the first (i.e., the pre-intervention period) and the post-intervention assessments are those conducted at 12, 16, 20 and 24 months.

In total, four outcome measures from the ÉPE-AD were analyzed: Communication, Self-Awareness, Cognition, and Expressive Vocabulary. As with the analyses for the program's first year (see Section 9.2), the treatment effect for comparison groups is measured for these outcomes by the DinD estimator, which captures the differences among the developmental trajectories of the three treatment groups. Two other vocabulary scales are added to the battery: the EOWPVT-F and the ÉVIP-R. These scales are analyzed using an ANCOVA estimator (i.e., an estimator derived from the covariance analysis), using the baseline score for Communication as covariate to statistically cancel any existing differences among children in the pre-intervention period.⁷⁴ In total, six outcomes are examined in assessing the program effect for the second year of program delivery.

9.3.1. Technical Details

As with the first-year analyses, the general strategy adopted here is to present two alternative specifications for each analysis. The first, the initial model, considers only the key predictors that explain the program effect. Specification of the initial model may vary depending on whether the DinD estimator or the ANCOVA estimator is used. The second specification is a complexification of the first and includes covariates in addition to key predictors. The two specifications yield results that can be compared in order to identify technical problems or nuance the interpretation of results. For this comparison exercise, the initial model serves mainly

⁷⁴ The ANCOVA estimator is less valid than the DinD estimator when two non-equivalent groups are compared (for a discussion, see Jamieson, 2003).

to enhance the interpretation of final results from the adjusted model. Note that when the results of the initial model are reported in appendix or are available upon request, that fact is stated in the text. That said, the results of the initial model were considered in the preliminary interpretation of the adjusted effects in all cases.

Details of Model Specification

Initial Model

The initial model of the DinD analysis by treatment group consists of a series of dummy variables representing the assessment period, group membership, and a term representing the interaction between these dummy variable indicators (i.e., the DinD estimators). The time factor or “assessment period” is represented by four dummy variables comparing the follow-up assessments to the first assessment (or baseline assessment). In other words, the baseline assessment period (or pre-intervention period) is used as the benchmark. The “group” factor is represented by two dummy variables used to compare the comparison groups to the Program Daycare group. The Program Daycare group is therefore used as the benchmark. The initial DinD model is modified slightly in each section depending on the research question being examined. Changes made to the initial model are noted in the text. Note that the dosage, fidelity and quality variables are interpreted in the usual manner (i.e., positive effects indicate an increase in the outcome in question based on an increase in the associated variable).

The initial model of the covariance analysis (ANCOVA) by treatment group includes a series of dummy variables representing membership in a treatment group as well as the score for the Communication dimension from the first assessment. The Program Daycare group remains the benchmark category for these analyses, which means that a negative value indicates a positive effect of the tested program, as with the DinD model. The initial ANCOVA model is modified slightly in each section depending on the research question being examined. Changes made to the initial model are noted in the text for each analysis. Note that the dosage, fidelity and quality variables are interpreted in the usual manner (i.e., positive effects indicate an increase in the outcome based on an increase in the associated variable). There was no need to include the “time” dimension in the model because there is only one observation per child for each outcome.

Adjusted Model

The initial model (DinD or ANCOVA) is nested in the adjusted model, which also includes all the covariates identified in Section 7.3 of Chapter 7 as well as their interaction with the time factor.

Analysis Samples

Baseline Samples

The sample that is analyzed varies based on the outcome being studied. The total sample of participants is used for analyses of the Communication scale, the EOWPVT-F and the ÉVIP-R. For the other scales, the sample used for the analyses is that which excludes children with missing data due to English-language administration of the ÉPE-AD (i.e., the EYE-DA) and for whom the missing values in the French version of the ÉPE-AD could not be imputed. Finally,

children who withdrew from the study before the 24-month assessment are excluded from the analyses to maximize the internal validity of the DinD estimator. The other precautions introduced for data preparation are the same as those used for the first-year analyses.

9.3.2. Analyses by Group

This section presents the analyses by treatment group. See Appendix H for tables summarizing the analysis of the initial model and the adjusted model for each outcome. In this section, the figures report only the results of the analyses by group from the adjusted model. Readers will recall that the Program Daycare group, the Comparison Daycare group and the Informal Care group are identified respectively as G1, G2 and G3. *Contrary to the analyses reported in the tables, a positive program effect is indicated in the figures by a positive value.*

A figure is presented below for each outcome with estimates derived from the adjusted model. The program effect is presented in the form of 90% confidence intervals. The confidence intervals (CIs) represent a range of scores that probably include the real value of the program impact (i.e., 9 out of 10 times in this case). *If the interval does not include the value 0, the effect is considered significant at $p = 10\%$.*⁷⁵ As with the tables, we report the effects for each crossing of the time and comparison group variables. To facilitate the longitudinal interpretation of results, the figures present the DinD estimates for the second year along with those of the first-year analyses.⁷⁶

Communication Scale

Initial Model

The results of the initial model are reported in Appendix H (Table H.1).

Adjusted Model

The results of the adjusted model are presented in Figure 9.2 for both years of program delivery. First, it should be noted that the program effect never invalidates our fundamental hypothesis that the program has a positive effect on development: all the confidence intervals have positive values.

The program effect for the Informal Care group appears to diminish over time. The confidence intervals for this effect overlap the zero value at 16, 20 and 24 months, indicating a lack of significant effects. The program effect for the Comparison Daycare group is non-significant only at 16 months.

However, it would be useful to nuance the interpretation of this lack of significance at certain assessment periods. Each DinD effect can be considered a reproduction of the test for program effect. Based on this perspective, each DinD estimate attempts to estimate the same real value of the program effect.⁷⁷ If we accept this assumption, it is important to interpret not only the significance of each confidence interval, but overlap among the intervals as well. We might

⁷⁵ The complete results presented in Appendix C distinguish between the significance levels of $p = 10\%$, 5% and 1% .

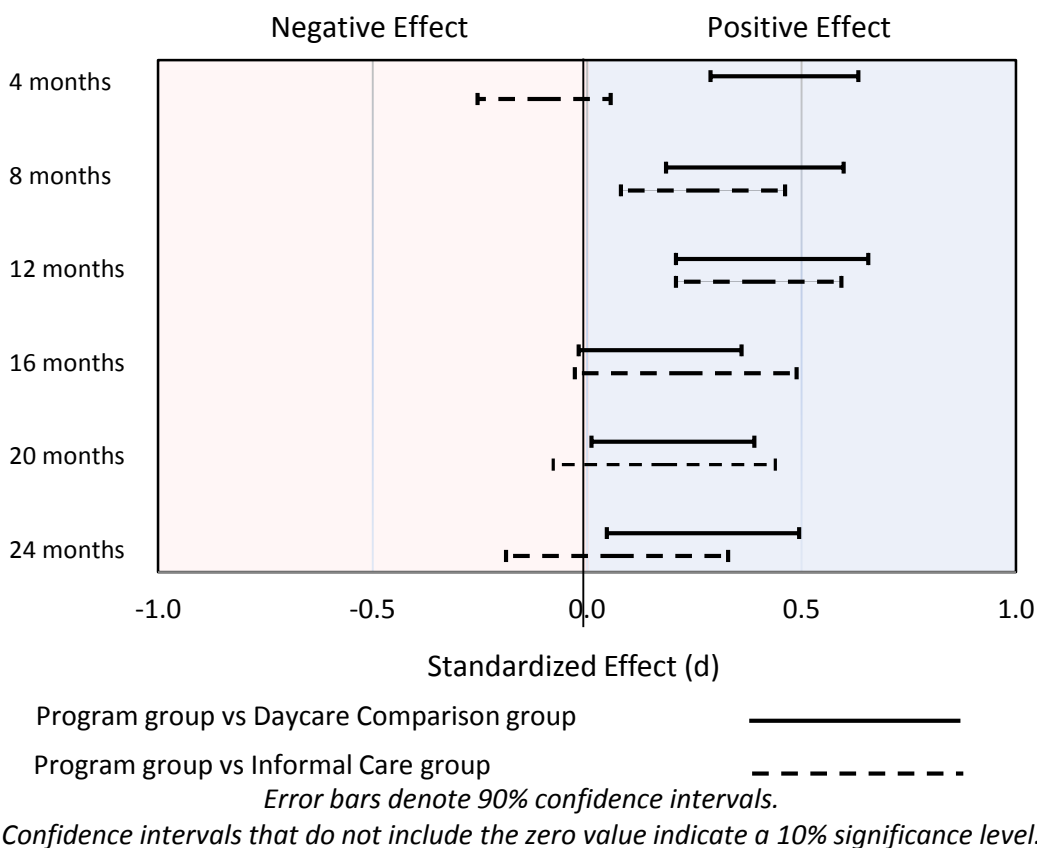
⁷⁶ Note that the results of the first-year assessments are the same as those presented in Section 9.2 for the first-year analysis.

⁷⁷ We recognize that the real program effect may vary over time. We take the perspective of a straightforward reproduction to illustrate the logic of the overall interpretation of the confidence intervals. Since there is no clear trend from one assessment to the next (the program effect is neither systematically enhanced nor diminished over time), this perspective is reasonable.

therefore ask: What values might the program effect take when there is overlap among the confidence intervals observed at each assessment period?

In fact, there is an area of overlap in the confidence intervals ranging from 0.26 to 0.36 standard deviations for the DinD effects representing the comparison of the Program Daycare group and the Comparison Daycare group. In other words, although the size of the effect varies from one assessment to the next (and is sometimes non-significant, as with the 16-month assessment period), the estimates never invalidate the hypothesis of a positive program effect of the order of about 0.30 standard deviations. Excluding the effect at 4 months, the overlap of confidence intervals for the DinD effects representing the comparison of the Program Daycare group with the Informal Care group suggests an effect of about the same magnitude. The possibility of interpreting the results in this way is among the key advantages of working with confidence intervals instead of a simple significance test.

Figure 9.2: Program Effect (DinD) on Communication at Each Assessment



Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance with the Communication scale is expressed as the probability of being classified in a higher quintile: the success category. The non-parametric analysis partly confirms the significance of the DinD effects for comparison of the Program Daycare group to the Comparison Daycare group. The

odds ratio is 0.76 (90% CI from 0.46 to 1.26), 0.60 (90% CI from 0.38 to 0.96) and 0.55 (90% CI from 0.28 to 1.08) in favour of the Program Daycare group respectively for the DinD effects at 16, 20 and 24 months. Thus, the effect at 20 months reported in Figure 9.2 is the only one reproduced. However, the effect at 24 months approaches the 10% significance level. Finally, as with the parametric analysis reported in Figure 9.2, none of the DinD effects at 16, 20 and 24 months is significant for the comparison with the Informal Care group.

Self-Awareness Scale

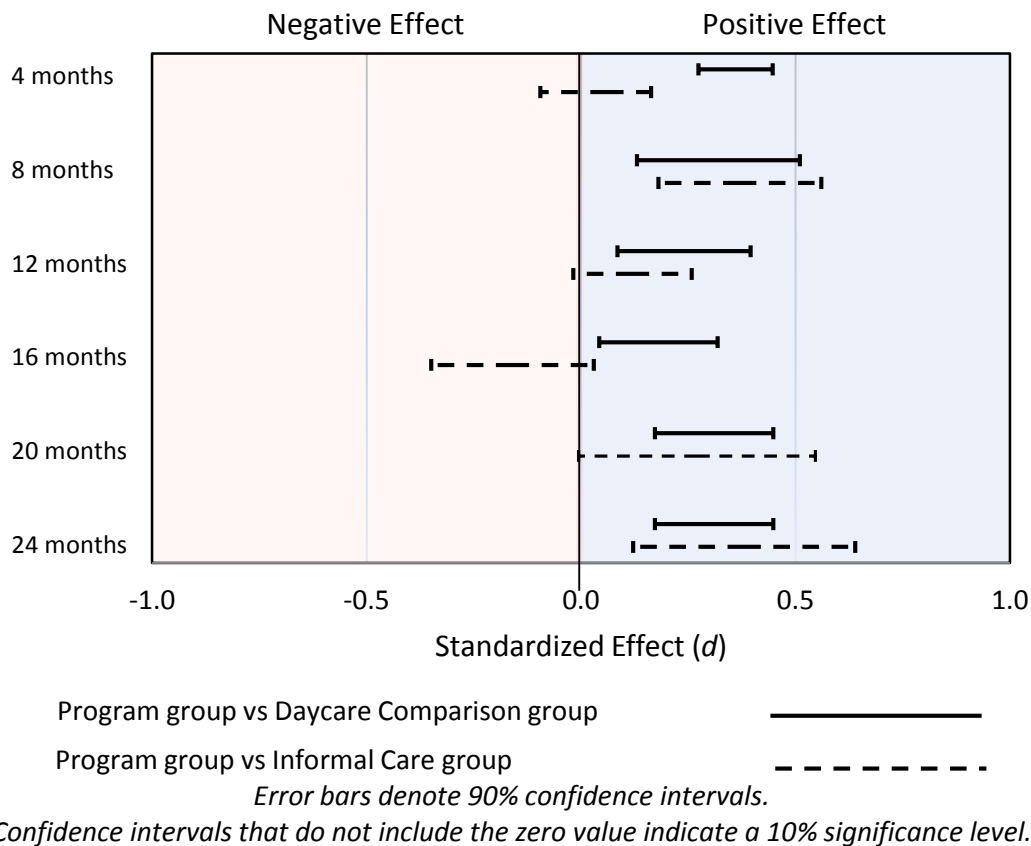
Initial Model

The results of the initial model are reported in Appendix H (Table H.1).

Adjusted Model

The results of the adjusted model are presented in Figure 9.3 for the two years of program delivery. First, note that the confidence intervals never invalidate our fundamental hypothesis that the program has a positive effect on development. The program effect for the Informal Care group is the least accurate (larger confidence intervals) in addition to being unstable from one assessment to the next in terms of significance and direction (positive versus negative). However, it is significant at 24 months ($d = 0.38$). Comparison of the daycare groups reveals a stable and significant effect in favour of the program for both years of program delivery. The area of CI overlap ranges from 0.27 to 0.32 standard deviations, indicating that the reported confidence intervals never invalidate the hypothesis that the program's real effect is found between these values (see the analysis of Communication for further explanations).

Figure 9.3: Program Effect (DinD) on Self-Awareness at Each Assessment



Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance with the Self-Awareness scale is expressed as the probability of being classified in a higher quintile: the success category. The non-parametric analysis reproduces the significance of the DinD effects observed for comparisons of the daycare groups reported in Figure 9.3. Thus, the odds ratio is 0.67 (90% CI from 0.46 to 0.99), 0.48 (90% CI from 0.34 to 0.69), and 0.59 (90% CI from 0.41 to 0.82) respectively for the DinD effects at 16, 20 and 24 months. However, the non-parametric analysis does not reproduce the significance of the observed DinD effects for comparisons with the Informal Care group. The odds ratio is 0.68 (90% CI from 0.46 to 1.01) for the assessment at 24 months. Although this effect approaches the 10% significance level, it is nevertheless non-significant.

Cognition Scale

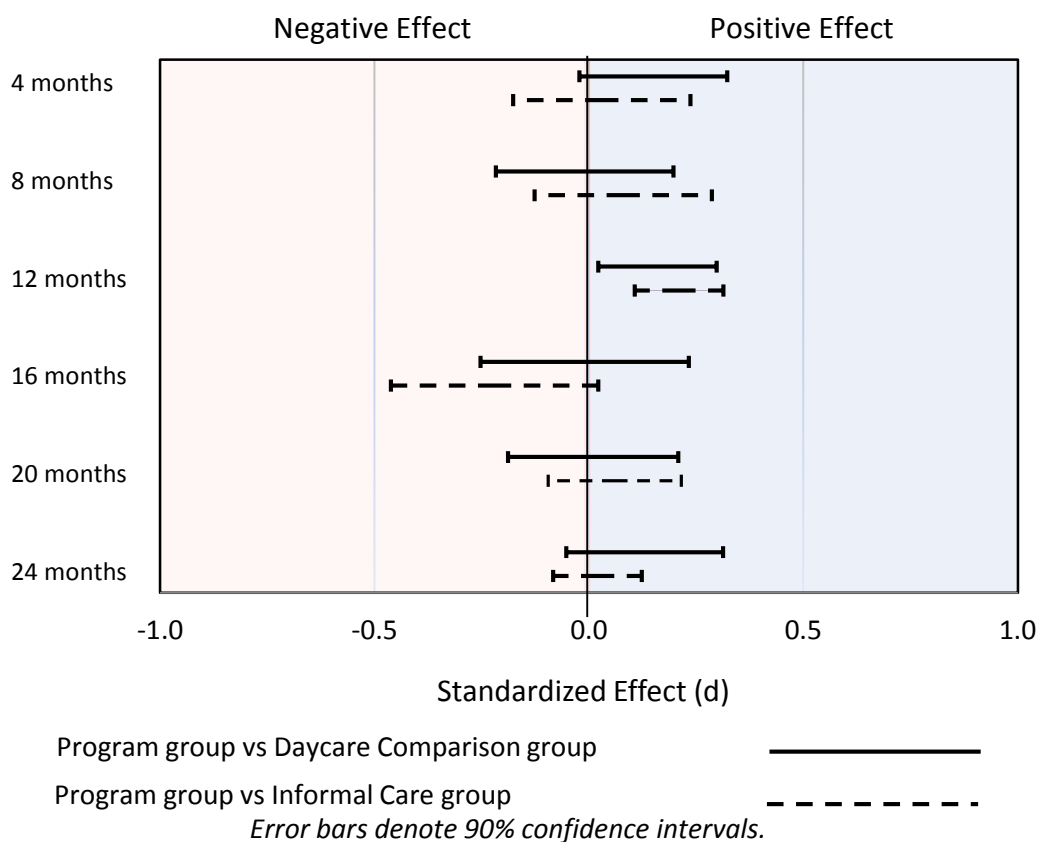
Initial Model

The results of the initial model are reported in Appendix H (Table H.1).

Adjusted Model

The results of the adjusted model are presented in Figure 9.4 for both years of program delivery. It shows no significant effects in favour of the Program Daycare group for the two comparison groups.

Figure 9.4: Program Effect (DinD) on Cognition at Each Assessment



Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance with the Cognition scale is expressed as the probability of being classified in a higher quintile: the success category. The non-parametric analysis reproduces the lack of significance of the observed DinD effects for this outcome in the comparisons with the two comparison groups.

Expressive Vocabulary Scale

Initial Model

The results of the initial model are reported in Appendix H (Table H.1).

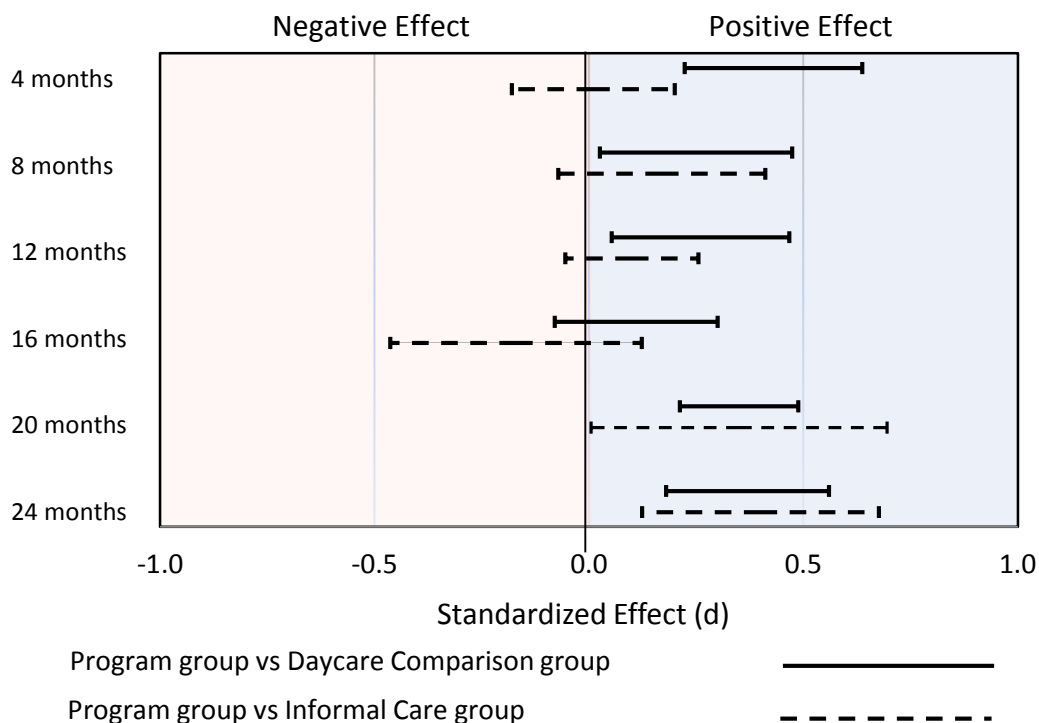
Adjusted Model

The results of the adjusted model are presented in Figure 9.5 for both years of program delivery. All confidence intervals have positive values, supporting the notion that the program has a positive effect on child vocabulary development.

When children begin senior kindergarten (24-month assessment period), we observe a significant advantage of 0.37 standard deviations for the Program Daycare group versus the Comparison Daycare group and of 0.41 standard deviations versus the Informal Care group. These benefits exceed the initial deficit of the Program Daycare group and represent gains of 3.3 and 3.7 months in vocabulary development for children in the Program Daycare group versus the comparison groups when we refer to the data of Hill and colleagues (2007). This estimate is most likely conservative if we refer instead to ÉVIP-R standards (Dunn, et al., 1993, Table 4.4, Form A), according to which a gain in vocabulary of 0.49 standard deviations is expected between the age of three-and-a-half to age four-and-a-half. Applying the developmental curve for Receptive Vocabulary captured by the ÉVIP-R to that of Expressive Vocabulary, the effect reported here represents a gain of about 9.6 months over the Comparison Daycare group and of about 10 months over the Informal Care group. It remains to be seen whether comparable results will be obtained with the ÉVIP-R in the analyses to follow.

These results confirm a positive and stable program effect when compared to the Comparison Daycare group, and the effect persists until the 24-month assessment. Moreover, the significance of the program effect for the Informal Care group is observed only in the second year, emerging late.

Figure 9.5: Program Effect (DinD) on Expressive Vocabulary at Each Assessment



A line represents a confidence interval of 90%.

Confidence intervals that do not include the zero value indicate a 10% significance level.

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance with the Expressive Vocabulary scale is expressed as the probability of being classified in a higher quintile: the success category. The non-parametric analysis reproduces the significance of the DinD effects for the comparison of daycare groups reported in Figure 9.5. The odds ratio is 0.81 (90% CI from 0.48 to 1.33), 0.52 (90% CI from 0.36 to 0.75), and 0.48 (90% CI from 0.26 to 0.87) in favour of the Program Daycare group respectively for the DinD effects at 16, 20 and 24 months. The DinD effect for comparison with the Informal Care group at 24 months is less robust, as it is not significant based on the non-parametric analysis, at 0.80 (90% CI from 0.48 to 1.31). In fact, this effect is far from the 10% significance level.

EOWPVT-F and ÉVIP-R Scales

The ANCOVA estimates for the initial model and the adjusted model are reported for the EOWPVT-F (Expressive Vocabulary) and the ÉVIP-R (Receptive Vocabulary) in Table G.1 (Appendix G). Note that the distinction between the ANCOVA estimator and the DinD estimator is that with the ANCOVA estimator, a pre-intervention variable (in the case at hand, Communication) serves as covariate in the model, playing a role that is conceptually similar to that of the baseline assessment for the DinD estimator (the two methods are not statistically equivalent). Only the result for the adjusted ANCOVA model is reported graphically in Figure 9.6. We address the results for both outcomes at the same time.

Initial Model

The results of the initial model are reported in a table in Appendix G. Since the results of the adjusted model are interpreted based on the initial model, it should be noted that the “positive” program effects on the development of school readiness take on a negative value in the tables and a positive value in the figures. The two types of results are compared on the basis of the absolute size of the effect (i.e., the size of the effect ignoring the negative or positive sign since in both cases the program has a positive impact on development).

Adjusted Model

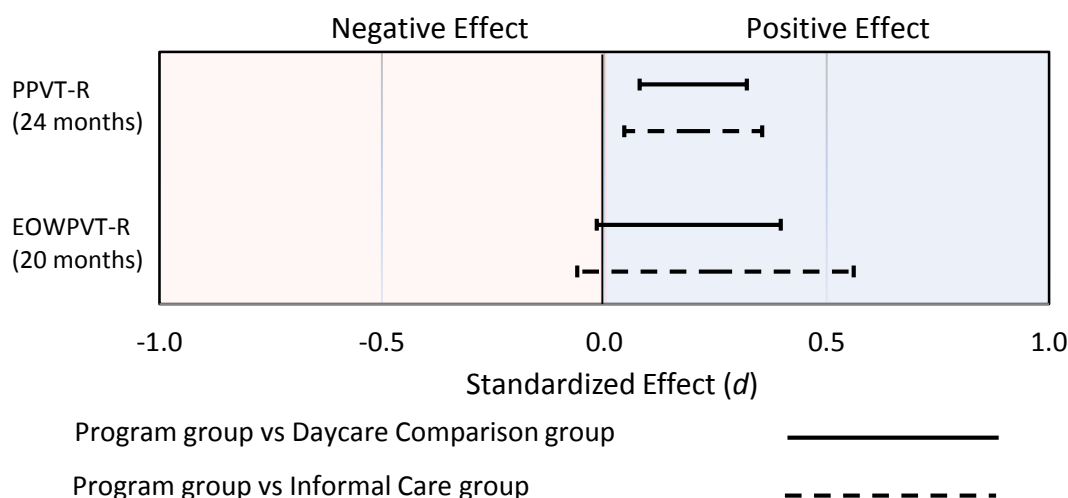
According to the ANCOVA estimates for the adjusted model, there is a difference of about 0.20 standard deviations between the daycare groups for the ÉVIP-R and the EOWPVT. This difference is significant for the ÉVIP-R and non-significant for the EOWPVT. We note that the size of the program effect on vocabulary for the Comparison Daycare group is robust to changes in specifications and outcomes (ÉVIP-R versus EOWPVT).

The adjusted model shows a significant difference between the Program Daycare group and the Informal Care group for the ÉVIP-R (see the reported confidence interval that does not include the zero value in Figure 9.6). An advantage of 0.20 standard deviations is observed in favour of the Program Daycare group for this comparison. A comparable, but non-significant, effect is observed with the EOWPVT-F for the Informal Care group. In short, the analyses indicate that for this sample, the tested program has a positive effect on the development of Receptive Vocabulary.

We use the ÉVIP-R standards to express the gains observed in terms of “months” of development. According to the ÉVIP standards (Dunn et al., 1993, Table 4.4, Form A), a vocabulary gain of about 0.38 standard deviations is typical between the ages of four and five. The effects observed with the Readiness to Learn project would therefore represent about 6.6 months of development for Receptive Vocabulary (ÉVIP-R).⁷⁸ The size of the non-significant effects reported in Figure 9.6 suggest that a comparable effect would be observed with Expressive Vocabulary (EOWPVT), that is, a gain of about six months.

⁷⁸ Conversion formula: $[(0.21 \text{ program effect} / 0.38 \text{ typical change}) * 12 = 6.6 \text{ months}]$

Figure 9.6: Confidence Intervals for the Impact of the Standardized Program (ANCOVA) on French Vocabulary: Standardized Receptive Vocabulary (ÉVIP-R) and Expressive Vocabulary (EOWPVT) Scales



A line represents a confidence interval of 90%.

Confidence intervals that do not include the zero value indicate a 10% significance level.

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance with the ÉVIP-R and the EOWPVT-F is expressed as the probability of being classified in a higher quintile: the success category. This analysis confirms the results reported in Figure 9.6. The non-parametric analysis reproduces the significant effects observed with the ÉVIP. We obtain an odds ratio of 0.64 (90% CI from 0.47 to 0.87) in favour of the Program Daycare group for comparison with the Comparison Daycare group and an odds ratio of 0.59 (90% CI from 0.38 to 0.90) in favour of the Program Daycare group for comparison with the Informal Care group. Likewise, the non-parametric analysis reproduces the program effect's lack of significance for the EOWPVT. The odds ratio is 0.63 (90% CI from 0.34 to 1.18) and 0.63 (90% CI from 0.28 to 1.41) respectively for comparisons with the Comparison Daycare group and the Informal Care group.

Summary: Analyses by Group

The results indicate that the tested program appears to have a positive effect on children's school readiness. For Communication, Self-Awareness, Expressive Vocabulary and the ÉVIP-R, the tested program is observed to have a positive effect when the participating children begin senior kindergarten (24 months). This effect is observed for both comparison groups for each of these outcomes except Communication. For Communication, the effect is observed only with the Comparison Daycare group. No stable program effect is observed from one assessment to the next for Cognition in this analysis. Likewise, the program did not have a significant effect on the EOWPVT-F in this analysis. The anticipated effects will be observed with other models presented below.

9.3.3. Analyses by Dosage

The results of the analysis of the dosage effect based on membership in the Program Daycare group or the Comparison Daycare group are reported in Table 9.14 for the initial model and in Table 9.15 for the adjusted model. Since both models generated essentially the same pattern of results, we are interpreting only the results of the adjusted model (see Table 9.15). Moreover, we discuss only the dosage effects for the Cognition scale due to the absence of significant dosage effects for other scales. *Note that, contrary to the figures presented in the last section, the tested program's "positive" effects on development take on negative values for the DinD estimates reported in the tables.*

Cognition Scale

Initial Model

As indicated earlier, no interpretation is provided for this model's results. These are reported in Table 9.14.

Adjusted Model

The analysis results indicate that first-year dosage is not very predictive of development in the second year for ÉPE-AD scales. The only observed exception is for the Cognition scale. The DinDinD estimates for this variable indicate that the dosage effect is differentiated by treatment group for all the assessment periods in the second year except the 24-month assessment. When significant, the effect is about 0.30 to 0.40 standard deviations for every 10-hour increase/reduction in exposure per week. It is interesting to note that the program effect is completely absent in the analysis by treatment group. This suggests that for the 16- and 20-month assessments, contrary to the data in Figure 9.4, a program effect is observed for the Cognition scale only for children who received a very high dosage (at least 37 hours per week).

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance with the Cognition scale is expressed as the probability of being classified in a higher quintile: the success category. The non-parametric analysis reproduces only the significant DinDinD effects at 16 and 20 months. The odds ratios for the assessment periods at 16, 20 and 24 months are 0.94 (90% CI from 0.89 to 0.98), 0.92 (90% CI from 0.86 to 0.99) and 0.97 (90% CI from 0.89 to 1.05) in favour of the Program Daycare group.

Table 9.14: Number of Hours Spent at Daycare in the First Year (Dosage) and Its Unadjusted Effect on the Development of School Readiness in the Second Year for the Two Daycare Groups — Standardized Scores

Type of Difference	School Readiness Indicator							
	Communication		Self-Awareness		Cognition		Expressive Vocabulary	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Dosage								
Baseline	0.029***	0.009	0.015	0.017	-0.023	0.016	0.009	0.015
Dosage x Group								
G1 vs. G2 (DinD)	-0.016	0.015	-0.005	0.022	0.024	0.020	-0.001	0.020
Dosage x Time								
12 months (DinD)	0.004	0.021	0.021	0.015	0.040**	0.017	0.016	0.015
16 months (DinD)	0.009	0.013	0.002	0.007	0.040***	0.008	0.018	0.012
20 months (DinD)	-0.012	0.017	0.015	0.009	0.029**	0.012	0.021**	0.007
24 months (DinD)	-0.009	0.011	0.015**	0.007	0.011	0.009	0.022***	0.004
Dosage x Time x Group								
12 months (DinDinD)	0.012	0.022	-0.011	0.019	-0.045**	0.019	-0.004	0.019
16 months (DinDinD)	0.000	0.012	0.003	0.012	-0.039**	0.015	-0.008	0.014
20 months (DinDinD)	0.005	0.020	-0.013	0.013	-0.034*	0.017	-0.014	0.011
24 months (DinDinD)	0.013	0.010	-0.009	0.015	-0.008	0.016	-0.009	0.011
Dose x Time Wald F	2.02		1.29		4.35**		2.32*	
Dose x Time x Group Wald F	0.74		0.92		5.09***		0.54	

Note: The negative values of DinD estimates for all group comparisons are a positive treatment effect (i.e., an advantage for the Program Daycare group). The degrees of freedom for the Wald F-tests are 4 and 19. The Huber-White robust standard errors are reported with error terms clustered by daycare. The difference between the groups is significant at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Table 9.15: Number of Hours Spent at Daycare in the First Year (Dosage) and Its Adjusted Effect on the Development of School Readiness in the Second Year for the Two Daycare Groups — Standardized Scores

Type of Difference	School Readiness Indicator							
	Communication		Self-Awareness		Cognition		Expressive Vocabulary	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Dosage								
Baseline	0.007	0.008	0.002	0.012	-0.015	0.011	-0.005	0.009
Dosage x Group								
G1 vs. G2 (DinD)	-0.017*	0.008	-0.005	0.014	0.016	0.012	-0.002	0.011
Dosage x Time								
12 months (DinD)	0.006	0.021	0.016	0.015	0.028*	0.014	0.010	0.016
16 months (DinD)	0.010	0.014	-0.004	0.011	0.026***	0.008	0.005	0.010
20 months (DinD)	0.020	0.015	0.009	0.015	0.033***	0.011	0.013	0.011
24 months (DinD)	0.000	0.014	0.010	0.015	0.013	0.014	0.009	0.010
Dosage x Time x Group								
12 months (DinDinD)	0.014	0.020	-0.013	0.019	-0.038**	0.017	-0.002	0.018
16 months (DinDinD)	0.005	0.012	0.005	0.014	-0.032**	0.012	0.000	0.011
20 months (DinDinD)	-0.012	0.015	-0.015	0.016	-0.038**	0.014	-0.014	0.012
24 months (DinDinD)	0.014	0.013	-0.012	0.018	-0.012	0.017	-0.002	0.012
Dose x Time Wald F	5.37***		1.92		2.00		0.32	
Dose x Time x Group Wald F	6.25***		1.64		5.77***		1.33	

Note: The negative values of DinD estimates for all group comparisons are a positive treatment effect (i.e., an advantage for the Program Daycare group). The degrees of freedom for the Wald F-tests are 4 and 19. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

EOWPVT-F and ÉVIP-R Scales

The ANCOVA estimates for the initial model and the adjusted model are reported for the EOWPVT-F (Expressive Vocabulary) and the ÉVIP-R (Receptive Vocabulary) in Table 9.16. Note that the distinction between the ANCOVA estimator and the DinD estimator is that for the ANCOVA estimator, a pre-intervention variable (in this case, Communication) serves as a covariate in the model, playing a role that is conceptually similar to that of the baseline assessment in the DinD estimator (the two methods are not statistically equivalent). We address the results of the two scales at the same time.

Initial Model

The results from the initial model (Table 9.16) indicate that dosage has a positive effect on Receptive and Expressive Vocabulary, as measured by the ÉVIP-R and the EOWPVT. The dosage effect is significant, but the effect of dosage–treatment group interaction is not. Based on

these results, the number of hours spent at daycare brings positive gains in vocabulary, whatever the daycare program (Daycare Program or comparison program).

Adjusted Model

The dosage effect observed with the initial model is cancelled in the adjusted model (Table 9.16). This pattern of results suggests that the effect observed with the initial model is an artefact associated with the confounding variables. Moreover, the effect of the treatment group is significant in the adjusted model for both outcome variables. The size of the observed effects is similar to that observed in the analysis by group, about 0.20 standard deviations (see Figure 9.6). Observation of this difference in the adjusted model is attributable to greater statistical power after including dosage in the specification.

In short, the analysis results indicate that first-year dosage is not very predictive of Expressive and Receptive Vocabulary development in the second year. That said, the greater statistical power that results from including dosage in the model nevertheless allowed us to statistically detect the difference of 0.20 standard deviations in favour of the Program Daycare group. Finally, we estimate using ÉVIP-R standards that this effect represents a gain of about 6.6 months in Receptive and Expressive Vocabulary.

Table 9.16: Impact of Daycare Dosage on French Vocabulary: Standardized Receptive Vocabulary (ÉVIP-R) and Expressive Vocabulary (EOWPVT) Scales

Type of Difference	Receptive Vocabulary (ÉVIP-R)				Expressive Vocabulary (EOWPVT)			
	Unadjusted (N = 217)		Adjusted (N = 212)		Unadjusted (N = 215)		Adjusted (N = 209)	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Group (G1 vs. G2)								
Raw score	-4.029	2.409	-4.648***	1.129	-2.595	1.538	-2.595**	1.076
Standardized score	-0.224	0.134	-0.258***	0.063	-0.193	0.114	-0.193**	0.080
Dosage								
Raw score	0.236***	0.057	-0.011	0.096	0.231**	0.090	0.072	0.087
Standardized score	0.013***	0.003	-0.001	0.005	0.017**	0.007	0.005	0.006
Dosage x Group								
Raw score	-0.138	0.120	-0.112	0.132	-0.076	0.130	-0.068	0.091
Standardized score	-0.008	0.007	-0.006	0.007	-0.006	0.010	-0.005	0.007

Note: The negative values of *DinD* estimates for all group comparisons are a positive treatment effect (i.e., an advantage for the Program Daycare group). Standardized scores are raw data converted to a Z-score with a mean of zero and a standard deviation of 1. Assignment to the study groups for scale administration is used for the comparisons reported in this table. The ÉVIP-R was administered at 20 months and the EOWPVT, at 24 months. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$; ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Non-Parametric Cross-Validation

For the non-parametric analysis, the children's observed performance with the ÉVIP-R and the EOWPVT-F is expressed as the probability of being classified in a higher quintile: the

success category. As with the parametric results presented in Table 9.16, no differentiated dosage effect (DinDinD) by treatment group is detected in this analysis.

Summary: Analyses by Dosage

Based on the reported analyses, dosage in the first year of the study is not very useful in predicting the program impact in the second year of program delivery. Exceptionally, the program impact on cognitive development appears to depend on dosage, at least for the assessments at 12, 16 and 20 months (see Table 9.15). As such, it would seem that the long-term program effects are more sensitive to dosage for cognitive development than for the development of other competencies. Finally, controlling for dosage shows the tested program's positive impact on the development of Expressive (EOWPVT) and Receptive (ÉVIP-R) Vocabulary. This last finding shows the usefulness of dosage as a covariate in enhancing the accuracy of statistical testing.

9.3.4. Analyses by Quality and Fidelity of the Daycare Program in the First Year

This section presents analyses examining the effect of Daycare Program quality and fidelity. The following sections present the results for two series of analyses: (a) a verification of the predictive value in the second year of the tested program's key dimensions for fidelity and quality of implementation; and (b) a mediation analysis ascertaining whether Daycare Program implementation is entirely or partially responsible for the program effects reported in Section 9.3.2. These analyses estimate the importance of Daycare Program elements in accounting for the observed effects of the tested program.

Impact of Fidelity and Quality

The results of the analysis by quality and fidelity of the Daycare Program are presented respectively in Tables 9.17 and 9.18 for ÉPE-AD outcomes and for the vocabulary scales in Table 9.19. We describe the results for each school readiness indicator separately.

Structural Quality

At the baseline, the relationship between Structural Quality and two outcomes, Communication and Cognition, is significantly negative (see Table 9.17). When children begin senior kindergarten (24-month assessment period), we observe higher gains (see the DinD effects) for daycares characterized by better Structural Quality, but only for Cognition. The size of the effect is 0.13 standard deviations for each quality "point" in the second year of program delivery, which constitutes a slight softening of the effect observed at 12 months. Note that the Structural Quality effect is generally positive and significant for certain DinD tests at other times and for other outcomes. The adjusted model reported in Table 9.19 indicates that Structural Quality also predicts the development of Expressive Vocabulary as measured by the EOWPVT.

Educative Quality

We observe a negative relationship at baseline between Educative Quality and all outcomes in Table 9.17. Based on this finding, children attending a high-quality daycare tend to be disadvantaged in these outcomes. However, the Educative Quality received in the first year at daycare had a positive effect on child development in the second year of the project. Almost all

DinD effects are significant and positive for all assessment periods. The model predicts approximately that a one-point difference in Educative Quality would result in a gain of 0.09 to 0.17 standard deviations in Communication when children begin senior kindergarten (24-month assessment period).

According to the DinD estimates of the adjusted model reported in Table 9.19, Educative Quality also predicts the development of Expressive Vocabulary as measured by the EOWPVT. For this outcome, the size of the effect is about 0.10 standard deviations per quality point.

Educator Sensitivity

An examination of Table 9.17 shows a lack of significant association between Educator Sensitivity and each outcome at the baseline. The developmental gains subsequently observed are not systematically associated with this factor other than a few positive DinD effects (at 16 months for Self-Awareness; at 12 and 16 months for Expressive Vocabulary). Finally, when children begin senior kindergarten, no DinD effect for Educator Sensitivity is detected.

The adjusted model reported in Table 9.19 indicates that Educator Sensitivity also predicts the development of Expressive Vocabulary as measured by the EOWPVT. For this outcome, the size of the effect is about 0.05 standard deviations per quality point.

Reading Quality

At baseline, we observe a negative relationship between Reading Quality and Expressive Vocabulary (see Table 9.17). When children begin senior kindergarten (24-month assessment period), we observe higher gains at baseline for Self-Awareness and Expressive Vocabulary. According to the model presented, a difference of three points on the Reading Quality scale is equivalent to an effect of about 0.20 standard deviations on the Self-Awareness and Expressive Vocabulary scales.

The adjusted model reported in Table 9.19 indicates that Reading Quality also predicts the development of Expressive Vocabulary as measured by the EOWPVT-F and Receptive Vocabulary as measured by the ÉVIP-R. For these outcomes, the size of the effect is about 0.03 standard deviations per quality point. This is the only aspect of quality significantly associated with these two measures of vocabulary.

Overall Quality

Overall Quality (i.e., the average of the above four quality indices) is significantly associated with poor school readiness (see Table 9.17). The DinD estimates indicate that when children begin senior kindergarten (24-month assessment period), the program's Overall Quality is positively associated with the developmental gains of children who attend daycare for Self-Awareness, Cognition, and Expressive Vocabulary. The significant effect most stable over time is that observed for Expressive Vocabulary, which is found for all assessment periods. According to the model presented, a difference of three quality points (roughly the range observed) would correspond to an effect of 0.60 standard deviations.

The adjusted model reported in Table 9.19 indicates that Overall Quality significantly predicts the development of Expressive Vocabulary as measured by the EOWPVT. For this

outcome, the size of the effect is about 0.13 standard deviations per quality point. It is no surprise that this overall index alone captures the observed effect given all the elements it encompasses.

Structural Fidelity

At the baseline, we observe in Table 9.18 a negative association between Structural Fidelity and Communication, Self-Awareness, and Expressive Vocabulary. Note that this relationship is cancelled or reversed at 20 months. For example, for Communication, the positive DinD coefficient of 0.537 is greater than the deficit of -0.492 observed at the baseline, $0.537 - (-0.49) = +0.047$. Moreover, the DinD coefficients for the period when children begin senior kindergarten (24-month assessment period) are significantly positive for Communication, Self-Awareness, and Expressive Vocabulary. The largest effect is observed with Expressive Vocabulary, for which it is estimated that a 50% difference in fidelity corresponds to a gain of about 0.50 standard deviations.

The adjusted model reported in Table 9.19 indicates that Structural Fidelity is predictive of vocabulary development as measured by the ÉVIP-R and the EOWPVT. In both cases, a difference of 40 percentage points for fidelity would correspond to an effect of about 0.15 standard deviations.

Content Fidelity

At baseline, Content Fidelity is negatively associated with school readiness (see Table 9.18). This relationship can be attributed to the fact that children who are less prepared tend to enrol at daycares with greater Content Fidelity. The hypothesis that program content is the most important determinant of developmental gains is invalidated by the absence of significant DinD estimates for this indicator. The adjusted model reported in Table 9.18 also indicates the absence of any positive effects associated with this aspect.

Overall Fidelity

The pattern of results observed in Table 9.18 for overall fidelity resembles that of Structural Fidelity. A detailed description of the results would therefore be redundant. The adjusted model reported in Table 9.19 indicates a positive relationship only for the EOWPVT.

Non-Parametric Cross-Validation

With few exceptions, the non-parametric analyses reproduce the results of the effect of program quality and fidelity on child outcomes stemming from the DinD and ANCOVA analyses (see Tables 9.17 and 9.18).

Summary: Analyses by Quality and Fidelity of the Daycare Program in the First Year

This section presents results indicating that certain program elements, particularly Structural Fidelity, Structural Quality and Educative Quality, account for child development since the baseline assessment (pre-intervention assessment period), especially for Cognition and Expressive Vocabulary. Having established this to be so, in the next section we verify whether these same indicators explain the tested program's effect (i.e., the effect of the treatment group) reported in Section 9.3.2. This analysis is a way to determine the extent to which the program

effect can be attributed to the daycare component alone. It is interesting to note that Reading Quality- the index with the greatest difference between the tested Daycare Program and the program at other daycares - is significantly associated with gains in Self-Awareness and Expressive Vocabulary when children begin senior kindergarten, as well as with the ÉVIP-R and EOWPVT-F vocabulary scales.

Table 9.17: Quality of the Daycare Program in the First Year and Its Adjusted Effect on the Development of School Readiness in the Second Year — Standardized Scores

Type of Difference	School Readiness Indicator							
	Communication		Self-Awareness		Cognition		Expressive Vocabulary	
Structural Quality	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.17*	0.09	-0.09	0.08	-0.16**	0.07	-0.13	0.10
12 months (DinD)	0.21*	0.11	0.17**	0.06	0.18**	0.07	0.26**	0.11
16 months (DinD)	0.11	0.08	-0.01	0.03	0.20*	0.10	0.08*	0.04
20 months (DinD)	0.24**	0.11	0.16**	0.08	0.13**	0.06	0.18*	0.10
24 months (DinD)	0.10	0.07	0.14	0.08	0.12*	0.07	0.16	0.09
<i>Effect by Time Wald F</i>	1.13		4.56**		1.56		1.42	
Educative Quality	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.13***	0.03	-0.11**	0.04	-0.18***	0.03	-0.15***	0.04
12 months (DinD)	0.13**	0.05	0.09*	0.05	0.11***	0.03	0.17**	0.06
16 months (DinD)	0.09**	0.03	0.04	0.04	0.14**	0.05	0.09**	0.04
20 months (DinD)	0.14***	0.04	0.14***	0.04	0.16***	0.04	0.19***	0.05
24 months (DinD)	0.08	0.05	0.08**	0.03	0.13***	0.03	0.17***	0.03
<i>Effect by Time Wald F</i>	2.47*		3.49**		6.49***		4.68***	
Educator Sensitivity	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.03	0.02	-0.01	0.03	-0.04	0.03	-0.04	0.03
12 months (DinD)	0.03	0.04	0.01	0.02	-0.03	0.03	0.07*	0.03
16 months (DinD)	0.04	0.03	0.04*	0.02	0.01	0.03	0.07**	0.02
20 months (DinD)	0.04	0.03	0.03	0.03	0.03	0.02	0.06	0.04
24 months (DinD)	0.04	0.04	-0.00	0.03	0.01	0.03	0.04	0.03
<i>Effect by Time Wald F</i>	0.49		2.08		7.47***		2.06	
Reading Quality	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.02	0.01	-0.02	0.01	-0.01	0.02	-0.04**	0.01
12 months (DinD)	0.01	0.01	0.01	0.02	0.00	0.02	0.01	0.02
16 months (DinD)	-0.01	0.01	0.01	0.02	-0.04	0.02	0.01	0.02
20 months (DinD)	0.01	0.01	0.05***	0.01	0.00	0.02	0.06***	0.01
24 months (DinD)	0.02	0.02	0.03**	0.01	0.01	0.02	0.06**	0.02
<i>Effect by Time Wald F</i>	0.36		5.68***		3.80**		3.98**	

Type of Difference	School Readiness Indicator							
	Communication		Self-Awareness		Cognition		Expressive Vocabulary	
Overall Quality	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.13***	0.04	-0.10*	0.05	-0.13**	0.05	-0.16***	0.05
12 months (DinD)	0.13*	0.06	0.08	0.05	0.03	0.04	0.16**	0.06
16 months (DinD)	0.07	0.05	0.06	0.04	0.02	0.06	0.11**	0.04
20 months (DinD)	0.12*	0.06	0.17***	0.05	0.10**	0.03	0.22***	0.07
24 months (DinD)	0.10	0.07	0.10*	0.05	0.09*	0.04	0.20***	0.06
<i>Effect by Time Wald F</i>	1.15		5.31***		2.64*		3.42**	

Note: Contrary to the analyses by treatment group, positive values for the DinD effects indicate a positive relationship between the fidelity index and the school readiness indicators. The degrees of freedom for the Wald F-test are 4 and 18. Overall Quality is the average of the four quality indicators. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff. = difference, SE = standard error.

Table 9.18: Fidelity of the Daycare Program in the First Year and Its Adjusted Effect on the Development of School Readiness in the Second Year — Standardized Scores

Type of Difference	School Readiness Indicator							
	Communication		Self-Awareness		Cognition		Expressive Vocabulary	
Structural Fidelity	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.49***	0.15	-0.33**	0.15	-0.34	0.24	-0.65***	0.15
12 months (DinD)	0.64***	0.21	0.14	0.14	0.22	0.17	0.46*	0.24
16 months (DinD)	0.29	0.22	0.15	0.17	0.00	0.30	0.25	0.23
20 months (DinD)	0.35*	0.17	0.54***	0.14	0.01	0.22	0.88***	0.18
24 months (DinD)	0.53**	0.21	0.49***	0.14	0.34	0.21	0.93***	0.17
<i>Effect by Time Wald F</i>	2.37*		4.61***		3.37**		8.82***	
Content Fidelity	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.65***	0.21	-0.51*	0.27	-0.64	0.37	-0.61*	0.33
12 months (DinD)	0.33	0.22	0.10	0.35	0.22	0.34	-0.09	0.39
16 months (DinD)	0.07	0.31	-0.12	0.37	-0.27	0.47	-0.23	0.46
20 months (DinD)	0.12	0.31	0.46	0.45	0.25	0.47	0.58	0.44
24 months (DinD)	0.36	0.47	0.28	0.38	0.20	0.41	0.75	0.48
<i>Effect by Time Wald F</i>	0.53		1.21		0.60		1.81	
Overall Fidelity	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Baseline	-0.66***	0.18	-0.47**	0.20	-0.52	0.32	-0.79***	0.21
12 months (DinD)	0.69**	0.26	0.17	0.21	0.27	0.24	0.37	0.31
16 months (DinD)	0.29	0.30	0.08	0.25	-0.08	0.41	0.14	0.35
20 months (DinD)	0.36	0.26	0.64**	0.23	0.10	0.31	0.98***	0.27
24 months (DinD)	0.60*	0.34	0.53**	0.20	0.38	0.27	1.08***	0.27
<i>Effect by Time Wald F</i>	1.71		2.20		2.40*		5.17***	

Note: Contrary to the analyses by treatment group, positive values for the DinD effects indicate a positive relationship between the Fidelity index and the school readiness indicators. The degrees of freedom for the Wald F-tests are 4 and 18. Overall Fidelity is the average of the four fidelity indicators. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Table 9.19: Impact of Program Quality and Fidelity on French Vocabulary: Standardized Receptive Vocabulary (ÉVIP-R) and Expressive Vocabulary (EOWPVT) Scales

Type of Difference	Program Quality							
	Receptive Vocabulary (ÉVIP-R)				Expressive Vocabulary (EOWPVT)			
	Unadjusted (N = 210)		Adjusted (N = 207)		Unadjusted (N = 209)		Adjusted (N = 203)	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Structural Quality								
Raw score	-0.30	1.87	1.55	1.35	-0.10	1.00	1.48**	0.57
Standardized score	-0.02	0.10	0.09	0.08	-0.01	0.07	0.11**	0.04
Educative Quality								
Raw score	0.86	0.93	0.66	0.94	0.49	0.92	1.28***	0.45
Standardized score	0.05	0.05	0.04	0.05	0.04	0.07	0.10***	0.03
Educator Sensitivity								
Raw score	0.34	0.49	0.16	0.56	0.01	0.50	0.71**	0.27
Standardized score	0.02	0.03	0.01	0.03	0.00	0.04	0.05**	0.02
Reading Quality								
Raw score	0.65	0.67	0.58*	0.29	0.52	0.44	0.40**	0.15
Standardized score	0.04	0.04	0.03*	0.02	0.04	0.03	0.03**	0.01
Overall Quality								
Raw score	1.25	1.29	1.40	0.85	0.71	0.79	1.76***	0.41
Standardized score	0.07	0.07	0.08	0.05	0.05	0.06	0.13***	0.03
Type of Difference	Program Fidelity							
	Receptive Vocabulary (ÉVIP-R)				Expressive Vocabulary (EOWPVT)			
	Unadjusted (N = 210)		Adjusted (N = 207)		Unadjusted (N = 209)		Adjusted (N = 203)	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
Structural Fidelity								
Raw score	9.12**	3.94	6.99*	3.44	3.76	3.24	5.10***	1.64
Standardized score	0.51**	0.22	0.39*	0.19	0.28	0.24	0.38***	0.12
Content Fidelity								
Raw score	6.63	9.34	3.32	4.92	5.89	6.63	4.33	3.99
Standardized score	0.37	0.52	0.18	0.27	0.44	0.49	0.32	0.30
Overall Fidelity								
Raw score	10.80*	6.22	7.33	4.64	5.47	4.85	5.97**	2.34
Standardized score	0.60*	0.35	0.41	0.26	0.41	0.36	0.44**	0.17

Note: Contrary to the analyses by treatment group, positive values for the DiD effects indicate a positive relationship between the fidelity element and the vocabulary index. Standardized scores are raw data converted to a Z-score with a mean of zero and a standard deviation of 1. Assignment to study groups for scale administration is used for the comparisons reported in this table. The ÉVIP-R was administered at 20 months and the EOWPVT, at 24 months. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Mediator Effect of Fidelity and Quality in the First Year

The logic of the mediation analysis presented here is the same as in Section 9.2.4 on first-year data analysis. Structural Fidelity and Structural Quality are the indices retained as the mediators that best predict child development. The two questions we ask are: (a) Does the fidelity/quality of project implementation account for all or part of the program's effect on outcomes? If not, we must conclude that other aspects of the program are responsible for the observed effects on the school readiness indicators (e.g., the Family Workshop Program). The results of a mediation analysis designed to answer these questions are reported in Tables 9.20 and 9.21.

Fidelity

An examination of Table 9.20 shows that Structural Fidelity accounts for a significant share of the program effect on all child outcomes (see indirect effect). It is observed that the program effect is reduced entirely or considerably (from 23% to 88.9%) when fidelity of program delivery is controlled for. Controlling for fidelity in the analysis reduces the difference between groups at 24 months (DinD) by 88.9% and 81.1% respectively for Communication and Expressive Vocabulary. The only significant direct effect is that associated with the ÉVIP-R, for which a significant share of the program effect has yet to be explained.

Table 9.20: Mediation Test Using Fidelity as Mediator of the Program Effect at 24 Months

Outcome	Direct Effect			Indirect Effect	
	Effect at 24 months	SE	% reduction	T-test	Sobel SE
Communication ¹	-0.03	0.22	-88.9%	2.49**	0.41
Self-Awareness ¹	-0.35	0.20	Supp.	3.41***	0.31
Cognition ¹	0.10	0.17	-23%	1.61*	0.45
Expressive Vocabulary ¹	-0.07	0.24	-81.1%	5.14***	0.39
EOWPVT-F (20 months) ¹	-0.09	0.16	-56.2%	3.02***	0.26
ÉVIP-R (24 months) ¹	-0.26**	0.11	Supp.	1.98**	0.41

Note: The negative values of effects comparing groups (i.e., direct effects) are a positive treatment effect (i.e., an advantage for the Program Daycare group). % reduction means the reduction in difference between the daycare groups in relation to the estimates in Table D2. Supp. = Suppression effect according to which that the program effect is stronger after controlling for fidelity. This means there is a low negative correlation between fidelity and another useful element of the program (e.g., the impact of the Family Workshop Program on parents). This effect is so low that a detailed interpretation is not relevant. ¹The mediator is Structural Fidelity. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$.

Quality

For ÉPE-AD outcomes, Educative Quality is selected as the mediator. For the ÉVIP-R and the EOWPVT, Structural Quality was chosen because it was stronger in the analyses presented in the previous section. Based on the results in Table 9.21, the program effect for all outcomes is partly attributable to implementation quality, except for Communication and the ÉVIP-R because their indirect effect is non-significant. We find that the direct program effect on Communication, Cognition and the EOWPVT-F is entirely attributable to Daycare Program implementation. For

the other indices, part of the program effect has yet to be explained because estimates of the direct effect are significant for these outcomes.

Table 9.21: Mediation Test Using Quality as Mediator of the Daycare Program Effect at 24 Months

Outcome	Direct Effect			Indirect Effect	
	Effect at 24 months	SE	% reduction	T-test	Sobel SE
Communication ¹	-0.16	0.11	-40.7%	1.13	0.04
Self-Awareness ¹	-0.28***	0.07	-9.8%	1.38*	0.04
Cognition ¹	-0.05	0.10	-61.5%	1.51*	0.05
Expressive Vocabulary ¹	-0.32***	0.09	-13.5%	1.55*	0.07
EOWPVT-F (20 months) ²	-0.14	0.09	-26.3%	1.79**	0.04
ÉVIP-R (24 months) ²	-0.22***	0.06	Supp.	1.08	0.06

Note: The negative values of effects comparing groups (i.e., direct effects) are a positive treatment effect (i.e., an advantage for the Program Daycare group). % reduction means the reduction in the difference between daycare groups in relation to the estimates in Table D2. Supp. = Suppression effect according to which that the program effect is stronger after controlling for quality. This means there is a low negative correlation between quality and another useful element of the program (e.g., the impact of the Family Workshop Program on parents). This effect is so low that a detailed interpretation is not relevant. ² The mediator is Structural Quality. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$.

Summary: Mediator Effect of Fidelity and Quality

The analyses show that Daycare Program fidelity and quality are associated with child development. Fidelity in particular is especially useful in explaining the effect of the tested program. The particularly strong mediation effects observed with fidelity can be attributed to the close association between the tested program and fidelity. However, the quality indices include general aspects that are not necessarily specific to the tested program. For the ÉVIP-R alone, fidelity does not entirely explain the effect of the tested program, as suggested by the significant direct effect. For this outcome alone, part of the tested program's effect must yet be accounted for with other variables (e.g., capturing the impact of the Family Literacy Workshops on parents).

9.3.5. Analyses by Linguistic Profile

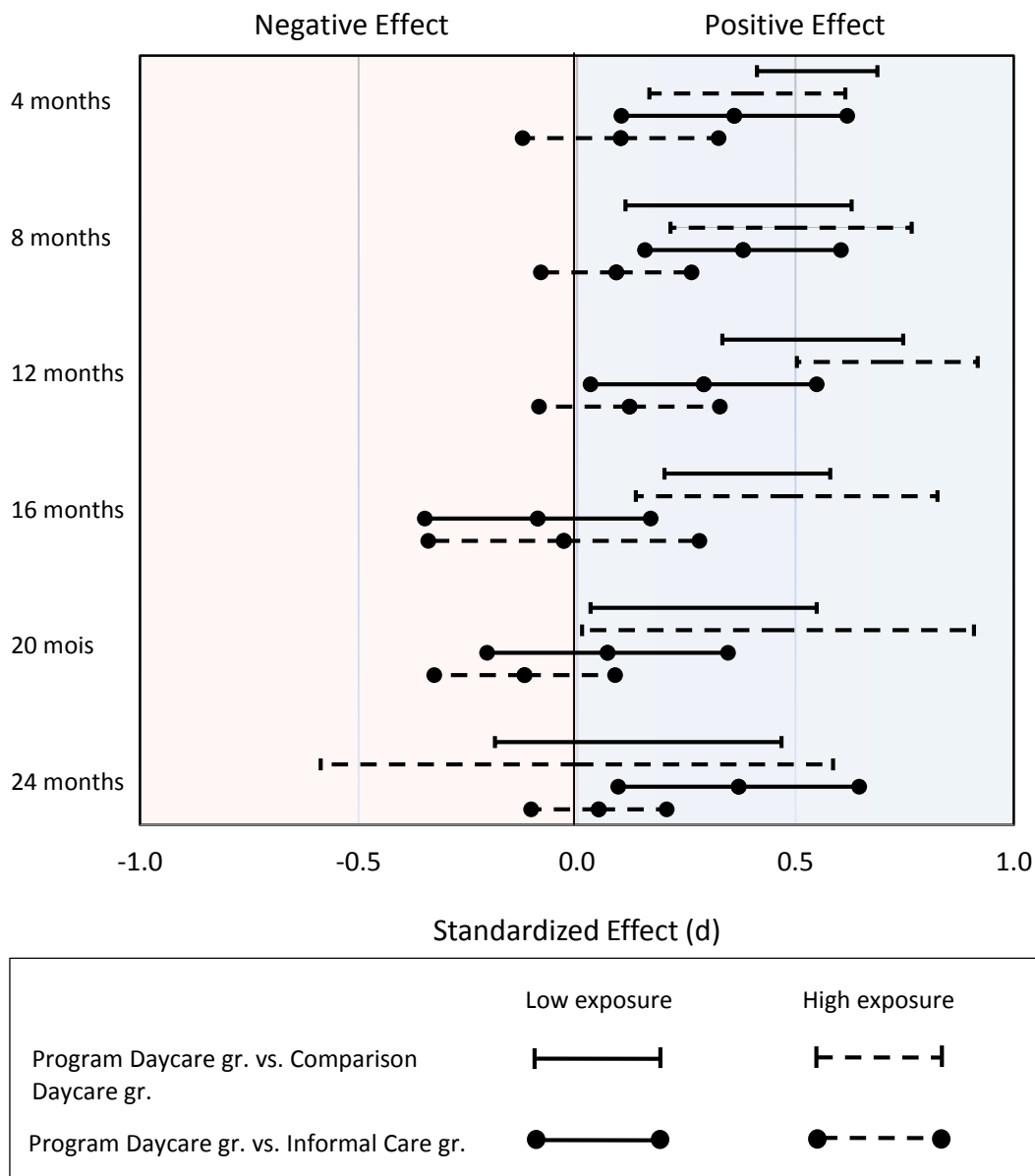
The analysis by linguistic profile is reported in Figures 9.7 to 9.10 for ÉPE-AD outcomes. Figure 9.11 shows the analysis results for other measures of vocabulary (ÉVIP-R and EOWPVT). Appendix I provides a detailed breakdown of all these results. ***Note that in the figures, a positive value indicates that the tested program has a positive effect on development.*** The results are described below for each outcome.

Communication Scale

The confidence intervals for the standardized effects of the tested program (DinD) at each assessment period are presented in Figure 9.7. The figure distinguishes between households based on the child's level of exposure to French (high or low). A visual inspection of the figure shows that, as anticipated, the effects for this variable tend to be smaller for the high-exposure condition, with the moderating effect of exposure most marked at 12, 16 and 20 months.

For children with low exposure, the program impact is significant for both comparison groups at 4, 8, 12, 16, and 20 months. For the high-exposure group, the program impact is never significant for the Informal Care group, but is significant for the Comparison Daycare group at 4, 8, 12 and 24 months. Overall, the program impact for the Communication scale is more stable for children with low exposure, although positive effects are also observed for the high-exposure group. It is interesting to note that the tested program's effect is greater than that of an informal care setting only with respect to Communication development in a context of low exposure to French (outside daycares).

Figure 9.7: Confidence Intervals for the Program's Standardized Effect (DinD) on Communication by Household Type

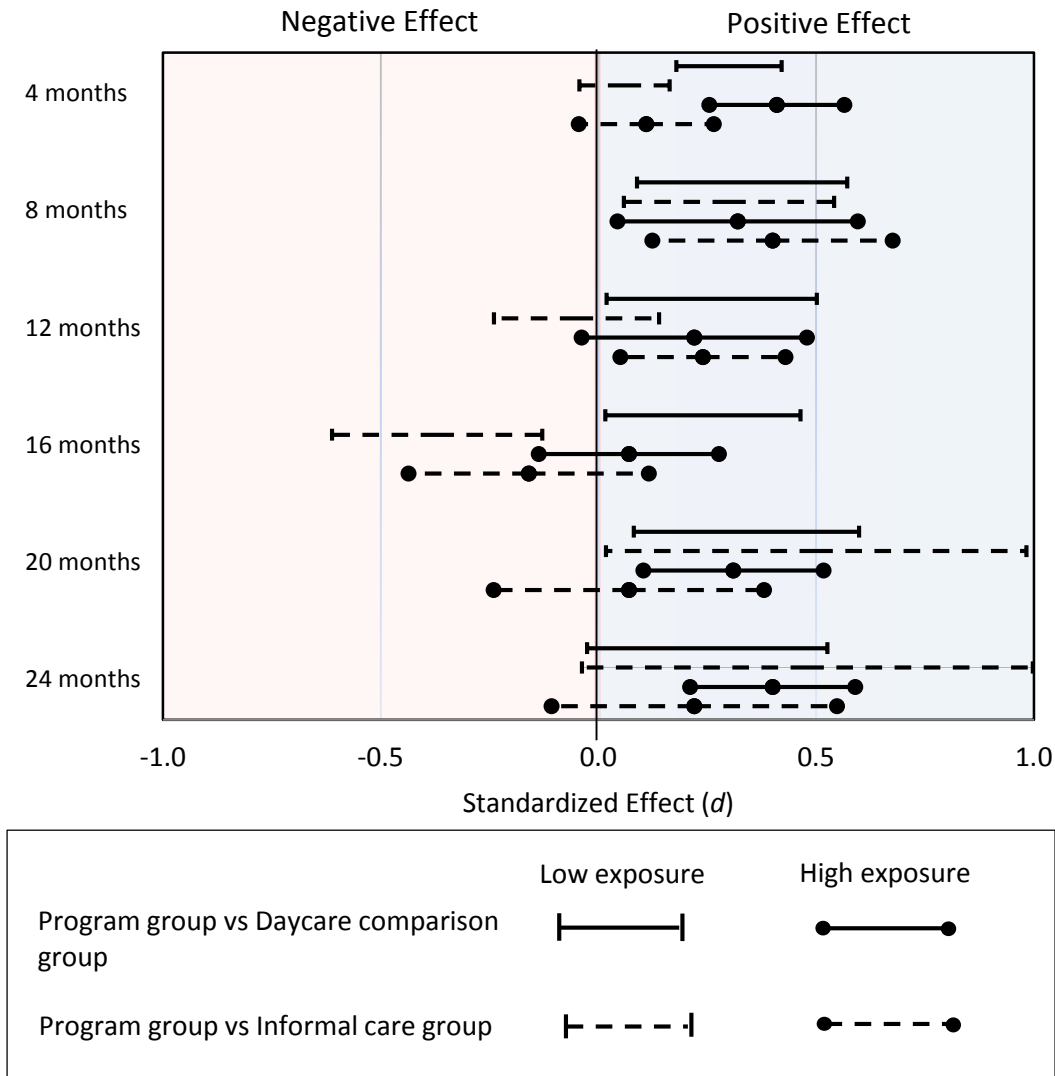


A line represents a confidence interval of 90%. Confidence intervals that do not include the zero value indicate a 10% significance level.

Self-Awareness Scale

The confidence intervals for the standardized effects of the tested program (DinD) at each assessment period are presented in Figure 9.8. The figure distinguishes between low-exposure and high-exposure households. A visual inspection of the results does not show a clear relationship between a family's linguistic profile and the size of the observed effects. This may be due to the fact that this scale targets Communication and Cognition elements, which may cancel out the program effects differentiated by linguistic profile.

Figure 9.8: Confidence Intervals for the Program's Standardized Effect (DinD) on Self-Awareness by Household Type



A line represents a confidence interval of 90%.

Confidence intervals that do not include the zero value indicate a 10% significance level.

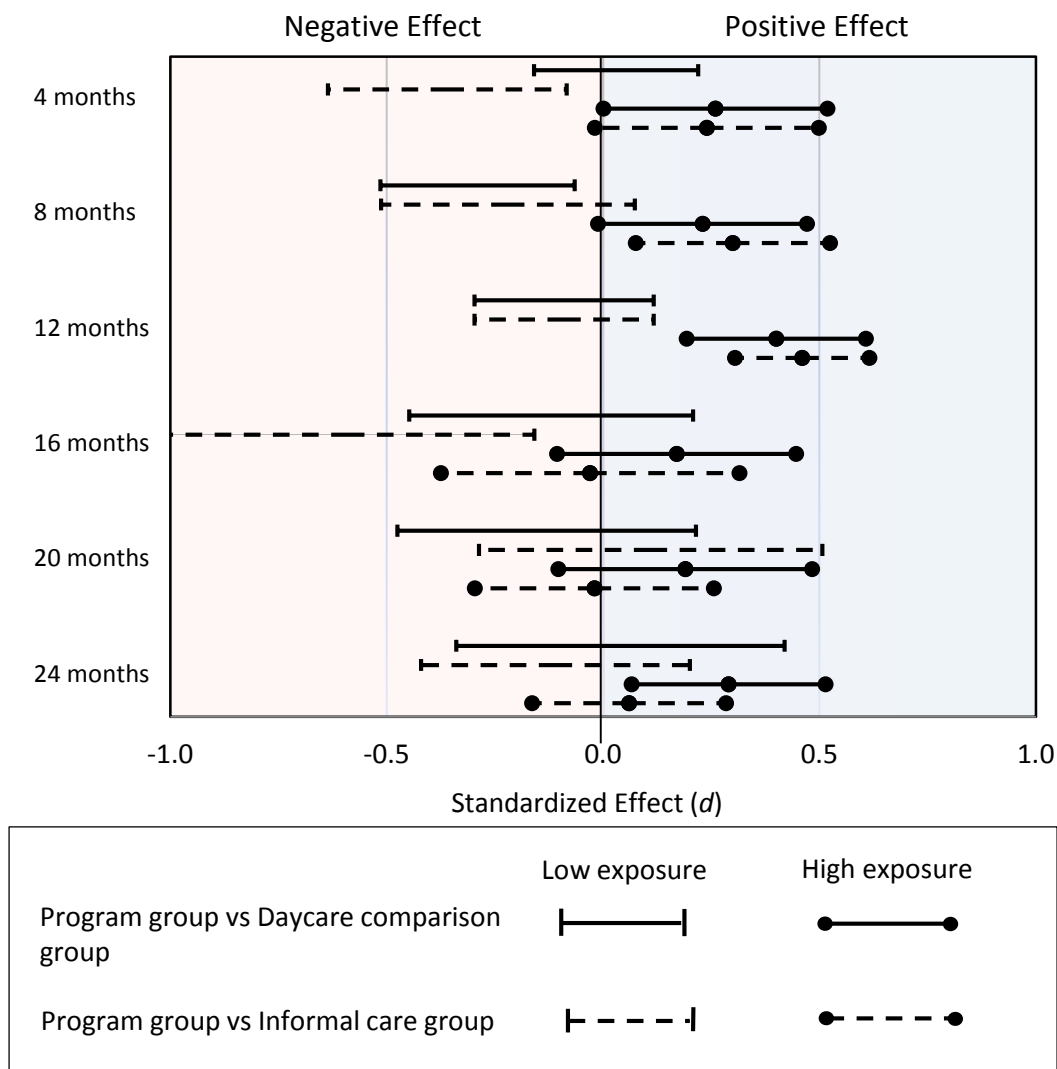
Cognition Scale

The confidence intervals for the standardized effects of the tested program (DinD) at each assessment period are presented in Figure 9.9. Yet again, a distinction is made between households with low and high exposure to French. A visual inspection of the results reveals a clear pattern: children with high exposure benefit more in terms of cognitive development.

For children with high exposure to French, the effect in relation to the Comparison Daycare group is significant at 4, 12 and 24 months. In relation to the Informal Care group, the effect for the same group of children is significant at 8 and 12 months. However, a positive and significant program effect is never observed for children with low exposure to French. In fact, for such children, the tested program appears to have a negative effect on their cognitive development for

several assessment periods. Fortunately, this effect seems temporary since it is not observed at 20 and 24 months. What would explain this effect? Insisting on French-only communication at program daycares may be an obstacle for certain children with low exposure to French, at least with respect to the acquisition of cognitive skills. Again, this effect appears temporary and offset by language skill acquisitions.

Figure 9.9: Confidence Intervals for the Program's Standardized Effect (DinD) on Cognition by Household Type



A line represents a confidence interval of 90%.

Confidence intervals that do not include the zero value indicate a 10% significance level.

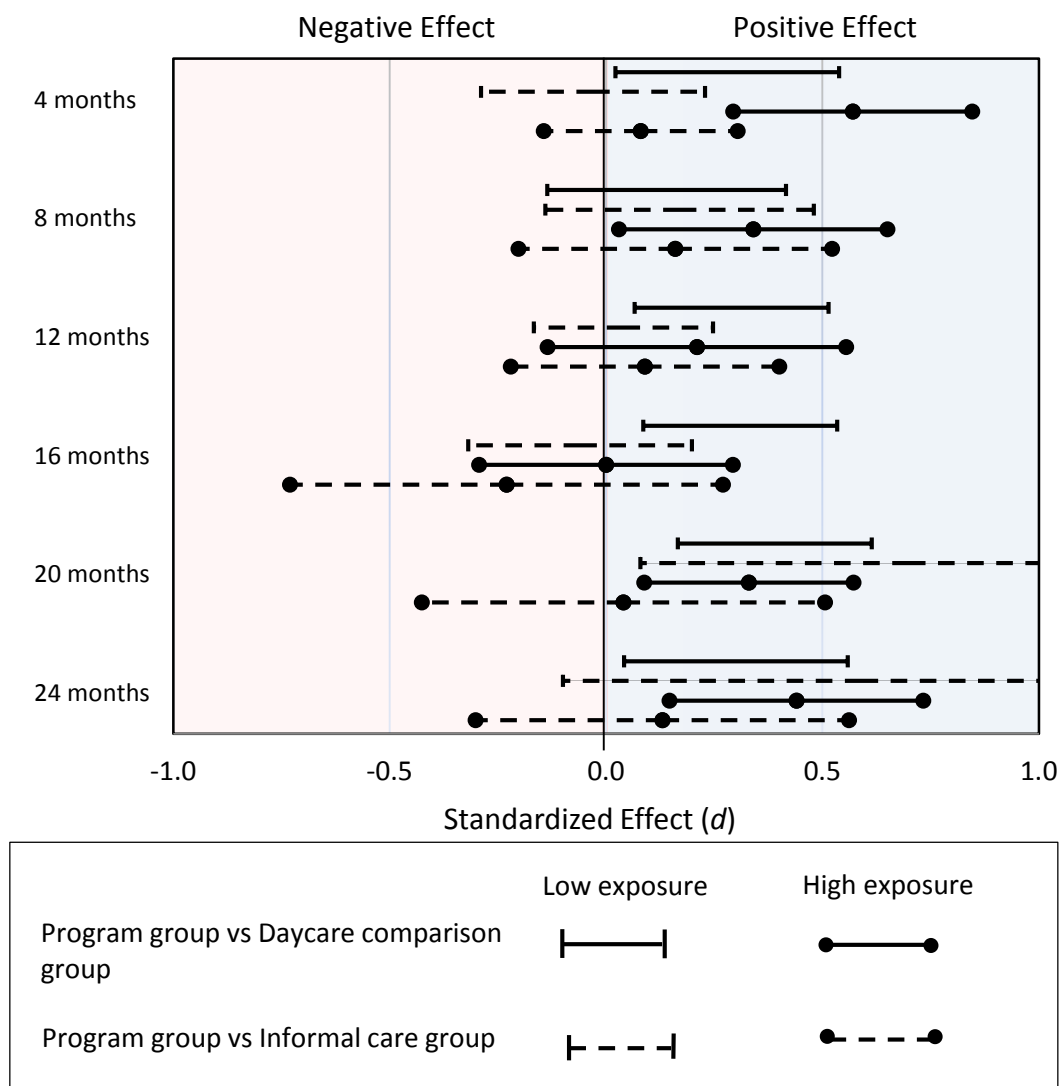
Expressive Vocabulary Scale

The confidence intervals for the standardized effects of the tested program (DinD) at each assessment period are presented in Figure 9.10. Yet again, a distinction is made between low-exposure and high-exposure families. Contrary to our expectations, at 4 months we observe a greater program effect among children with high exposure to French. In the second year, the

anticipated effect is observed: children with low exposure experience greater program effects for the development of expressive vocabulary.

For children with low exposure, the program effect for the Comparison Daycare group is significant at 4, 12, 16, 20 and 24 months. For the same group, the program effect is significant for the Informal Care group at 20 months. In contrast, among children with high exposure, the program effect is significant for the Comparison Daycare group at 4, 8, 20 and 24 months. For the same group, the program effect for the Informal Care group is never significant. In short, the program effect for Expressive Vocabulary, as measured by that scale, appears not to be closely tied to a child's exposure to the French language. That said, the program impact seems to be more stable for this measure among children with low exposure.

Figure 9.10: Confidence Intervals for the Program's Standardized Effect (DinD) on Expressive Vocabulary by Household Type



A line represents a confidence interval of 90%.

Confidence intervals that do not include the zero value indicate a 10% significance level.

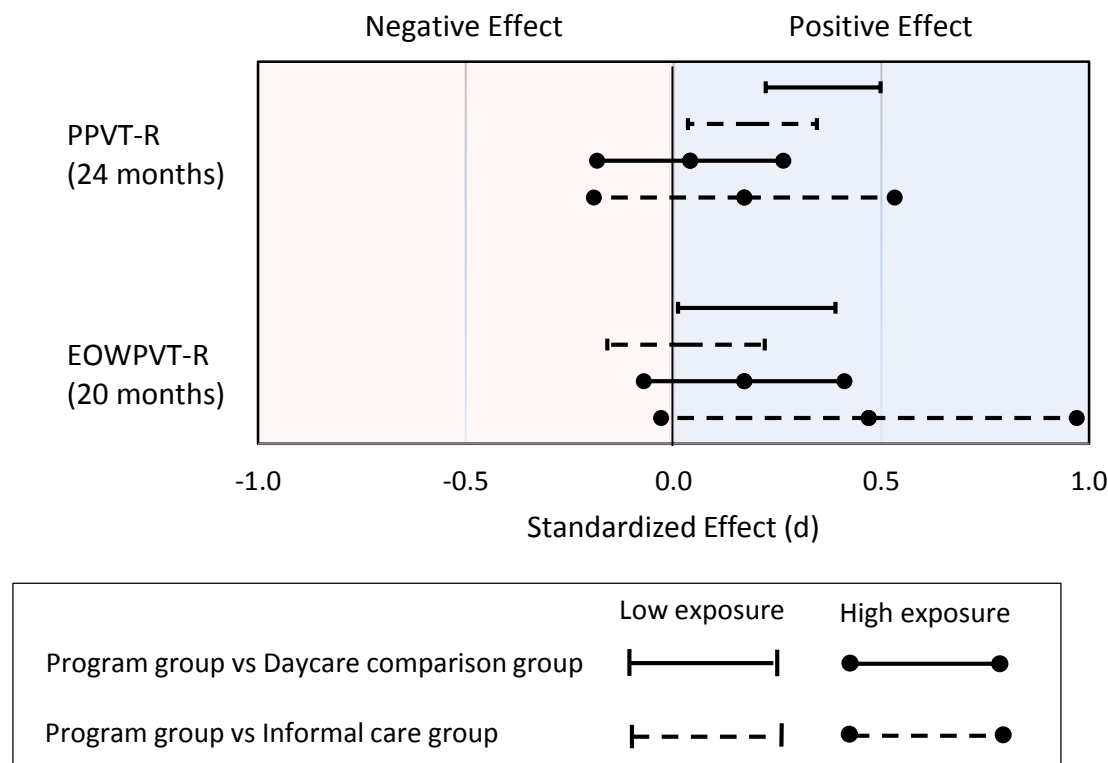
ÉVIP-R and EOWPVT-F Scales

The confidence intervals for the standardized effects of the tested program (ANCOVA) for the ÉVIP-R and the EOWPVT are presented in Figure 9.11. As with ÉPE-AD outcomes, a distinction is made between children with low and with high exposure. All confidence intervals include positive values indicating that the hypothesis of a positive program impact is never invalidated.

As regards Receptive Vocabulary, the results show that, as anticipated, children with low exposure benefit more from the program than do those with high exposure. Significant effects (those without the zero value in the confidence intervals) are observed only for the subsample of low-exposure children. As regards Expressive Vocabulary, children with low exposure develop

their vocabulary more quickly than their peers in the Comparison Daycare group. No effect is observed for children with high exposure. The hypothesis that program impact on French vocabulary varies based on exposure to the language appears to be confirmed. However, this conclusion must be nuanced by the fact that all confidence intervals overlap: one effect is no different than another (statistically), although certain effects differ significantly from zero.

Figure 9.11: Confidence Intervals for the Program's Standardized Effect (ANCOVA) on the ÉVIP-R and the EOWPVT-F by Household Type



A line represents a confidence interval of 90%.

Confidence intervals that do not include the zero value indicate a 10% significance level.

Summary: Analyses by Linguistic Profile

The results tend to support the hypotheses proposed at the beginning of this section. The program effects on language skills tend to be significantly more pronounced among children whose exposure to French at home is low. However, program effects on cognitive skills (Cognition) tend to be significantly more pronounced among children from households characterized by high exposure to the French language. However, for other outcomes, the confidence intervals for effects differentiated by level of language exposure usually overlap, indicating that the program effect is nevertheless similar whatever the linguistic profile. Although a larger participant sample would be necessary to obtain a definitive answer, the results tend to support the original hypothesis.

9.4. SUMMARY

The purpose of this chapter was to evaluate the combined effect of the two components of the tested program on children's school readiness. Analysis of the data collected over the two years of program delivery showed positive and significant program effects for various dimensions of children's school readiness. The comparison of treatment groups identified an advantage for children who received the program versus their peers in terms of the dimensions targeted by the ÉPE-AD scales: Communication, Cognition, Receptive Vocabulary and Expressive Vocabulary. The positive and significant program effect on language-skill development is reproduced for Receptive Vocabulary (ÉVIP-R) and Expressive Vocabulary (EOWPVT). Whatever the comparison group, children in the Program Daycare group appear better prepared for a French-language school environment. This is the case when they begin junior kindergarten (at the 12-month assessment) and senior kindergarten (at the 24-month assessment).

Although significant effects are observed, it is important to determine whether they are important. The size of the effects is relatively constant over the two years of the project, ranging, based on Cohen's (1988) benchmarks, from a small (0.20) to a medium (0.50) effect. Expressed otherwise, according to the data of Hill and colleagues (2008), the size of these effects is tantamount to accelerating the children's development and learning by several months. The literacy disadvantage for 15-year-old minority Francophones happens to be roughly the same magnitude (Allen, 2004). That is not to say that the tested intervention would alone be sufficient to offset the observed achievement gap in literacy between minority Francophones and their Anglophone peers. Its aim is simply to give concrete expression to the size of the reported effects.

The robustness of the results is examined in several ways. The validity of the analyses is first verified by reproducing the results of parametric analyses through non-parametric analyses. Next, the credibility of the findings from the comparison of treatment groups is enhanced through a series of additional analyses associating the intensity of program exposure (dosage, fidelity) to the observed effects. The result based upon which the advantage of the Program Daycare group seems to depend on dosage and implementation fidelity heightens our confidence that the observed effects stem from the tested program (more specifically, the Daycare Program) rather than other factors. While most of the tested program's effect is accounted for by Daycare Program implementation, the contribution of the family literacy component (see Chapter 10) has yet to be determined.

Finally, the result that predicts a program effect differentiated by Household Type (low or high exposure to French) consolidates the study's internal validity. Thus, children from households characterized by low exposure to French show greater and more significant gains in language-skill development. Children from households characterized by high exposure to French benefit mainly in terms of their cognitive development. These results highlight the program's beneficial effects regardless of the linguistic profile of the sample and of the minority Francophone population. These points are discussed in greater detail in Chapter 11.

10. Impact of the Family Literacy Component

Chapter 10 concerns the results of analyses assessing the contribution of the Family Literacy Program. Two research questions are considered: a) Did the Family Literacy Workshops affect parents' attitudes and behaviours? and b) To what degree are the changes observed among the parents responsible for the tested program's effect on the children? These questions are addressed respectively in Sections 10.1 and 10.2.

The following points should be noted:

1. The analyses reported in Section 10.1.2 are based on four communities: Orleans, Cornwall, Durham and Edmundston. These analyses include all three groups of children in the study. These four communities alone were included in the analyses in order to maximize the comparability of results with those stemming from the Daycare Program analyses reported in Chapter 9.
2. The analyses reported in Section 10.1.3 are based on the project's six communities: Edmonton, Saint John, Orleans, Cornwall, Durham and Edmundston. The study's outcomes are based only on children in the Program Daycare group.
3. For most analyses, the first cohort of participants consists of families from the four communities (six for the analyses in Section 10.1.3), while the second cohort consists exclusively of families from Orleans and Cornwall.

10.1. DIRECT IMPACT ON PARENTS

This section aims to determine whether the program offered as part of the Family Literacy Workshops has a positive effect on the attitudes and behaviours of parents in the Program Daycare group.

10.1.1. Variables Retained for the Analysis

In the analyses that follow, two classes of outcomes are examined. Data for the first class of outcomes were collected through a post-program survey administered only to parents in the Program Daycare group. The examined outcomes are: questions to determine parents' opinions regarding their child's development (Opinions), the frequency at which literacy activities are modeled (Modeling), the parents' sense of self-efficacy (Self-Efficacy), self-assessment of the parents' own knowledge of general child development and the resources available to them in the community (Knowledge), and their sense of belonging to the Francophone community (Sense of Belonging).⁷⁹ Data for the second class of outcomes were collected through the baseline survey and the follow-up surveys administered to all parents participating in the project. These data are analyzed in the same way as the data from the children's assessment periods (i.e., DiD estimate and non-parametric analysis). The examined outcomes include Frequency and Language of Literacy Activities.

⁷⁹ Exceptionally, the number of participants is maximized by including children from the six project communities for the analysis of certain parent outcomes.

10.1.2. Analyses by Group

Two outcomes were collected from parents in all three treatment groups at each follow-up survey: Frequency of Literacy Activities and Language of Literacy Activities. This section presents the results of an analysis intended to determine how the Family Literacy Program affects these outcomes. Complete results for the initial model and the adjusted model are presented in Appendix J. The result of the program's impact analyses is presented in Figure 10.4 for Frequency of Literacy Activities and in Figure 10.8 for Language of Literacy Activities.

Note that in the *First Cohort Findings Report* (2014), the result of the impact analyses revealed that an increase in Frequency of Literacy Activities was impossible due to a ceiling effect.⁸⁰ This information is therefore important to consider in order to properly interpret the result of the impact analyses for these outcomes. We therefore begin by examining the empirical distribution of Frequency of Literacy Activities and Language of Literacy Activities for three assessment periods: at baseline (Figures 10.1 and 10.5), at 12 months (Figures 10.2 and 10.6), and at 24 months (Figures 10.3 and 10.7).

Frequency of Literacy Activities

This section presents the results of three analyses: (1) the empirical distribution of the Frequency of Literacy Activities scale; (2) the results of the impact analyses; and (3) the result of a non-parametric confirmation of the results obtained with this scale.

Empirical Distributions

The empirical distribution of the Frequency of Literacy Activities scale is presented in Figures 10.1, 10.2 and 10.3 for the baseline, 12-month and at 24-month assessment periods respectively. Note that each choice of answer is assigned a numerical value from 1 to 5 (where 1 = Never and 5 = Several times a day). The data reported in the figures reflect the parents' "average" response to each item of the scale.⁸¹ To facilitate figure interpretation, the label associated with each numerical value has been provided. An examination of the figures allows us to conclude that this variable has an approximately normal distribution for the study sample.

⁸⁰ The ceiling is the result of an unrealistic choice of answers for the upper spread of the distribution. Few parents have the time to perform five separate literacy activities per day, at a frequency of three or more times each.

⁸¹ Mean scores are rounded to the closest number in the figures. However, the impact analyses are based on unrounded scores.

Examination of Figures 10.2 and 10.3 reveals that the frequency of Literacy Activities increases from year to year. In particular, we note that the distribution of Figure 10.3 is skewed to the right in comparison with that of Figure 10.2. These two distributions are skewed to the right in comparison with Figure 10.1. The skew is more pronounced for the Program Daycare group (G1) due to a higher percentage of respondents who say they perform literacy activities once or more a day.

Figure 10.1: Distribution of Frequency of Literacy Activities at the Baseline

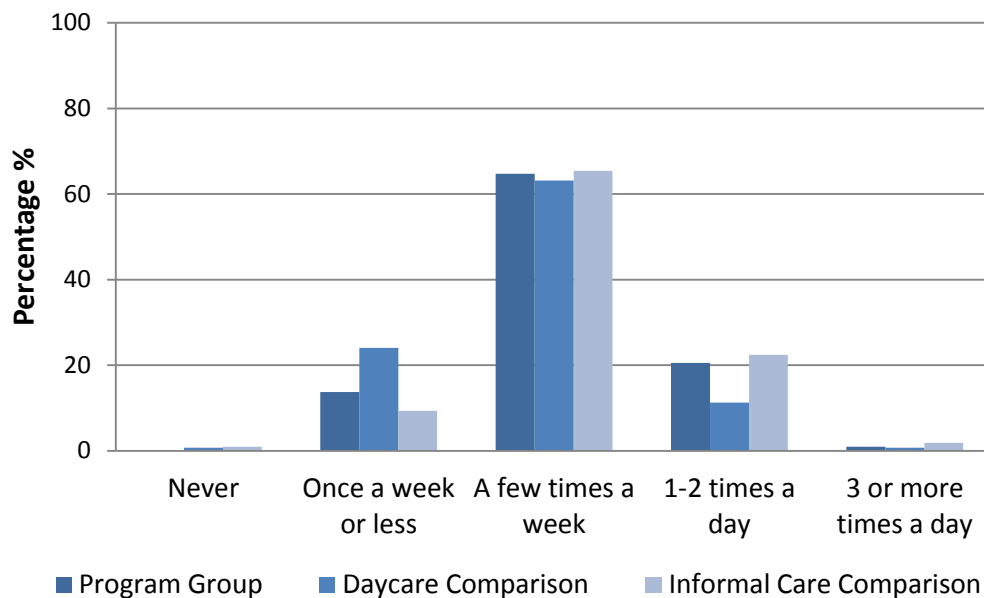


Figure 10.2: Distribution of Frequency of Literacy Activities at 12 Months

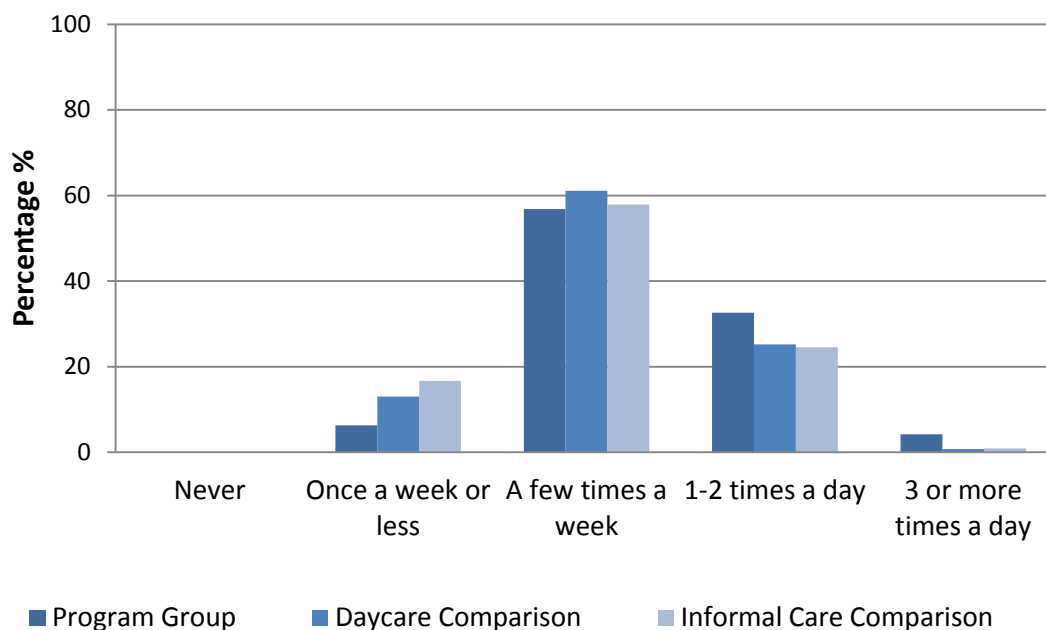
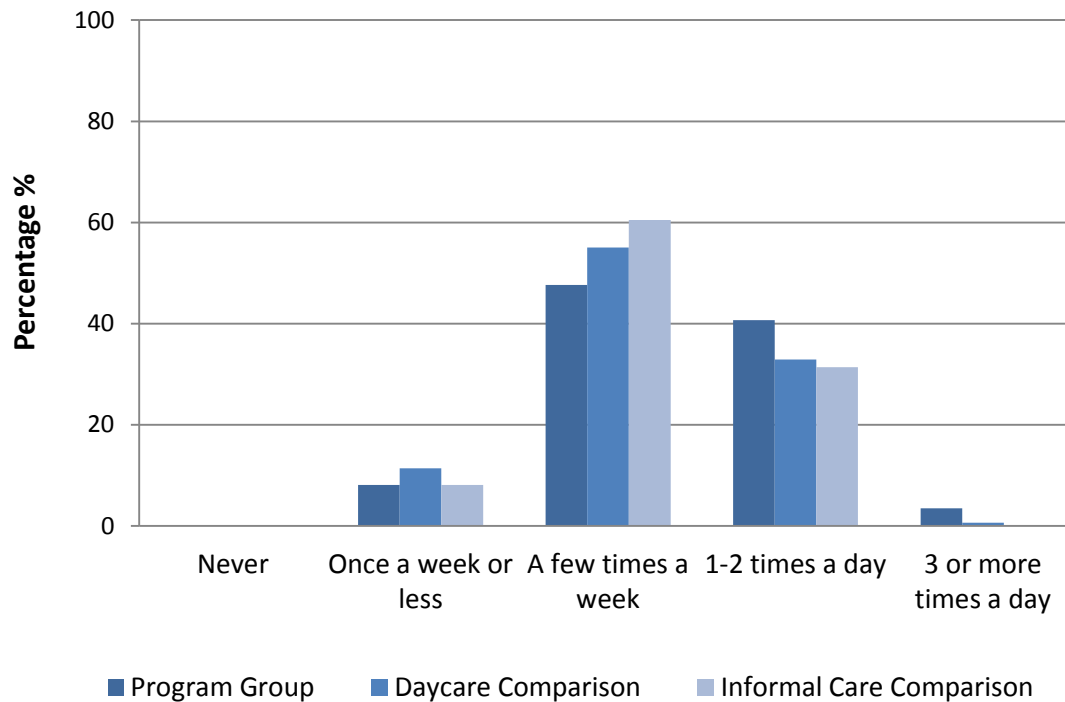


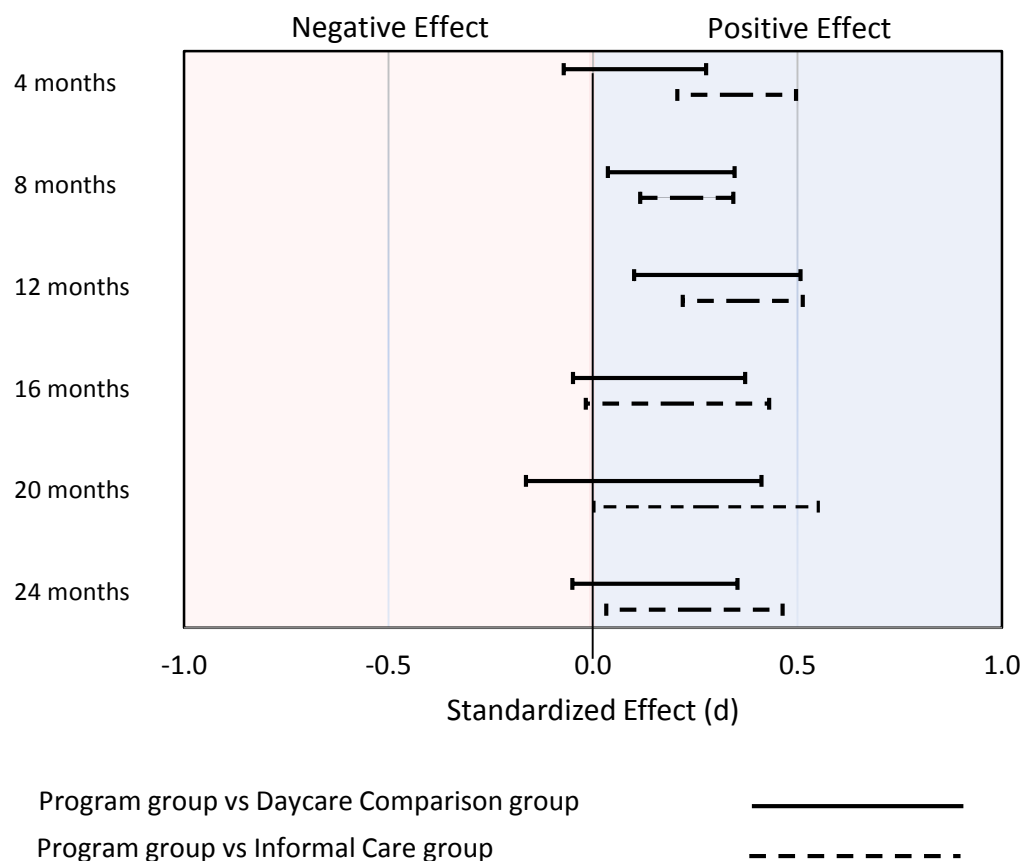
Figure 10.3: Distribution of Frequency of Literacy Activities at 24 Months



Adjusted Model

The confidence intervals for the standardized DinD effects calculated based on the adjusted model are reported in Figure 10.4. For the Informal Care group, we observe a program effect emerging at 4 months and persisting until 24 months. For the Comparison Daycare group, a significant program effect is observed at 8 and 12 months. This result gives empirical support to the hypothesis that the Family Literacy Program's positively influence on parental behaviour. Verification of this hypothesis is a condition necessary yet insufficient to assert that this component of the tested program contributes positively to child development. The second condition, that changes observed among parents be predictive of child development, is the object of analyses reported in Section 10.2.

Figure 10.4: Effect of the Family Literacy Program on Frequency of Literacy Activities



A line represents a confidence interval of 90%.

Confidence intervals that do not include the zero value indicate a 10% significance level.

Non-Parametric Cross-Validation

The non-parametric analysis of data was conducted using the same strategy as in Chapter 9, Section 9.31. The result is expressed as the probability of being classified in a higher quintile: the success category. The odds ratio for the DinD effects corresponding to the effects reported in Figure 10.4 is reported in Table 10.1.

An examination of the confidence intervals reported in Table 10.1 partially confirms the significance of the DinD effects for comparison of the Program Daycare group with the Informal Care group. The DinD estimates for this effect are significant and in favour of the Program Daycare group in the first year and at 24 months. However, the non-parametric analysis partially confirms the significance of the DinD effects for comparison of the Program Daycare group with the Comparison Daycare group. The odds ratios are always in favour of the Program Daycare group, but only the effect at 12 months is significant, based on the non-parametric analysis. An effect representing a change in the child's environment lasting several months may be enough to create positive effects for the children.

Table 10.1: Odds Ratio for the Standardized DiD Effects of the Adjusted Model for Frequency of Literacy Activities

Effect	Odds Ratio	90% Confidence Intervals	
		Lower Bound	Upper Bound
G1 vs. G2			
4 months	0.69	0.40	1.19
8 months	0.79	0.54	1.14
12 months	0.55	0.35	0.87
16 months	0.89	0.41	1.92
20 months	0.88	0.41	1.88
24 months	0.72	0.42	1.23
G1 vs. G3			
4 months	0.43	0.27	0.69
8 months	0.71	0.55	0.91
12 months	0.55	0.38	0.80
16 months	0.61	0.29	1.29
20 months	0.59	0.27	1.30
24 months	0.52	0.28	0.95

Language of Literacy Activities

This section presents the results of the following three analyses: (1) the empirical distribution of the Language of Literacy Activities scale; (2) the results of the impact analyses; and (3) the result of a non-parametric confirmation of the results obtained with this scale.

Empirical Distributions

The empirical distribution of the Language of Literacy Activities scale is presented in Figures 10.5, 10.6 and 10.7 for baseline, 12-month and 24-month assessment periods respectively. Note that each choice of answer for this scale is assigned a numerical value from 1 to 5 (where 1 = Never in French and 5 = Only in French). The data reported in the figures reflect parents' "average" response to each item of the scale.⁸² To facilitate figure interpretation, the label associated with each numerical value is provided. An examination of the figures allows us to conclude that the distribution for this variable is asymmetric, since the chosen language for literacy activities is usually French.

⁸² Mean scores are rounded to the closest number in the figures. However, the impact analyses are based on unrounded scores.

Moreover, a comparison of the distributions does not reveal any clear program effect. The effect, if any, is found in the lower spread of the score distribution for this scale. We observe that it includes the smallest percentage of respondents who report using English and /or another language during literacy activities. At 24 months, when children begin senior kindergarten (Figure 10.7), a slight trend toward greater use of English and /or another language during literacy activities is observed.

Figure 10.5: Distribution of Language of Literacy Activities at the Baseline

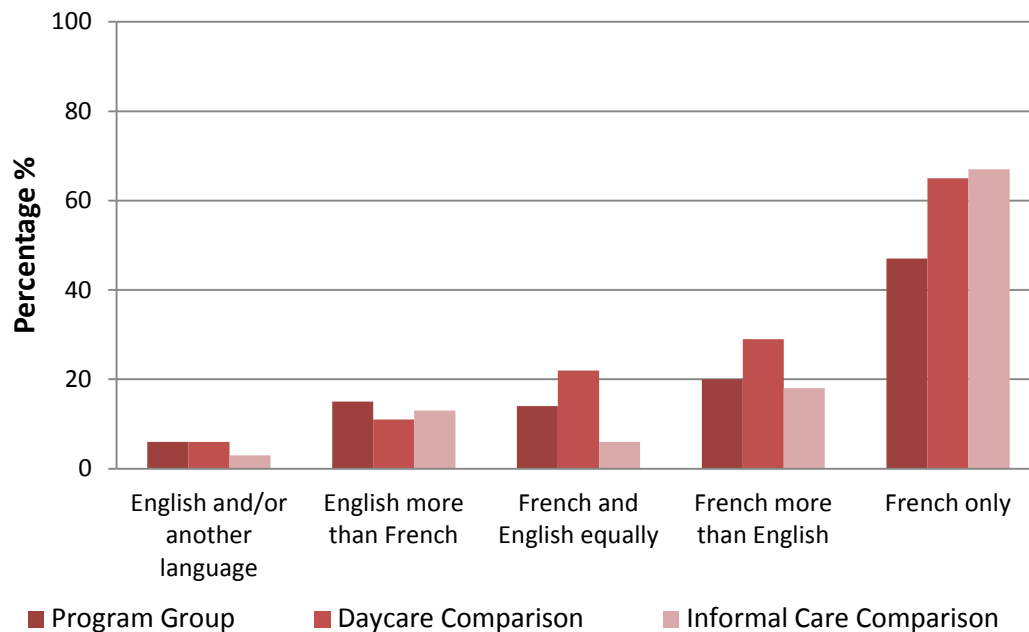


Figure 10.6: Distribution of Language of Literacy Activities at 12 months

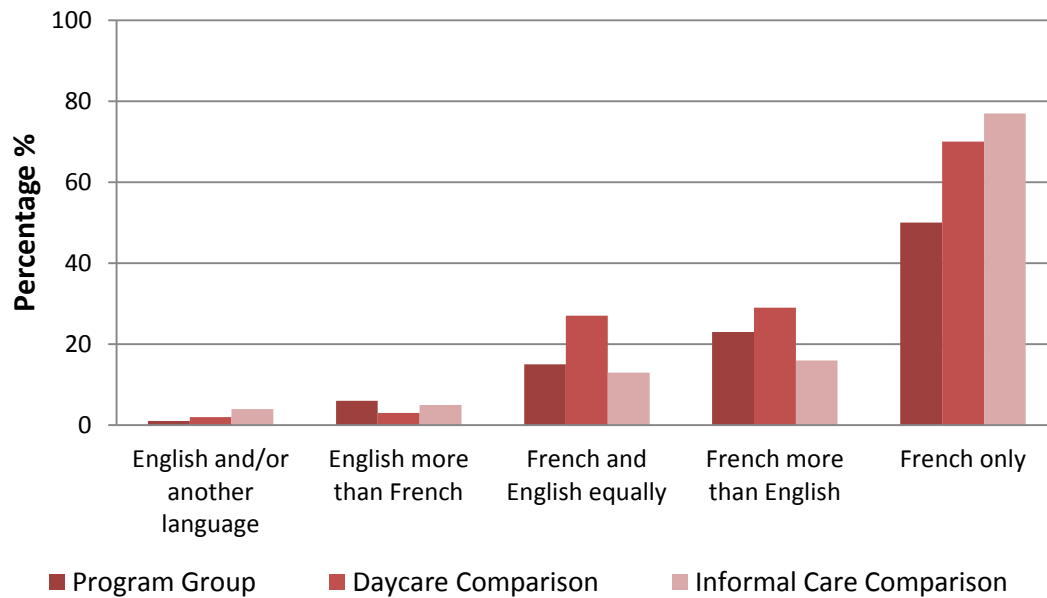
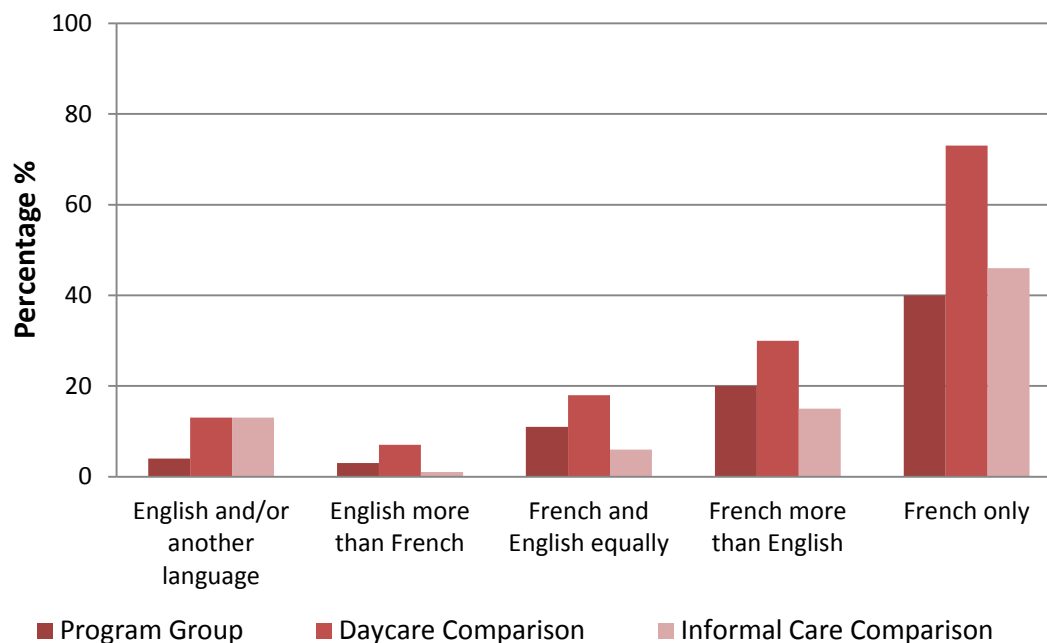


Figure 10.7: Distribution of Language of Literacy Activities at 24 Months

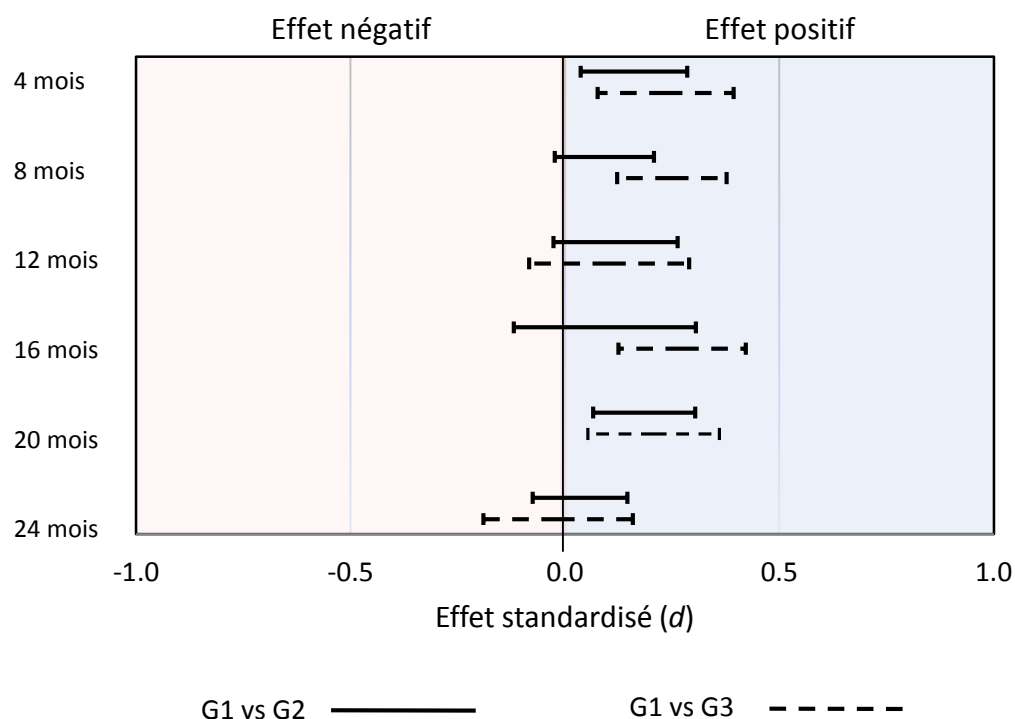


Adjusted Model

The confidence intervals for the standardized DinD effects calculated based on the adjusted model are presented in Figure 10.8. For the Comparison Daycare group, a program effect emerges at 4 months and reappears at 20 months. For the Informal Care group, a program impact

is observed at assessments other than 12 months and 24 months. Regardless of the comparison group used as a benchmark, the program effect at 24 months appears to be completely mitigated. It remains to be verified whether the effects observed from 4 to 20 months are sufficient to result in positive effects for the children.

Figure 10.8: Effect of the Tested Program on Language of Literacy Activities



A line represents a confidence interval of 90%. Confidence intervals that do not include the zero value indicate a 10% significance level.

Non-Parametric Cross-Validation

The non-parametric analysis of data was conducted using the same strategy as that used to analyze Frequency of Literacy Activities. The result is expressed as the probability of being classified in a higher quintile: the success category. The odds ratio for the DinD effects reported in Figure 10.8 is reported in Table 10.2.

The hypothesis that the Family Literacy Program significantly affects the linguistic behaviours of parents is invalidated by the non-parametric analysis. All the confidence intervals reported in Table 10.2 include the “1” value, thereby indicating the absence of a significant program effect. These results do not reproduce the DinD effects reported in Figure 10.8, perhaps due to the asymmetric distribution of this variable. Additional non-parametric analyses (unreported) show that the program effect is found mostly in the second and third quintiles. We consider the non-parametric analysis most credible⁸³ and conclude that the program has no significant effect on parents’ linguistic behaviours.

⁸³ The non-normality of distribution for this variable may negatively affect the results of the parametric analyses.

Table 10.2: Odds Ratio for the Standardized DiD Effects of the Adjusted Model for Language of Literacy Activities

Effect	Odds Ratio	90% Confidence Intervals	
		Lower Bound	Upper Bound
G1 vs. G2			
4 months	0.78	0.44	1.38
8 months	0.91	0.50	1.67
12 months	0.99	0.43	2.30
16 months	0.82	0.40	1.67
20 months	0.85	0.62	1.14
24 months	0.78	0.53	1.16
G1 vs. G3			
4 months	0.69	0.44	1.08
8 months	0.98	0.61	1.56
12 months	1.29	0.70	2.39
16 months	0.64	0.37	1.11
20 months	0.80	0.52	1.23
24 months	0.88	0.57	1.35

10.1.3. Analyses Concerning Parents in the Program Daycare Group

The following outcomes were measured only among parents in the Program Daycare group and their analysis is based on six communities. These data are analyzed based on three complementary strategies. First, participating parents (i.e., those who participated in three or more workshops, $n = 116$) are compared with parents classified as non-participants (i.e., those who participated in two or fewer workshops, $n = 14$).⁸⁴ Next, we examine whether attendance by participating parents is associated with the magnitude of program impact. Finally, we report the result of a series of analyses in connection with the fidelity and quality indices for implementation of the Family Literacy Workshops as well as the impact of the Family Literacy Program on parent outcomes.

⁸⁴ A statistical power analysis concludes that this comparison with uneven sample size can detect true effects of 0.20, 0.50 and 0.80 respectively 28%, 68% and 94% of the time. A 10% significance level is assumed.

The small sample of non-participants makes it impossible to “statistically” correct the pre-program differences among the groups. Note that the t-test results indicate a significant difference between the groups (participants and non-participants) only for Social Capital and for Sense of Belonging to the cultural group.

Analyses by Participation (Participants Versus Non-Participants)

In the next subsections, the effect of parental participation or non-participation in the Family Literacy Workshops is examined for every parent outcome.

Parents’ Opinions

Parents were asked to answer four questions concerning their beliefs about factors linked to their child’s development. The analysis results for three of the four questions are presented in Tables 10.3, 10.4 and 10.5 respectively. The fourth statement, “I believe that my child is too young to hear stories read to him/her,” was not analyzed because all parents said they absolutely disagreed with that statement.

In Table 10.3, no significant pre-program difference is observed between participating parents and non-participating parents. The size of the effect ranges from small to medium (i.e., from an absolute value of 0.10 to 0.40). After the Family Literacy Workshops, we observe a non-significant trend in the anticipated direction (i.e., a greater belief among participating parents that they are their child’s first educator).

The DinD estimator shows that the Family Literacy Workshops have a medium effect ($d = 0.54$ standard deviations), although non-significant according to the t-test. This effect stems mainly from a slight negative and non-significant change observed among non-participants ($d = -0.38$) along with a slight non-significant increase observed among participants ($d = 0.15$).

Table 10.3: Summary of the Family Literacy Program’s Effect on Parents’ Opinions: The Parent as a Child’s First Educator

	Means (SD)		Change		T-test
	Pre	Post	Diff (Post- Pre)	Diff (d)	
Participants (P)	2.91 (0.89)	3.05 (0.88)	0.14 (0.87)	0.15	1.70
Non-participants (NP)	3.21 (0.89)	2.86 (0.95)	-0.36 (1.15)	-0.38	-1.16
Difference (P – NP)	-0.30	0.19	0.49		
Standardized difference (d)	-0.33	0.21	0.54		
T-test	1.19	0.73	1.30		

Note: Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$.

In Table 10.4, no significant difference is observed between participating parents and non-participating parents, either before or after program delivery. Likewise, the DinD estimator indicates the absence of a significant change over time among participants and non-participants alike based on the t-test.

Table 10.4: Summary of the Family Literacy Program's Effects on Parents' Opinions: The Importance of Fostering a Francophone Environment in the Home

	Means (SD)		Change		
	Pre	Post	Diff (Post- Pre)	Diff (d)	T-test
Participants (P)	3.00 (0.82)	3.05 (0.88)	0.05 (0.76)	0.05	0.62
Non-participants (NP)	3.08 (0.49)	3.38 (0.51)	0.31 (0.75)	0.61	1.47
Difference (P – NP)	-0.08	-0.34	-0.26		
Standardized difference (d)	-0.09	-0.39	-0.34		
T-test	0.48	2.07	1.11		

Note: Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$.

In Table 10.5, no significant program effect is observed. Participants do not differ significantly from non-participants either before or after program delivery or with respect to the gains observed before and after program delivery.

Table 10.5: Summary of the Family Literacy Program's Effects on Parents' Opinions: The Parent and Educator Working Together to Prepare the Child for School

	Means (SD)		Change		
	Pre	Post	Diff (Post- Pre)	Diff (d)	T-test
Participants (P)	3.37 (0.78)	3.52 (0.72)	0.15 (0.83)	0.19	1.91
Non-participants (NP)	3.00 (0.91)	3.00 (1.22)	0.00 (0.82)	0.00	0.00
Difference (P – NP)	0.37	0.52	0.14		
Standardized difference (d)	0.47	0.66	0.17		
T-test	1.42	1.50	0.32		

Note: Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$.

Modeling

In Table 10.6, no significant difference is observed between the groups before or after delivery of the Family Literacy Program. The frequency of modeling activities does not change significantly at any time for either group. However, participating parents tend to perform more modeling activities on average than do non-participating parents, both before and after program delivery.

Table 10.6: Summary of the Family Literacy Program's Effects on Modeling (Two Weeks Post-Program)

	Means (SD)		Change		T-test
	Pre	Post	Diff (Post- Pre)	Diff (d)	
Participants (P)	18.46 (3.28)	18.80 (2.84)	0.37 (3.07)	0.12	1.28
Non-participants (NP)	17.92 (3.15)	18.62 (3.01)	0.69 (2.87)	0.22	0.87
Difference (P – NP)	0.53	0.19	-0.32		
Standardized difference (d)	0.16	0.06	-0.10		
T-test	0.57	0.21	0.24		

Note: Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$.

An examination of Table 10.7 shows no significant difference between participating parents and non-participating parents either before or after delivery of the Family Literacy Program. In short, no program effect manifests immediately after program delivery ends or eight months after the intervention ends.

Table 10.7: Summary of the Family Literacy Program's Effects on Modeling (Eight Months Post-Program)

	Means (SD)		Change		T-test
	Pre	Post	Diff (Post- Pre)	Diff (d)	
Participants (P)	18.46 (3.28)	18.67 (3.15)	0.23 (3.33)	0.07	0.73
Non-participants (NP)	17.92 (3.15)	18.56 (3.38)	0.63 (3.18)	0.19	0.72
Difference (P – NP)	0.53	0.11 (0.98)	-0.40		
Standardized difference (d)	0.16	0.03	-0.12		
T-test	0.57	0.11	0.29		

Note: Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$.

Self-Efficacy

At the last workshop, participating parents were asked to estimate their self-efficacy before and after the Family Literacy Workshops. A review of Table 10.8 shows a medium-sized significant difference in favour of non-participating parents versus participating parents before program delivery. This difference is reproduced by a non-parametric test (rank z-test). However, after the program ended, participating parents reported a level of self-efficacy comparable to that reported by non-participating parents.

This catch-up effect is attributable to the gains reported by parents who participated in the Family Literacy Workshops. A positive and significant change of 0.92 standard deviations is observed among participating parents before and after program delivery. This result is reproduced by a non-parametric analysis. In other words, participating parents reported feeling significantly more confident in their ability to properly fulfill their role as parents after the Family Literacy Program.

Table 10.8: Summary of the Family Literacy Program's Effects on Parents' Self-Efficacy

	Means (SD)		Change		
	Pre	Post	Diff (Post - Pre)	Diff (d)	T-test
Participants (P)	15.48 (2.61)	17.51 (1.69)	2.03 (1.7)	0.92	12.75***
Non-participants (NP)	16.93 (1.64)				
Difference (P – NP)	-1.45	0.58			
Standardized difference (d)	-0.57	0.34			
T-test	-2.88**	1.24			

Note: Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$.

Knowledge

At the last workshop, participating parents were asked to estimate their state of knowledge before and after the Family Literacy Workshops. The questions concerned their knowledge of overall child development, strategies to help their child better prepare for school, the work of educators and, finally, the availability of French-language services and resources in their community.

According to Table 10.9, participating parents report a state of knowledge equivalent to that reported by non-participating parents before program delivery. After the program, participating parents report knowing considerably more than did non-participating parents. The size of the effect is large, at about 1.0 standard deviations, regardless of whether the calculation was based on a group comparison (DinD) or a change in the state of knowledge reported by participants before and after the program (t-test). These results are reproduced by the non-parametric analysis. In short, the results converge to indicate that the Family Literacy Program has a positive and significant effect on parents' knowledge.

Table 10.9: Summary of the Family Literacy Program's Effects on Parents' Knowledge

	Means (SD)		Change		
	Pre	Post	Diff (Post - Pre)	Diff (d)	T-test
Participants (P)	17.19 (3.32)	20.77 (2.1)	3.58 (2.53)	1.29	15.24***
Non-participants (NP)	18.54 (2.73)				
Difference (P – NP)	-1.35	2.23			
Standardized difference (d)	-0.41	1.02			
T-test	-1.65	2.85*			

Note: Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$.

Sense of Belonging

In Table 10.10, no significant difference is observed between the parent groups before or after program delivery. However, there is a medium-sized yet non-significant difference between the groups following program delivery. The analysis comparing the participants' scores before and after the program reveals a significant change among participating parents. The effect is medium in size, that is, 0.55 standard deviations, and is reproduced by the non-parametric analysis.

Table 10.10: Summary of the Family Literacy Program's Effects on Parents' Sense of Belonging to the Francophone Community

	Means (SD)		Change		
	Pre	Post	Diff (Post - Pre)	Diff (<i>d</i>)	T-test
Participants (P)	2.99 (1.21)	3.60 (1)	0.61 (0.74)	0.55	8.86***
Non-participants (NP)	2.93 (1.14)				
Difference (P – NP)	0.06	0.67			
Standardized difference (<i>d</i>)	0.05	0.66			
T-test	0.19	2.11			

Note: Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$.

Analyses by Dosage

The frequency of workshop attendance by participating parents varies. This makes it possible to verify the hypothesis that assiduous parents (i.e., parents with the highest dosage) experience more changes after the workshops than do participating parents who are less assiduous. This hypothesis is tested in Tables 10.11 and 10.12 respectively for parents' opinions and other outcomes. These tables report the association between parent outcomes and dosage. Note that the analyses by dosage are conducted using several dosage indicators: attendance by child, attendance by mother, attendance by father, and attendance by one parent.

In Table 10.11, a positive and significant association is observed between dosage and parents' opinion regarding their role as their child's first educator. The father's rate of attendance is significantly associated with the opinion regarding the importance of the Francophone environment. Thus, parents' active participation in the workshops appears to be associated with the anticipated effects for these outcomes.

In Table 10.12, a positive and significant association is observed between dosage and the outcomes Self-Efficacy and Knowledge. When the dosage indicator in question is attendance by one parent, we observe a correlation of 0.34 with the change scores for the Self-Efficacy outcome and of 0.45 with the change scores for the Knowledge outcome. In other words, a 27% increase in attendance (versus the average of 67%) is associated with an increase of 0.34 standard deviations in Self-Efficacy and of 0.45 standard deviations in Knowledge.

Table 10.11: Correlations Between Dosage and Three Questions Eliciting Parents' Opinion

Indicator	Descriptive Statistics		First Educator (Table 10.3)			Francophone Environment (Table 10.4)			Parent-Educator (Table 10.5)		
	M	SD	Pre	Post	Diff	Pre	Post	Diff	Pre	Post	Diff
Attendance by child	0.67	0.27	-0.01	0.19**	0.20**	0.09	0.14*	0.06	0.02	0.04	-0.05
Attendance by mother	0.56	0.3	0.10	0.16**	0.05	0.00	-0.01	-0.01	-0.05	0.08	0.06
Attendance by father	0.29	0.31	-0.11	0.10	0.21***	0.09	0.22***	0.15**	0.06	0.06	-0.03
Attendance by one parent	0.67	0.27	-0.02	0.18**	0.20**	0.10	0.15**	0.07	0.02	0.04	-0.05

Note: N = 111. Pearson correlations. The Spearman correlations are comparable (unreported). The key findings are the correlations with the Diff (post-pre difference). Significance levels set at *** p < 1%, ** p < 5%; * p < 10%.

Table 10.12: Correlations Between Dosage and Modeling, Self-Efficacy, Knowledge and Sense of Belonging

Indicator	Modeling (Table 10.6)			Self-Efficacy (Table 10.8)			Knowledge (Table 10.9)			Sense of Belonging (Table 10.10)		
	Pre	Post	Diff	Pre	Post	Diff	Pre	Post	Diff	Pre	Post	Diff
Attendance by child	-0.02	0.04	-0.02	-0.12*	0.15	0.36***	-0.24**	0.14	0.44***	0.07	0.18**	0.13*
Attendance by mother	-0.02	0.00	0.01	-0.10	0.10	0.26***	-0.05	0.21***	0.25***	0.02	0.11	0.12*
Attendance by father	0.03	0.07	-0.01	0.01	0.12	0.12*	-0.22**	-0.09	0.20**	0.12*	0.13*	-0.01
Attendance by one parent	-0.02	0.05	-0.03	-0.11	0.15	0.34***	-0.24**	0.15**	0.45***	0.08	0.18**	0.12*

Note: N = 111. Pearson correlations. The Spearman correlations are comparable (unreported). The key findings are the correlations with the Diff (post-pre difference). The key findings are the correlations with the Diff (post-pre difference). Significance level : *** p < 1%, ** p < 5%; * p < 10%.

Analyses by Quality

In this subsection, we examine whether the quality of workshop delivery accounts for the change observed among participating parents. The results of this analysis are presented in Table 10.13 for parents' opinions and in Table 10.14 for other outcomes.

In Table 10.13, Delivery Quality is observed to be associated with a positive and significant change in parents' opinion regarding the collaborative work between parent and educator in fostering the child's development and school readiness. Likewise, a positive and significant association is observed between the quantity of content covered and the change in the opinion of participating parents as regards their role as their child's first educator. Finally, a positive and significant association is observed between parents' positive reactions (observed during the workshops) and a change with respect to their role as their child's first educator (see Table 10.3).⁸⁵

An examination of Table 10.14⁸⁶ shows that the percentage of content covered is associated with an increase in modeling activities. Delivery Quality is linked to gains observed among participants for Self-Efficacy and Knowledge. Session length is positively correlated with an increase in Modeling and in Sense of Belonging.

⁸⁵ There is a high probability that one of the 12 correlation tests for the "Diff" scores is significant simply by chance.

⁸⁶ There is a high probability that two of the 24 correlation tests for the "Diff" scores is significant simply by chance.

Table 10.13: Correlations Between Quality of Program Delivery and Three Questions Eliciting Parents' Opinion

Indicator	Descriptive Statistics		First educator (Table 10.3)			Francophone Environment (Table 10.4)			Parent-Educator (Table 10.5)		
	M	SD	Pre	Post	Diff	Pre	Post	Diff	Pre	Post	Diff
% of content covered	0.97	0.03	-0.11	0.11	0.21**	0.05	0.08	0.04	-0.01	-0.11	-0.13
Delivery Quality	0.71	0.30	-0.10	-0.15	-0.05	-0.22***	-0.12*	0.09	0.03	0.11	0.26***
Positive reactions observed among parents	0.77	0.18	-0.08	0.09	0.17**	-0.06	-0.05	0.02	0.05	0.03	0.07
Negative reactions observed among parents	0.60	0.18	0.08	0.12*	0.04	-0.02	-0.05	-0.03	-0.12*	-0.15	-0.09

Note: The key findings are the correlations with the Diff (post-pre difference). N = 111. Pearson correlations. The Spearman correlations are comparable (unreported). Significance levels set at *** p < 1%, ** p < 5%; * p < 10%.

Table 10.14: Correlations Between Quality of Program Delivery and Modeling, Self-Efficacy, Knowledge and Sense of Belonging.

Indicator	Modeling (Table 10.6)			Self-Efficacy (Table 10.8)			Knowledge (Table 10.9)			Sense of Belonging (Table 10.10)		
	Pre	Post	Diff	Pre	Post	Diff	Pre	Post	Diff	Pre	Post	Diff
% of content covered	-0.19	0.21	0.33***	-0.17*	-0.16*	0.08	-0.15	-0.17**	0.04	-0.06	0.01	0.11
Delivery Quality	-0.05	-0.08	-0.14*	-0.08	0.15**	0.28***	-0.04	0.21**	0.23***	0.13*	0.06	-0.12*
Positive reactions observed among parents	-0.11	0.65	0.10	-0.16**	-0.13	0.11	-0.14*	-0.13	0.07	-0.31***	-0.27***	0.13*
Negative reactions observed among parents	0.04	0.06	0.01	-0.05	-0.06	0.00	-0.02	-0.07	-0.03	-0.16	-0.06	0.16
Session length	-0.13	0.10	0.13*	0.00	-0.18	-0.19**	0.08	-0.08	-0.17*	-0.39***	-0.21**	0.34***
Overall quality	-0.15	0.34	0.03	-0.20**	-0.09	0.20**	-0.11	0.00	0.14	0.30***	0.11	-0.33***

Note: The key findings are the correlations with the Diff (post-pre difference). N = 111. Pearson correlations. The Spearman correlations are comparable (unreported). Significance levels set at *** p < 1%, ** p < 5%; * p < 10%.

10.1.4. Summary of the Impact Analyses on Parents

Analysis results show that the Family Literacy Program has a positive influence on the frequency of literacy activities conducted by parents, their knowledge and their sense of self-efficacy. Subsequent analyses show that the participation rate (in particular, a parent's rate of workshop attendance), Content Fidelity, Delivery Quality and session length are the main determinants of the Family Literacy Program's impact, whatever the parent outcome in question. In short, the analyses show that this component of the tested program achieved one of its goals: to generate a change in parents' knowledge, attitudes and/or behaviours. It remains to be determined whether the changes observed among parents are, in turn, associated with the developmental gains observed for children in the Program Daycare group.

10.2. INDIRECT IMPACT ON CHILDREN

The purpose of the analyses reported in this section is to determine the extent to which the program's impact on children can be attributed to the Family Literacy Program. This objective is achieved through a mediation analysis that ascertains whether the implementation of the Family Literacy Program is responsible for all or part of the program's effects on child outcomes. A formal mediation analysis is possible for both Frequency and Language of Literacy Activities (see Figure 9.1). For other parental variables, the preferred strategy is to associate the parents' change scores (pre- and post-workshop) with the change scores for child development.⁸⁷ This is a way of estimating the "indirect" *ab* link illustrated in Figure 9.1. If this link is not found, the conclusion is that there is no mediation of the program effect.

In Section 10.2.1, we determine whether the changes observed among parents are associated with the changes observed in the children's development for the first year (change score = score for the 12-month assessment period – score for the baseline assessment). In Section 10.2.2, this analysis is repeated but this time the change scores capture the gains observed from the baseline assessment to the 24-month assessment periods.

10.2.1. In the First Year

Change scores were calculated for four standardized ÉPE-AD scales: Communication, Self-Awareness, Cognition, and Expressive Vocabulary. These scores represent the change in the relative position of each child in the distribution of scores observed at the baseline assessment period and in the distribution of scores observed at the 12-month assessment period. Change scores were also calculated for the following parent outcomes: Frequency of Literacy Activities, Language of Literacy Activities, Knowledge, Self-Efficacy, Modeling, and Sense of Belonging.

The analysis was designed to determine whether the changes observed among parents in the Program Daycare group could better predict child development. This analysis includes an initial model and an adjusted model.⁸⁸

⁸⁷ A mediation analysis is conducted only for Frequency and Language of Literacy Activities. Data for the comparison groups were not collected for other parent outcomes.

⁸⁸ The adjusted model adds covariates as well as baseline parent outcomes to the initial model.

Communication

The analysis results show that the Family Literacy Program has a significant effect that fosters an increase in parents' sense of self-efficacy and the subsequent positive impact of this parental sense of self-efficacy on the communication skills of children in the Program Daycare group, that is, a gain of 1.6%. No other parental indicator for the effect of the Family Literacy Program reaches the 10% significance level. We attribute little credibility to this effect since, *a priori*, the other indicators (e.g., changes in Frequency and Language of Literacy Activities) are deemed to be more important determinants of French-language communication skills and no corresponding effect is observed for Expressive Vocabulary.

Self-Awareness

No significant effect was found for any parental indicator.

Cognition

No significant effect was found for any parental indicator.

Expressive Vocabulary

No significant effect was found for any parental indicator.

10.2.2. In the Second Year

Change scores were calculated for the following four standardized ÉPE-AD scales: Communication, Self-Awareness, Cognition, and Expressive Vocabulary. These scores represent the change in the position of each child in relation to the distribution of scores observed at the baseline assessment and 24-month assessment periods. Change scores were also calculated for the following parent outcomes: Frequency of Literacy Activities, Language of Literacy Activities, Knowledge, Self-Efficacy, Modeling, and Sense of Belonging. For the ÉVIP and the EOWPVT, the raw scores were associated with the parents' change scores.

Communication

No significant effect was found for any parental indicator.

Self-Awareness

No significant effect was found for any parental indicator.

Cognition

A series of independent regression analyses show that the parental gains in Self-Efficacy and Knowledge account for respectively 3.9% and 6.3% of the gains observed in the cognitive development of children in the Program Daycare group. The gains observed in Sense of Belonging account for 8.6% of the gains observed in the cognitive development of children in the Program Daycare group. This series of results suggests a significant link between the effect of the Family Literacy Workshops on parental attitudes and their subsequent impact on children's cognitive development. It is interesting to note that this relationship emerges over one year later.

Expressive Vocabulary

No significant effect was found for any parental indicator.

ÉVIP-R and EOWPVT

No significant effect was found for any parental indicator.

10.3. SUMMARY

This chapter aimed foremost to determine the impact of the Family Literacy Workshops on the attitudes and behaviours of parents in the Program Daycare group. The results of impact analyses confirm that the tested program has a positive and significant effect on parents' knowledge, attitudes and behaviours. This finding is supported through analyses by treatment group, analyses concerning parents in the Program Daycare group (participants versus non-participants), analyses by dosage, and analyses by fidelity and quality of Family Literacy Program implementation.

The chapter then went on to ascertain whether the program impact on parents is partly transferred to child development. The results tend to invalidate this hypothesis. The sole effect of the change in parental attitudes - a change attributed to the Family Literacy Workshops - appears to be enhanced cognitive development for children in the second year of the project. Thus, while the Family Literacy Workshops are responsible for part of the program's impact on children, this effect is limited to one dimension of school readiness, at least based on the variables considered in this report's analyses. Overall, the Daycare Program is a more credible and more important source of the tested program's effect according to the results of the mediation analysis reported in Chapter 9.

11. Discussion

This research involved two objectives: to conduct an implementation review for a new dual-component intervention and to evaluate the intervention's impact on minority Francophone children and their parents. The tested program was intended to promote the development of language skills and school readiness of Francophone children living in a minority language context. Theorists have proposed that the driving force behind the acquisition of such competencies is a child's interaction with his or her environment (Bronfenbrenner and Morris, 1998). Drawing upon Bronfenbrenner's (1979) ecological theory of child development, we conceived a child's environment as a series of partially overlapping spheres of influence, which includes the home, the daycare (or school), and the community. As discussed in the introduction, the majority language can influence speakers of the minority language within each of these spheres, including the two primary settings: home and daycare. Accordingly, a two-pronged approach was undertaken for maximum effect on child outcomes (Reese et al., 2010): a) Family Literacy Workshops (Family Literacy Program) to positively influence the home environment through the parents, and b) a high-quality Francophone daycare program (Daycare Program) designed to influence the children more directly.

The next sections concern the results of the implementation study (Section 11.1) and the results of the impact analyses (Section 11.2). The report's conclusion is presented in Section 11.3.

11.1. PROJECT IMPLEMENTATION

What are the main conclusions of the implementation study for the Readiness to Learn project? This dual-component program is based on best practices in the fields of early childhood development and family literacy. The program's implementation review entailed several objectives. First, it enabled us to properly understand how a program developed on paper translates to reality. An implementation study is typically used to understand the barriers, facilitators and adjustments required to facilitate program appropriation by the environment. It is also an important tool to better understand and nuance the program's effects. Thus, a growing number of studies show that an intervention's degree of implementation is linked to the program's observed effects on participants (Charlebois, et al., 2004; Conduct Problem Prevention Research Group, 1999; Dane and Schneider, 1998; Durlak and DuPre, 2008). The program's impact will differ if a number of elements are not implemented or if the quality of implementation is mediocre. In this regard, Durlak and DuPre (2008) concluded, based on data from five meta-analyses, that the magnitude of an intervention's effect is two to three times greater when the program is implemented as planned.

In addition to the study's approach, data from various sources was triangulated to come up with an overall portrait of implementation for the Readiness to Learn project. Triangulation was made possible by adopting a three-tiered methodological approach (Patton, 1990). In other words, we took advantage of various measurement tools, both quantitative (as with the fidelity and quality scales) and qualitative (as with the participant interviews). The benefit of this

approach - also referred to as the mixed-method model - is that it documents program effects in several ways and allows for a more in-depth program assessment (Patton, 2008). This methodology also gave stakeholders a voice, resulting in a range of perspectives on project implementation. Data were then gathered using a number of measurement instruments to evaluate a given phenomenon. These diverse data enhance the validity and reliability of findings due to the complementarity of the data collected.

The following findings take into consideration the integrity of project implementation, first for the daycare component, and then for the family literacy component.

11.1.1. The Daycare Program

The Daycare Program is innovative in several regards. Analysis results for implementation indicate that the tested program contributes new elements in terms of both the structural environment and its content. In other words, certain elements of the new program were not necessarily found in the childcare settings of the comparison group. The integrity of the Daycare Program was examined based on seven dimensions: Structural Fidelity, that is, correspondence between the applied intervention and the planned program; fidelity and quality of the program's educational content, for example, the elements to be implemented as part of the activities and implementation quality for those elements; dosage, that is, the participants' exposure to the program; differentiation between the implemented program and other interventions, that is, the program's new contribution versus what is already taking place; strategies that facilitate implementation; challenges associated with project implementation; and the participants' perceived benefits, that is, the benefits to children. Although all these dimensions were examined, the dimensions that fall under Structural Fidelity, as well as the fidelity and quality of the project's educational content, are the core elements of the study on the tested program's effects. One last core element of the impact study concerns the empirical verification used to quantitatively establish how the tested program differs from the programs at comparison daycares. For the purposes of the impact analyses, this differentiation between the two daycare groups can be considered *the Daycare Program*. If the comparisons are valid, the differentiation (observed and unobserved) will be deemed responsible for the program's observed effect.

The conclusions drawn from the implementation study are presented in Table 11.1. We find that project implementation is relatively stable over the two years of the project. Moreover, the degree of implementation integrity is high for the structural elements as much as the educational content for both cohorts. In terms of differentiation measurement, the tested Daycare Program distinguishes itself from comparison daycares through its fidelity and quality, showing a differentiation with respect to existing programs in the community. These results are supported by a Japel and colleagues (2005) study on the standard of quality in Quebec daycares (see Table 11.1). Moreover, the children's participation indicates a level of dosage that is high enough to allow for detection of the program effect. Overall, the educators appreciated the new program. In particular, they reacted favourably to circle time and to the routine chart, although they had difficulty implementing the program's creativity aspect. A number of facilitators were identified, mainly with regard to the support offered by various team members and the material supplied to the daycares. The main challenges associated with implementation involve time management in connection with program delivery, the difficulty of addressing certain themes, and educator turnover at daycares in the Program Daycare group.

In short, analysis of the Daycare Program's implementation suggests that the new program's fidelity and quality are good, which means it should have the anticipated effect on child outcomes. In fact, a qualitative analysis of the perceived effect corroborates this finding. The educators and trainer perceived several benefits from the program, not only with respect to child behaviour and development, but also with respect to their practices as educators.

11.1.2. The Family Literacy Program

Childcare educators can give only limited individual attention to any particular child (Wasik, 2008), which is why some researchers have argued that parents are an important resource in any intervention involving children (Reese et al., 2010). The workshops for parents were designed to make use of this resource in developing children's language and Francophone culture.

The integrity of the Family Literacy Program was assessed based on six dimensions: Content Fidelity, or the proportion of the program covered during the workshops; quality of workshop delivery by practitioners; dosage, or the parents' rate of workshop participation; parents' reaction to the workshops; facilitators and barriers encountered during project implementation; and the program's effects as perceived by parents and educators. The following paragraphs present the analysis results for each of these dimensions in turn.

Results show that the Family Literacy Workshops were delivered in manner faithful to the planned program. Observers and parents generally had a favourable view of the workshops' Delivery Quality, and parents report having benefited from their experiences. The participation rate for parents was acceptable, although it varied considerably by family. Note that the impact analyses established a link between quality of implementation and attendance, and effects observed among parents. Certain elements facilitated parental participation, especially the meal provided for families, the parent-child workshop, and the presence of educators at the workshops. Conversely, the demands associated with workshop scheduling and the amount of content were barriers to parental participation. A few perceived effects stemmed from the Family Literacy Workshops, mainly parents' awareness of their role and the adoption of new parenting techniques. What's more, parents were also informed of the importance of the French language in minority Francophone environments, which led to an increase in the use of French as part of family activities.⁸⁹

⁸⁹ However, this effect was not robust in the logistic regression analysis.

Table 11.1: Summary of Results for the Implementation Study

Daycare Program Component		Family Literacy Component	
Dimension	Conclusion	Dimension	Conclusion
Structural Quality	<ul style="list-style-type: none"> • Quality is good at program daycares and comparison daycares (see Japel and coll., 2005) • Quality at program daycares is significantly higher than that observed at comparison daycares (see Tables 5.3 and 9.5) 	Delivery Quality	<ul style="list-style-type: none"> • Generally favourable assessment by parents • Sessions tend to exceed the allotted time
Structural Fidelity	<ul style="list-style-type: none"> • Program elements are found more frequently at program daycares than at comparison daycares, and the difference is significant (see Tables 5.4 and 9.6) 	Content Fidelity	<ul style="list-style-type: none"> • On average, 98% of the mandatory material was covered
Educative Quality	<ul style="list-style-type: none"> • Quality is good at program daycares and comparison daycares • Quality at program daycares is significantly higher than that observed at comparison daycares (see Tables 5.5 and 9.5) 		
Fidelity of Educational Content	<ul style="list-style-type: none"> • Frequency of program elements is higher at program daycares than at comparison daycares; difference is significant at 8 months (see Tables 5.6 and 9.6) 		
Quality of Interactions	<ul style="list-style-type: none"> • Quality is good at program daycares and comparison daycares • No significant difference is observed between program daycares and comparison daycares (see Table 9.5) 		
Dosage	<ul style="list-style-type: none"> • Dosage is comparable for both groups, ranging on average from 27 to 30 hours (see Table 5.7) 	Dosage	<ul style="list-style-type: none"> • Average attendance is 63.5%, with a wide variation
Educators' Reactions	<ul style="list-style-type: none"> • Positive reactions to the content of circle time and the routine chart • Program's creative aspect is harder to implement 	Parents' Reactions	<ul style="list-style-type: none"> • Favourite topics identified for 6 of the 10 sessions • Access to French-language material, discussions with other parents and activities with their child as part of the parent-child component are appreciated
Facilitators and Barriers	<ul style="list-style-type: none"> • Facilitators: management support, teamwork, the material supplied, the program's flexibility, follow-up by trainer • Barriers: time management, difficult themes, educator turnover 	Facilitators and Barriers	<ul style="list-style-type: none"> • Facilitators: offering a meal, parent-child workshop, and presence of educators • Barriers: weekly schedule too demanding, too much content planned for the assigned session length

Daycare Program Component		Family Literacy Component	
Dimension	Conclusion	Dimension	Conclusion
Perceived Effects	<ul style="list-style-type: none"> • Better discipline, children are calmer, more independent and secure • Improved school readiness for French and literacy • Educators report being more knowledgeable, having changed their practices as educators (e.g., more child-centred) 	Parents' Perceived Effects	<ul style="list-style-type: none"> • Greater awareness of their role in their child's development and of the importance of French in a minority Francophone environment • More knowledgeable, new parenting techniques and more use of French for activities with their child

11.1.3 Recommendations for Project Implementation in the Future

Staff turnover can be considerable at some daycares. This was particularly problematic for this project in the community of Edmonton. Project implementation requires an investment of resources to train educators, which means the program is better adapted to daycares with a certain level of staff stability and where management actively supports the program. Excluding Edmonton from the impact analyses means that the reported results can be generalized only to a population of daycares where these conditions favourable to project implementation (at least) exist. The program impact may be lesser if the program is conducted on a wider scale without considering these aspects of implementation.

11.2. THE TESTED PROGRAM'S IMPACT

This dual-component program is evaluated using a quasi-experimental methodology with non-equivalent control groups. The methodology involves three treatment groups: the Program Daycare group, consisting of children enrolled in a French-language daycare that offers the new preschool program; the Comparison Daycare group, consisting of children enrolled in a French-language daycare that does not offer the new program; and the Informal Care group, consisting of children whose daytime care is provided at home or at an unregulated family daycare. The purpose of the Comparison Daycare group is to take into account how formal daycare affects child development, which is a subject unto itself. The purpose of the Informal Care group is to factor in how an informal childcare environment affects child development.

The intervention's two key components were the new Daycare Program and the Family Literacy Program. The main goal of the first component was to *directly* influence child outcomes, while that of the second component was to *indirectly* influence child outcomes by changing parents' attitudes and behaviours. The analyses conducted as part of the Readiness to Learn project do not clearly distinguish between the effects of the first component and the second component. A more complex experimental design would have been required to make this distinction. As such, the main analyses comparing the treatment groups test the effect of the dual-component program, that is, the *combined effect* of the program's two components on child development.

11.2.1. Analysis of Program Impact by Group

The key results of the impact analyses are presented in Table 11.2. These analyses show that the tested program has a positive effect on school readiness (the exact results depend on the comparison group used as a benchmark; see Table 11.2 for details). The impact translated to accelerated growth in language skills and cognitive skills for children in the Program Daycare group versus children in the comparison groups. Specifically, when children begin junior kindergarten, medium-sized effects are observed to range from 0.24 to 0.43 standard deviations for the Communication, Self-Awareness, and Expressive Vocabulary scales. The size of these effects is equivalent to an accelerated growth of 2.3 to 3.9 months according to the data of Hill and colleagues (2007). The effects are comparable in size one year later when children begin senior kindergarten. For the study sample, the tested program therefore appears to have had a non-negligible positive impact on these dimensions of children's school readiness as measured by the ÉPE-AD. The positive and significant program effects on language skills are reproduced in the analyses carried out for vocabulary scales: the ÉVIP-R and the EOWPVT. Effects of about 0.20 standard deviations are obtained with these tools, which is equivalent to a leap of about six months in vocabulary growth based on ÉVIP-R standards (Dunn and coll., 1993, Table 4.4, Form A). These reported effects are of practical importance in strengthening the use of French in minority contexts and, ultimately, in preserving the minority Francophone community.

Table 11.2: Summary of Results for the Tested Program's Impact Analyses

	First Year					Second Year				
	Effect at 12 months?	Increases with higher dosage?	Attributable to quality/fidelity?	Attributable to the effect on parents?	Greater benefit for a subsample?	Effect at 24 months?	Increases with higher dosage?	Attributable to quality/fidelity?	Attributable to the effect on parents?	Greater benefit for a subsample?
G1 vs. G2										
Communication	*	*	*		LE	*		*		LE
Self-Awareness	*	*				*		*		
Cognition	*	*	*		E		2	*	*	E
Physical	*					N/A	N/A	N/A	N/A	N/A
Receptive Vocabulary	*	*	*		LE	N/A	N/A	N/A	N/A	N/A
Expressive Vocabulary	*	*	*		E	*		*		LE
ÉVIP-R (at 24 months)	N/A	N/A	N/A	N/A	N/A	*		*		LE
EOWPVT-F (at 20 months)	N/A	N/A	N/A	N/A	N/A	*1		*3		LE
G1 vs. G3										
Communication	*	N/A	N/A		LE		N/A	N/A		LE
Self-Awareness		N/A	N/A			*	N/A	N/A		

	First Year					Second Year				
	Effect at 12 months?	Increases with higher dosage?	Attributable to quality/fidelity?	Attributable to the effect on parents?	Greater benefit for a subsample?	Effect at 24 months?	Increases with higher dosage?	Attributable to quality/fidelity?	Attributable to the effect on parents?	Greater benefit for a subsample?
Cognition	*	N/A	N/A		E		N/A	N/A	*	E
Physical		N/A	N/A			N/A	N/A	N/A	N/A	N/A
Receptive Vocabulary	*	N/A	N/A		LE	N/A	N/A	N/A	N/A	N/A
Expressive Vocabulary		N/A	N/A		E	*	N/A	N/A		LE
ÉVIP-R (at 24 months)	N/A	N/A	N/A	N/A	N/A	*	N/A	N/A	N/A	LE
EOWPVT-F (at 20 months)	N/A	N/A	N/A	N/A	N/A		N/A	N/A	N/A	

Note: An asterisk (*) indicates that the hypothesis in question has been verified. ¹ Effect detected after dosage was included in the model. ² Differentiated dosage effects at 16 and 20 months. ³ The effect is not fully explained. N/A = not applicable; LE = family environment with low exposure to French; E = family environment with high exposure to French; G1 = Program Daycare group; G2 = Comparison Daycare group; G3 = Informal Care group.

11.2.2. Daycare Program versus Family Literacy Program

Main analyses comparing the treatment groups test the *combined effect* of the program's two components on child development. Nevertheless, analyses can be conducted to give an idea of each component's relative contribution to the program impact. Thus, we may infer that if the program impact varies based on the child's exposure to the program (dosage) and if the fidelity/quality of program delivery appears to explain the program effect, then the daycare program component is mainly responsible for the program's observed effect. If the changes observed among parents after the Family Literacy Workshops are not followed by corresponding gains in their children's development, then the family literacy component has no observable effect on children in the Program Daycare group.

A first series of analyses is intended to show a close association between the tested program's impact and the Daycare Program received by the child. We find that the average program impact is based on an average dosage of roughly 27 hours per week at daycare. The effect at the end of the first year is 0.30 standard deviations greater for children who spend about 37 hours per week at daycare. However, the effect when children begin junior kindergarten is essentially null for children who spend only 17 hours per week at daycare. This strongly suggests that the observed effects depend on the daycare program, at least when children begin junior kindergarten. We then conducted a mediation analysis in order to verify whether implementation of the Daycare Program was responsible for all or part of the effects attributed to the program as the result of analyses comparing the study's groups. The finding is that most of the program effect can be attributed with confidence to the differentiation of the tested Daycare Program.⁹⁰ In

⁹⁰ Note that there are no pre-intervention observations for the daycare programs. Certain differences in programming may have existed prior to implementation of the Daycare Program. In this case the information is correlational rather than experimental. That said, in some communities the program daycares and comparison daycares were managed by the same childcare services provider, so there is reason to believe that the programs would have been fairly similar in the absence of the intervention.

other words, the Daycare Program delivered at program daycares differs from the program given at comparison daycares, and this difference accounts for the observed program effect.

Concerning the Family Literacy Program, we conclude that parents in the Program Daycare group experienced benefits. These effects are found in the parents' reported attitudes (i.e., sense of self-efficacy, knowledge about child development) and behaviours (i.e., frequency of literacy activities). These effects are also related to dosage and to the quality of project implementation, which enhances their credibility as true effects of the tested program. Otherwise stated, the Family Literacy Program appears to affect parents in the Program Daycare group.

Unfortunately, there is little empirical evidence that changes in parental attitudes and behaviours foster children's school readiness. Only children's cognitive development when they begin senior kindergarten appears to be affected by the change in parental attitudes attributable to the Family Literacy Workshops. Thus, the effects of the Family Literacy Program are limited to one dimension of school readiness, at least based on the variables considered in this report's analyses. Overall, the Daycare Program is a more credible and more important source of the tested program's effect.

That said, the quasi-experimental design does not allow us to rule out the possibility that the Family Literacy Workshops are necessary to obtain significant effects from the Daycare Program. Certain studies show that interventions targeting parents as well as children have a greater effect than do interventions focussing on only one of the two groups (Brooks-Gunn, Berlin, and Fuligni, 2000; Reese, et al., 2010). However, the unique contribution of the family literacy component may be very small. Family literacy programs that are more highly targeted have a generally positive impact, but one that is small in size, $d = 0.18$ (see the meta-analysis by van Steensel, McElvany, Kurvers, and Herppich, 2011). A meta-analysis by Sénéchal and Young (2008) concludes that family literacy programs have a much greater effect when parents are given concrete strategies they can use with their children, rather than just general advice (for more examples, see Reese, et al., 2010). With this project, it appears that the content of the Family Literacy Program was too general to have a significant effect on development. Note that the results are empirically supported by the literature, which seems to support the conclusion that the Daycare Program is the main driver of the reported effects, without dismissing the complementary role of the Family Literacy Workshops.

Overall, the picture sketched by this series of results supports the conclusion that much of the "combined" program effect is attributable to the Daycare Program. Although the Family Literacy Workshops likely also contribute positively, this link, if any, is not systematically expressed through the observed variables.

11.2.3. The Children's Linguistic Profile

The literature on bilingualism makes a clear distinction between cases where the acquisition of a second language benefits a child's general development and cases where development is affected (e.g., Landry, Allard, and Deveau, 2009). The research identifies two types of bilingualism: additive and subtractive. The form of bilingualism depends on the answer to the following question: Is the *mother tongue* developed enough to support the acquisition of a second language without delaying the age-appropriate development of the child's cognitive skills or language skills in the mother tongue (Ball, 2010)? In the affirmative, the form of bilingualism is additive. In the negative, the bilingualism is subtractive. This question is particularly important

when the mother tongue is the language of instruction. This is the case of many Francophones in minority settings.

The condition required to develop additive bilingualism is not met for many Francophone children living in minority environments, a population that experiences certain delays in literacy in comparison with their peers (Bussière, et al., 2001; Chartier et al., 2011; Canadian Council on Learning, 2008; Education Quality and Accountability Office, 2009; Knighton, Brochu, and Gluszynski, 2010). The underlying causes of this achievement gap are quite well understood. The bilingualism literature identifies a mechanism that is relatively clear: exposure to the language. For bilingualism to be additive, a minimal threshold of exposure to, or use of, the mother tongue must be exceeded (for a review, see Pearson 2007). For various reasons (e.g., motivational, greater exposure to the majority language in several environments, Landry et al., 2009), the required minimal threshold is higher when the mother tongue is a minority language (Pearson, Fernandez, Lewedag, and Oller, 1997; Vihman, Lum, Thierry, Nakai, and Keren-Portnoy, 2006). Thus, Francophone children who grow up in a bilingual environment need special support to improve their likelihood of achieving additive bilingualism.

This is the spirit in which we examined the possibility that children exposed to languages other than French (usually English) benefit most from the tested preschool program. This research question was the subject of an in-depth analysis in the *First Cohort Findings Report* (2014) and is repeated in this report. Here we refer explicitly to the combined effect of both components of the tested program (the Daycare Program and the Family Literacy Program) as a source of environmental influence on children's language development. Thus, we anticipate differentiated effects based on the children's linguistic profile. Children with low exposure to French at home are more likely to develop subtractive bilingualism. The risk is even greater with low exposure to French outside the home because these children live in minority linguistic setting.

Two hypotheses were put forward: At-risk children, or children with low exposure to French in the family environment, should benefit more from the program in terms of their language skills; Children at less risk, or with high exposure to French in the family environment, should benefit more in terms of their cognitive development. These hypotheses were investigated in a series of analyses where program effects were estimated for children with low versus high exposure to the French language when program delivery began (baseline assessment period). The outcomes studied included those that capture language skills (i.e., measures of Communication, Expressive Vocabulary, the ÉVIP-R and the EOWPVT) and those targeting more general cognitive skills (i.e., measures of Cognition). The Self-Awareness scale was not retained since it targeted a series of heterogeneous competences.

Outcomes Targeting Language Skills

As anticipated, the tested program's impact on Communication is greater for children with low exposure to French at every assessment except the last. Similarly, the tested program's impact on Expressive Vocabulary for the Comparison Daycare group is more pronounced for children with low exposure to French. When the Program Daycare group is compared with the Comparison Daycare group, the analyses of the ÉVIP-R and EOWPVT-F vocabulary scales confirm the tested program's effect on children with low exposure to French. The pattern is less clear when the Informal Care group is used as a benchmark, in which case a program effect is

observed only for the ÉVIP-R. No difference between the treatment groups is observed for these outcomes among children with high exposure to French at home.

This pattern of results tends to confirm our hypotheses. However, note that the confidence intervals for the estimated effects based on both household types (low-exposure, high-exposure) usually overlap, indicating that although the direction of the differences was consistent with our hypotheses, the size of the program effect is generally comparable whatever the household type. That said, the pattern of results reported for this project reproduces the results found in other research on this subject (e.g., Maltais, 2007). Moreover, in some cases, the program effect was indeed significantly stronger for children with low exposure (see, for example, the effect at 12 months for the Communication scale in comparison with the Comparison Daycare group).

Outcomes Targeting Other Cognitive Skills

The pattern of results obtained with the Cognition scale of the ÉPE-AD provides sounder empirical support for our hypotheses. The program effect is significantly greater for children with high exposure to French throughout the first year of program delivery. In comparison with their peers in the comparison groups, these children had an advantage in terms of their cognitive development. The advantage observed among these children, if we consider the size of this effect, persists until they begin senior kindergarten. At that time, the overlap in confidence intervals among the treatment groups indicates a program effect that is comparable in size for both household types (low-exposure, high-exposure) for that assessment period, whatever the comparison group used as a benchmark in estimating the program effect.

Is the Hypothesis Verified?

On the whole, these results suggest that basic language skills facilitate the development of cognitive skills (e.g., literacy). This explanation is supported by developmental theory on the acquisition of competencies (Bloom, 1976; Vygotsky, 1978) and the importance of language proficiency to academic achievement (Cummins, 1979; Doherty, 1997; Hindman, Skibbe, Miller, and Zimmerman, 2010). It is also congruent with the findings of Maltais (2007), according to which the effect of a full-day junior kindergarten program on language development depends on low exposure to language, while the program's effects on cognitive development (i.e., gains in reading skills) are observed among children whose exposure to French is high, and these are the children who experience the greatest effects.

In short, the tested program's high degree of fidelity and quality led to a more effective transmission of competencies and knowledge. Although every child benefits from the program, those benefits depend on their readiness to learn certain types of competencies in a French-language environment. These findings have implications with respect to expectations about program impact. It appears that all children benefit from the program in terms of school readiness, but that the type of benefit they experience varies based on the child's linguistic characteristics. The credibility of the findings is based, in part, on theoretical considerations and the findings of other research (Maltais, 2007), since the differentiation of program effects by linguistic profile is not as pronounced as expected.

Implications for the Generalization of Findings

The result of comparative analyses shows that the study sample differs from the SVOLM sample mainly as regards linguistic characteristics. However, we believe that the gap between the linguistic profile of this sample and that of the general population of minority Francophones is no obstacle to the generalization of results. If the population is comparatively more exposed to French at home, then the impact will be observed mainly on cognitive development. If, on the contrary, the population is comparatively less exposed to French, a greater impact on language skills (e.g., vocabulary) is expected. If there is doubt regarding the validity of the analyses indicating a differentiation based on linguistic profile, then a generalized positive effect on school readiness is at least expected based on key findings comparing the treatment groups. As such, there is no reason to believe the program impact would not be reproduced with a different sample of minority Francophone children.

11.2.4. Limitations of the Findings and Future Research

The use of a quasi-experimental design with non-equivalent groups is generally accompanied by a number of concerns in regards to the validity of findings from such a study. Several strategies were applied to ensure valid results. Among the greatest threats to internal validity is that findings may result from a bias linked to group composition rather than from the program effect. This source of bias is diminished by the special attention given to recruiting participants in the comparison groups with a sociodemographic profile similar to that of the Program Daycare group (e.g., socioeconomic level) and living in the same neighbourhood, thereby ensuring that they have access to the same French-language resources and services as the Program Daycare group.⁹¹ Thus, group composition was first controlled for when families signed up for the project, particularly as regards household location for potential members of the comparison groups. A second control was the use of pre-intervention measures, a topic addressed in the next section.

A second threat to the study's internal validity is that the evaluators, educators and parents were aware of who was receiving the treatment and who was not. This source of bias is inevitable when daycare status is known in the community (daycare offering the program, daycare not offering the program) and, by association, the status of children attending those daycares. This is less of a threat than one might first believe, since the daycares, rather than the children, were recruited and assigned to the treatment groups. It is nevertheless difficult to respond to criticisms that the results are due to a bias tied to this knowledge. However, it is difficult to imagine that these potential sources of bias would have, separately or together, produced the pattern of results obtained. No one knew the relative level of program fidelity and quality available at program or comparison daycares. No one knew the hypotheses on the moderating effect of dosage or linguistic profile. As such, it is unlikely that any bias (e.g., in group composition, from evaluators) would have given responses consistent with our research hypotheses.

A third potential criticism is the modest size of the sample on which the findings are based ($N \approx 330$), since statistical analyses are most robust with very large samples ($N > 1000$). Several strategies were used to verify the robustness of results, particularly with respect to the analyses. The analyses comparing the groups are most robust in this regard, as they are based on seven

⁹¹ For more information, see the Revised Work Plan and Methodology Report submitted to HRSDC on March 30, 2007.

data collection periods from over 300 participants (2,100 observations in all). The more nuanced analyses have 100 participants per group (e.g., analysis by dosage, quality of program delivery, linguistic profile). The sampling error associated with these additional analyses is thus comparatively larger. That is why we favoured an interpretation of results based on all results, not just one result in particular. At the community level, it is important to note that the results may be generalized only to the communities studied or to similar communities.⁹²

Finally, the impact analyses concerned only school readiness indicators and French vocabulary indicators. School readiness indicators are general in nature, which can obscure important effects specific to one dimension. In future reports, we will examine the tested program's impact on more targeted measures of cognitive and language skills that predict children's academic achievement. These analyses will serve to elucidate the tested program's contribution for children and will allow us to identify the program impact in the short and medium term.

11.3. CONCLUSION

This research entailed two objectives: to conduct an implementation study for a new dual-component intervention and to evaluate its effects on young minority Francophone children and their parents. The main finding of the implementation study is that program delivery is characterized by a high degree of integrity in terms of both structural elements and the Fidelity and Quality of Educational Content over the two years of the program. These findings are relevant for both cohorts. These dimensions of project implementation are central to study of the tested program's impacts. One final core element of the impact study concerns the empirical verification that quantitatively establishes that the tested program differs from existing programs at comparison daycares by its fidelity and quality. For the purposes of the impact analyses, this differentiation of the program for both daycare groups is viewed as empirically supporting the fact that the *Daycare Program* is considered responsible for the program's observed effect.

The impact study revealed *positive program effects for both child and parent outcomes*. For child outcomes, the positive effects were expressed as higher developmental gains in relation to the Comparison Daycare group and the Informal Care group. Generally, the effects appeared early and remained constant over the two years examined in this report. The nature of these gains depended on the child's exposure to French when the project began. Children with the lowest exposure benefited most from the program in terms of language development (e.g., Expressive Vocabulary), while children with the highest exposure experienced gains in cognitive development. The program component intended for parents was partly successful. Parents did indeed experience positive changes in terms of the frequency of literacy activities, their knowledge, and their sense of self-efficacy. However, it is less clear whether the Family Literacy Program's effects on parents influence child development in turn. While the Family Literacy Program is responsible for part of the program impact on children, that effect is limited to cognitive development, at least based on the variables considered in the analyses thus far. At this point, the pattern of the results suggests that the Daycare Program is the dominant source of the program effect for children.

⁹² The "community" factor was considered a fixed factor in the impact analyses due to the small number of communities.

We can therefore conclude that the tested program had a modest impact on school readiness for minority Francophone children. The size of the observed effect size is equivalent to a developmental gain of a few months of. In the next reports, we will address a new research question: *Does the new preschool program better equip Francophone children in minority environments to succeed in reading and mathematics, tasks essential to academic achievement?* This second research question focuses on the education period of Grades 1 and 2 of primary school, when children are six and seven years of age. In this second phase of the Readiness to Learn project, it will be a matter of establishing whether improved school readiness and greater French-language proficiency increase the likelihood of academic achievement. This question will be answered by studying the program effects on more immediate precursors of academic achievement, such as phonological awareness, knowledge of the alphabet, and vocabulary (Lonigan, 2008).

References

- Abadie, A. (2005). Semiparametric Difference-in-Differences estimators. *Review of Economic Studies*, 72(1), 1-19.
- Adams, G. R., & Ryan, B. A. (2000). *A longitudinal analysis of family relationships and children's school achievement in one- and two-parent families*. Hull, QC: Human Resources Development Canada.
- Allen, M. (2004). Profiles of minority language schools in five Canadian provinces. *Education Quarterly Review*, 9(4), 17-63.
- Arnold, D. H., Zeljo, A., Doctoroff, G. L., & Ortiz, C. (2008). Parent involvement in preschool: Predictors and the relation of involvement to preliteracy development. *School Psychology Review*, 37(1), 74-90.
- Ball, J. (2010). *Enhancing learning of children from diverse language backgrounds: Mother tongue-based bilingual or multilingual education in early childhood and early primary school years*. Washington, D.C., DC: UNESCO. Available online at: <http://www.ecdip.org/docs/pdf/UNESCO%20Mother-tongue%20based%20EY%202010.pdf>.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173-1182.
- Becker, G. S. (1993). Nobel lecture: The economic way of looking at behavior. *The Journal of Political Economy*, 101(3), 385-409.
- Benoît, S. (n.d.). *Questionnaire utilisés par CIRCEM dans l'étude de l'impact de l'alphabétisation familiale sur les familles vivant en milieu minoritaires*. Vanier, ON: Coalition francophone pour l'alphabétisation et la formation de base en Ontario.
- Berger, L. M., Paxson, C., & Waldfogel, J. (2005). *Income and child development*. Princeton, NJ: Princeton University, Center for Research on Child Wellbeing.
- Bertrand, J. (2007). Programmes d'enseignement efficaces au préscolaire. Commentaires sur Kagan et Kauertz et sur Schweinhart. In R. E. Tremblay, R. G. Barr, & R. de V. Peters (Eds.), *Encyclopédie sur le développement des jeunes enfants* [on the Internet] (pp.1-7). Montreal, QC: Centre d'excellence pour le développement des jeunes enfants. Available online at: <http://www.enfant-encyclopedie.com/Pages/PDF/BertrandFRxp.pdf>
- Bertrand, M., Duflos, E., & Mullianathan, S. (2004). How much should we trust Difference-in-Differences estimates?. *The Quarterly Journal of Economics*, 119(1), 249-275.
- Bérubé, A., Legault, L., Janisse, L., Carson, R., Saucier, M.-A., & Lefebvre, L. (2014). *Readiness to Learn in Minority Francophone Communities: Project Implementation Report*. Ottawa, ON: Social Research and Demonstration Corporation.
- Bloom, B. S. (1976). *Human characteristics and school learning*. Columbus, OH: McGraw-Hill.

- Bornstein, L., & Bornstein, M. H. (2007). Pratiques parentales et développement social de l'enfant. In R. E. Tremblay, R. de V. Peters, M. Boivin, & R. G. Barr (Eds.), *Encyclopédie sur le développement des jeunes enfants* [on the Internet] (pp. 1-2). Montreal, QC: Centre d'excellence pour le développement des jeunes enfants. Available online at: <http://www.enfant-encyclopedie.com/Pages/PDF/BornsteinFRxp.pdf>.
- Bredenkamp, S. & Copple, C. (Eds.). (1999). *Developmentally appropriate practice in early childhood programs*. Washington, D.C., DC: National Association for the Education of Young Children.
- Breton, R. (1964). Institutional completeness of ethnic communities and the personal relations of immigrants. *The American Journal of Sociology*, 70(2), 193-205.
- Brigham, J. C. (1991). Attitudes and persuasion. *Social psychology* (2nd ed., pp. 132-224). New York, NY: Harper Collins.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.
- Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (5th ed., Vol. 1, pp. 993-1028). New York, NY: Wiley.
- Brooks-Gunn, J., Berlin, L. J., & Fuligni, A. S. (2000). Early childhood intervention programs: What about the family?. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 549-588). New York, NY: Cambridge University Press.
- Burchinal, M. R., Roberts, J. E., Riggins, R., Jr., Zeisel, S. A., Neebe, E., & Bryant, D. (2000). Relating quality of center-based child care to early cognitive and language development longitudinally. *Child Development*, 71(2), 339-357.
- Bussière, P., Cartwright, F., Crocker, R., Ma, X., Oderkirk, J., & Zhang, X. (2001). *Measuring Up: The Performance of Canada's Youth in Reading, Mathematics and Science*. Ottawa, ON: Human Resources Development Canada, Statistics Canada, Council of Ministers of Education (Canada).
- Canadian Charter of Rights and Freedoms*, s. 23, Part I of the *Constitution Act, 1982*, as Schedule B of the *Canada Act 1982* (U.K.), 1982, c.11.
- Canadian Council on Learning. (2008). *Literacy among Canadian students in minority language contexts*. Unpublished report.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(40), 1-9. Available online at: <http://www.implementationscience.com/content/pdf/1748-5908-2-40.pdf>.
- Charlebois, P., Brendgen, M., Vitaro, F., Normandeau, S., & Bourdreau, J.-F. (2004). Examining dosage effects on prevention outcomes: Results from a multi-modal longitudinal preventive intervention for young disruptive boys. *Journal of School Psychology*, 42(3), 201-220.

- Chartier, M. J., Dumaine, J., & Sabourin, E. (2011). Vivre en français à la petite enfance et apprendre à l'école française en milieu minoritaire, y a-t-il un lien? *Cahiers franco-canadiens de l'Ouest*, 23(1-2), 3-61.
- Chen, H. T. (Ed.). (2005). *Practice program evaluation: Assessing and improving planning, implementation, and effectiveness*. Thousand Oaks, CA: Sage Publications.
- Cleveland, G., Corter, C., Pelletier, J., Colley, S., Bertrand, J., Ontario Institute for Studies in Education/University of Toronto, & Jamieson, J. (2006). *A review of the state of the field of early childhood learning and development in child care, kindergarten and family support programs*. Toronto, ON: Atkinson Centre for Society and Child Development.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioural sciences* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Commission nationale des parents francophones. (2005). *Le meilleur est...avenir! La petite enfance dans les communautés francophones en milieu minoritaire au Canada: une analyse transformative de la situation*. Ottawa, ON: Author. Available online at: http://cnpf.ca/documents/SCAN_FRANCAIS_2005.pdf.
- Committee for Economic Development Research and Policy Committee. (1993). *Why child care matters: Preparing young children for a more productive America. A statement by the Research and Policy Committee of the Committee for Economic Development*. New York, NY: Committee for Economic Development.
- Conduct Problems Prevention Research Group. (1999). Initial impact of the Fast Track prevention trial for conduct problems: I. The high-risk sample. *Journal of Consulting and Clinical Psychology*, 67(5), 631-647.
- Connor, S., & Brink, S. (1999). *Understanding the Early Years: Community impacts on child development*. Hull, QC: Applied Research Branch, Strategic Policy, Human Resources Development Canada. Available online at: <http://www.RHDCC.gc.ca/eng/cs/sp/sdc/pkrf/publications/nlscy/uey/1999-000092/w-99-6e.pdf>.
- Connors, L. J., & Epstein, J. L. (1995). Parent and school partnerships. In M. H. Bornstein (Ed.), *Handbook of parenting. Volume 4: Applied and practical parenting* (pp. 437-458). Mahwah, NJ: Lawrence Erlbaum Associates.
- Canadian Council on Learning. (2007). *Rapport sur l'apprentissage au Canada: rapport sur l'état de l'apprentissage chez les jeunes enfants*. Ottawa, ON: Author.
- Corbeil, J.-P., & Blaser, C. (2007). *The Evolving Linguistic Portrait, 2006 Census*. Ottawa, ON: Statistics Canada. Available online at: <http://www12.statcan.ca/census-recensement/2006/as-sa/97-555/pdf/97-555-XIE2006001-eng.pdf>.

- Corbeil, J.-P., Grenier, C., & Lafrenière, S. (2007). *Minorities Speak Up: Results of the Survey on the Vitality of Official-Language Minorities*. Ottawa, ON: Statistics Canada. Available online at: <http://www.statcan.gc.ca/pub/91-548-x/91-548-x2007001-eng.pdf>.
- Corter, C., & Pelletier, J. (2005). Parents and community involvement in schools: Policy panacea or pandemic? In N. Bascia, A. Cumming, A. Datnow, K. Leithwood, & D. Livingstone (Eds.), *International Handbook of Educational Policy* (pp. 295-327). Great Britain: Springer.
- Cummins, J. (1979). *Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. Working papers on bilingualism, No. 19*. Toronto, ON: The Ontario Institute for Studies in Education.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control?. *Clinical Psychology Review*, 18(1), 23-45.
- Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, 19(2), 294-304.
- Dearing, E., Berry, D., & Zaslow, M. (2006). Poverty during early childhood. In K. McCartney, & D. Phillips (Eds.), *The Blackwell handbook of early childhood development* (pp. 399-423). Malden, MA: Blackwell Publishing.
- Desrosiers, H., & Ducharme, A. (2006). *Commencer l'école du bon pied: facteurs associés à l'acquisition du vocabulaire à la fin de la maternelle*, 4(1). Quebec City, QC: Institut de la statistique Québec. Available online at: http://www.jesuisjeserai.stat.gouv.qc.ca/pdf/publications/feuille/fascicule_ecole_bon_pied.pdf.
- Dionne, A.-M. (2007). *Incidence des pratiques de littératie familiale sur les compétences en lecture de l'enfant dans les familles immigrantes*. Presented at L'Actualité de la Recherche en Éducation et en Formation, Stasbourg, France.
- Doherty, G. (1997). *Zero to six: The basis for school readiness*. Hull, QC: Applied Research Branch of Strategic Policy, Human Resources Development Canada. Available online at: <http://www.RHDCC.gc.ca/eng/cs/sp/sdc/pkrf/publications/1997-002557/SP-241-02-01E.pdf>.
- Donald, S. G., & Lang, K. (2007). Inference with Difference-in-Differences and other panel data. *The Review of Economics and Statistics*, 89(2), 221-233.
- Duncan, G. J., & Brooks-Gunn, J. (Eds.). (1997). *Consequences of growing up poor*. New York, NY: Russell Sage Foundation.
- Dunn, L. M., Thériault-Whalen, C. M., & Dunn, L. M. (1993). *Échelle de Vocabulaire en Images Peabody adaptation française du Peabody Picture Vocabulary Test-Revised: Manuel pour les Formes A et B*. Toronto, ON: Pearson Canada Assessment Inc.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3-4), 327-350.

- Education Quality and Accountability Office. (2009). *Ontario student achievement: EQAO's provincial report on the results of the 2008-2009 assessments of reading, writing and mathematics, primary division (grades 1-3) and junior division (grades 4-6), and the grade 9 assessment of mathematics. English-language students*. Toronto, ON. Available online at: http://www.eqao.com/pdf_e/09/Cpr PJ9e_0909_web.pdf.
- Engle, P. L., Black, M. M., Behrman, J. R., Cabral de Mello, M., Gertler, P., Kapiriri, L., ... International Child Development Steering Group. (2007). Child development in developing countries 3: Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *The Lancet*, 369, 229-242.
Recovered
at:http://www.who.int/child_adolescent_health/documents/pdfs/lancet_child_dev_series_paper3.pdf. Consulted on March 23, 2010.
- Erikson, E. H. (1994). *Identity and the life cycle*. New York, NY: W. W. Norton & Company.
- Forgues, É., & Landry, R. (2006). *Defining Francophones in Minority Situations: An Analysis of Various Statistical Definitions and their Implications*. Ottawa, ON: Canadian Institute for Research on Linguistic Minorities.
- Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology*, 51(1), 115-134.
- Gardner, M. (1979). *Expressive One-Word Picture Vocabulary Test: Manual and forms*. Novato, CA: Academic Therapy Publications.
- Gauthier St-Onge, D. (n.d.). *Questionnaire utilisés dans l'Évaluation du programme alpha District 3, Edmundston, Nouveau-Brunswick*.
- Gilbert, A. (2003). *La petite enfance: porte d'entrée à l'école de langue française: une vision nationale. Rapport finale de la recherche*. Ottawa, ON: Centre for Interdisciplinary Research on Citizenship and Minorities, University of Ottawa in partnership with the Canadian Teachers' Federation. Available online at: <http://www.ctf-fce.ca/documents/Resources/Francaise/petiteenfance/Visionnationale.pdf>.
- Guimond, L. (2003). *Revue des travaux sur la vitalité des communautés francophones du Canada*. Ottawa, ON: University of Ottawa.
- Hargrave, A. C., & Sénéchal, M. (2000). A book reading intervention with preschool children who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly*, 15(1), 75-90.
- Harms, T., & Clifford, R. M. (1990). *Infant/Toddler Environment Rating Scale*. New York, NY: Teachers College Press.
- Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale – Revised Edition*. New York, NY: Teachers College Press.
- Harms, T., Clifford, R. M., & Cryer, D. (2010). *Échelle d'évaluation de l'environnement préscolaire – révisée*. Translated by M. Baillargeon and H. Larouche. Quebec City, QC: Presses de l'Université du Québec.

- Haveman, R., & Wolfe, B. (1995). The determinants of children's attainments: A review of methods and results. *Journal of Economic Literature*, 33(4), 1829-1878.
- Hayes, A. F., & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods*, 39(4), 709-722.
- Hertzman, C., & Kohen, D. (2003). Neighbourhoods matter for child development. *Transition*, 33, 3-5. Available online at: http://www.vifamily.ca/media/node/339/attachments/Neighbourhoods_Matter_for_Child_development_Hertzman_Kohen.pdf.
- Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3), 172-177.
- Hindman, A. H., Skibbe, L. E., Miller, A., & Zimmerman, M. (2010). Ecological contexts and early learning: Contributions of child, family, and classroom factors during Head Start, to literacy and mathematics growth through first grade. *Early Childhood Research Quarterly*, 25(2), 235-250.
- Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review*, 4(1), 93-106.
- Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly*, 23(1), 27-50.
- Hox, J. (2002). *Multilevel analysis: Techniques and applications* (1st ed.). New York, NY: Routledge.
- Human Resources and Skills Development Canada (2006). *Child Care Pilot Project: Guiding Principles for the Pilot Project, 2006*. Available online at: http://www.hrsdc.gc.ca/eng/cs/comm/sd/child_care/2006/guiding_principles.shtml.
- Institut de la statistique Québec. (2003). *En 2002...j'avais 5 ans! Questionnaire informatisé rempli par l'intervieweuse (QIRI) – E6*. Quebec City, QC: Author.
- Izzo, C. V., Weissberg, R. P., Kasprow, W. J., & Fendrich, M. (1999). A longitudinal assessment of teacher perceptions of parent involvement in children's education and school performance. *American Journal of Community Psychology*, 27(6), 817-839.
- Jamieson, J. (2003). Analysis of covariance (ANCOVA) with difference scores. *International Journal of Psychophysiology*, 52(3), 277-283.
- Japel, C., Tremblay, R. E., & Côté, S. (2005). La qualité, ça compte! Résultats de l'Étude longitudinale du développement des enfants du Québec concernant la qualité des services de garde. *Choix !RPP*, 11(4), 1-46.
- Jordan, G. E., Snow, C. E., & Porche, M. V. (2000). Project EASE: The effect of a family literacy project on kindergarten students' early literacy skills. *Reading Research Quarterly*, 35(4), 524-546.
- Kane, T. J. (2004). *The impact of after-school programs: Interpreting the results of four recent evaluations*. New York, NY: William T. Grant Foundation. Available online at:

http://www.wtgrantfoundation.org/publications_and_reports/browse_reports/kane_working_paper.

- Klebanov, P. K., Brooks-Gunn, J., & Duncan, G. J. (1994). Does neighbourhood and family poverty affect mother's parenting, mental health, and social support?. *Journal of Marriage and Family*, 56(2), 441-455.
- Knighton, T., Brochu, P., & Gluszynski, T. (2010). *Measuring up: Canadian results of the OECD PISA study. The performance of Canada's youth in reading, mathematics, and science. 2009 first results for Canadians aged 15*. Available online at: <http://www.cmec.ca/Publications/Lists/Publications/Attachments/254/PISA2009-can-report.pdf>.
- Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, 23(4), 418-444.
- Lafrance, F. (1993). Les conditions sociolinguistiques de l'enseignement du français minoritaire et leurs conséquences sur la pédagogie du français langue maternelle en Ontario. *Revue du Nouvel-Ontario*, 15, 11-34.
- Lamb, T., & Tschillard, R. (2003). *An underutilization design in applied research: The retrospective pre-test*. Presented at the Annual Meeting of the American Sociological Association.
- Lamb, T. A., & Tschillard, R. (2005). Evaluating learning in professional development workshops: Using the retrospective pre-test. *Journal of Research in Professional Learning*. Recovered at: http://www.mdecgateway.org/olms/data/resource/6794/Evaluating%20Learning%20in%20PD%20Workshops_OST-PD.pdf. Consulted in February 2008.
- Landry, R. (1994). Diagnostic sur la vitalité de la communauté acadienne du Nouveau-Brunswick. *Égalité*, 36, 11-39.
- Landry, R. (2010). *Petite enfance et autonomie culturelle: Là où le nombre le justifie... V*. Research Report for the Commission nationale des parents francophones. Moncton, NB: Canadian Institute for Research on Linguistic Minorities.
- Landry, R., & Allard, R. (1997). L'exogamie et le maintien de deux langues et de deux cultures: le rôle de la francité familioscolaire. *Revue des sciences de l'éducation*, 23(3), 561-592.
- Landry, R., & Allard, R. (2000). Langue de scolarisation et développement bilingue: le cas des acadiens et francophones de la Nouvelle-Écosse, Canada. *DiversCité Langues*, V. Recovered at: <http://www.telug.quebec.ca/diverscite/entree.htm>. Consulted on April 23, 2009.
- Landry, R., Allard, R., & Deveau, K. (2007). Bilingual schooling of the Canadian francophone minority: A cultural autonomy model. *International Journal of the Sociology of Language*, 185, 133-162.
- Landry, R., Allard, R., & Deveau, K. (2009). Self-determination and bilingualism. *Theory and Research in Education*, 7(2), 203-213.

- Laperrière, A. (1997). Les critères de scientificité des méthodes qualitatives. In J. Poupard, J.-P. Deslauriers, L. H. Groulx, A. Laperrière, R. Mayer & A. P. Pires (Eds.). *La recherche qualitative: enjeux épistémologiques et méthodologiques* (pp. 365-389). Boucherville, QC: Gaëtan Morin.
- Legault, L., Mák, G., Verstraete, M., & Bérubé, A. (2014). *Readiness to Learn in Minority Francophone Communities: Reference Report*. Ottawa, ON: Social Research and Demonstration Corporation.
- Lemelin, J.-P., & Boivin, M. (2007). *Mieux réussir dès la première année: l'importance de la préparation à l'école*. Quebec City, QC: Institut de la statistique Québec. Available online at: <http://www.jesuisjeserai.stat.gouv.qc.ca/pdf/publications/feuille/Fasc2Vol4.pdf>.
- LeTouzé, S. (2005). *Pour mon enfant d'abord: étude de l'impact de l'alphabétisation familiale sur les familles vivant en milieu minoritaire. Étape 1: 2004-2005*. Ottawa, ON: Centre for Interdisciplinary Research on Citizenship and Minorities.
- LeTouzé, S. (2006). *Pour mon enfant d'abord: étude de l'impact de l'alphabétisation familiale sur les familles vivant en milieu minoritaire. Étape 2: Cohortes 2 et 3: 2005-2006*. Ottawa, ON: Centre for Interdisciplinary Research on Citizenship and Minorities.
- LeTouzé, S. (2007). *Pour mon enfant d'abord: étude de l'impact de l'alphabétisation familiale sur les familles vivant en milieu minoritaire. Étape 3: Cohortes 4 et 5. 2006-2007*. Ottawa, ON: Centre for Interdisciplinary Research on Citizenship and Minorities.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. *American Psychologist*, 48(12), 1181-1209.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York, NY: John Wiley & Sons.
- Lonigan, C. J. (2008). *Effective strategies for promoting young children's literacy skills*. Paper presented at the 40th Banff International Conferences on Behavioural Science. Effective Early Learning Programs: Research, Policy and Practice, Banff, AB.
- Lonigan, C. J., Schatschneider, C., Westberg, L., & The National Early Literacy Panel. (2008). Identification of children's skills and abilities linked to later outcomes in reading, writing, and spelling. In *Developing early literacy: Report of the National Early Literacy Panel* (pp. 55-101). Jessup, MD: National Institute for Literacy.
- Maltais, C. (2007). Relation entre la littératie familiale en milieu francophone et l'incidence d'un programme de maternelle quatre ans à temps plein sur le développement du langage et de la lecture des enfants. *Éducation francophone en milieu minoritaire*, 2(1), 6-18. Available online at: <http://www.ustboniface.mb.ca/cusb/reefmm/notrerevue/v2n1Maltais.pdf>.
- Martel, A. (2001). *Rights, Schools and Communities in Minority Contexts: 1986 - 2002. Toward the Development of French through Education, An Analysis*. Ottawa, ON: Commissariat aux langues officielles.

- Meisels, S. J., & Atkins-Burnett, S. (2006). Evaluating early childhood assessments: A differential analysis. In K. McCartney & D. Phillips (Eds.), *The Blackwell handbook of early childhood development* (pp. 533-549). Malden, MA: Blackwell Publishing.
- Moore, T. (2005). *Towards improved support for families of young children: The role of communities*. Presented at the 9th Australian Institute of Family Studies Conference. Available online at: <http://aifs.gov.au/institute/afrc9/moore1.pdf>.
- Moore, D., & Tananis, C. A. (2009). Measuring change in a short-term educational program using a retrospective pretest design. *American Journal of Evaluation*, 30(2), 189-202.
- Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The Review of Economics and Statistics*, 72(2), 334-338.
- National Association for the Education of Young Children. (1984). *Accreditation criteria and procedures: Position statement of the National Academy of Early Childhood Programs*. Washington, D.C., DC: National Association for the Education of Young Children.
- National Institute of Child Health and Human Development Early Child Care Research Network, & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, 74(5), 1454-1475.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage Publications.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics*, 28(3), 399-410.
- Pearson, B. Z., Fernandez, S., Lewedag, V., & Oller, D. K. (1997). The relation of input factors in lexical learning by bilingual infants. *Applied Psycholinguistics*, 18(1), 41-58.
- Pelletier, J., & Corter, C. (2005). Design, implementation, and outcomes of a school readiness program for diverse families. *The School Community Journal*, 17(3), 89-116.
- Phillips, L. M., Hayden, R., & Norris, S. P. (2006). *Family literacy matters: A longitudinal parent-child interaction study*. Calgary, AB: Temeron Books.
- Raikes, H. H., Summers, J. A., & Roggman, L. A. (2005). Father involvement in Early Head Start programs. *Fathering: A Journal of Theory, Research, and Practice about Men as Fathers*, 3(1), 29-58.
- Reese, E., Sparks, A., & Leyva, D. (2010). A review of parent interventions for preschool children's language and emergent literacy. *Journal of Early Childhood Literacy*, 10(1), 97-117.
- Rockwell, S. K., & Kohn, H. (1989). Post-then-pre evaluation. *Journal of Extension*, 27(2) [online]. Recovered at: <http://www.joe.org/joe/1989summer/a5.php>.
- Rogers, K. N., Fernandez, M., Thurber, L., & Smitley, A. (2004). Exploring differential attrition rates among system of care evaluation participants. *Journal of Community Psychology*, 32(2), 167-176.

- Sanders, M. R., & Morawska, A. (2006). Peut-on améliorer les résultats des enfants en modifiant les connaissances des parents, leurs attentes dysfonctionnelles et la régulation de leurs émotions? In R. E. Tremblay, R. de V. Peters, M. Boivin & R. G. Barr (Eds.), *Encyclopédie sur le développement des jeunes enfants* [on the Internet] (pp. 1-13). Montreal, QC: Centre d'excellence pour le développement des jeunes enfants. Available online at: <http://www.enfant-encyclopedie.com/Pages/PDF/Sanders-MorawskaFRxp.pdf>.
- Sénéchal, M., & LeFevre, J.-A. (2002). Parental involvement in the development of children's reading skills: A five-year longitudinal study. *Child Development*, 73(2), 445-460.
- Sénéchal, M., & Young, L. (2008). The effect of family literacy interventions on children's acquisition of reading from kindergarten to grade 3: A meta-analytic review. *Review of Educational Research*, 78(4), 880-907.
- Ska, B. (1995). *Épreuve de compréhension de CARROW-WOOLFOLK. Épreuve de dénomination EO-WPVT-R. Données normatives*. Montreal, QC: Groupe coopératif en orthophonie pour la région des Laurentides.
- Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, D.C., DC: National Academic Press.
- Statistics Canada. (2006). *Survey on the Vitality of Official-Language Minorities (SVOLM): Questionnaire*. Ottawa, ON: Author. Available online at: http://www.statcan.gc.ca/imdb-bmdi/instrument/5099_Q1_V1-eng.pdf.
- Statistics Canada, & Human Resources and Skills Development Canada. (2005). *National Longitudinal Survey of Children and Youth. Survey Overview for 2002/2003 Data Collection. Cycle 5*. Ottawa, ON: Author.
- Statistics Canada, & Human Resources and Skills Development Canada. (2006). *National Longitudinal Survey of Children and Youth. Survey Instruments for 2004/2005 Data Collection. Cycle 6. Book 1: Parent, Child and Youth*. Ottawa, ON: Author.
- Strauss, A. L., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon.
- Tamis-LeMonda, C. S., & Rodriguez, E. T. (2008). Rôle des parents pour favoriser l'apprentissage et l'acquisition du langage chez les jeunes enfants. In R. E. Tremblay, R. G. Barr, R. de V. Peters & M. Boivin (Eds.), *Encyclopédie sur le développement des jeunes enfants* [on the Internet] (pp. 1-11). Montreal, QC: Centre d'excellence pour le développement des jeunes enfants. Available online at: <http://www.ccl-cca.ca/pdfs/ECLKC/encyclopedia/TamisLemondaRodriguezFRxpCSAJELangage.pdf>.
- Taylor, P. J., Russ-Eft, D. F., & Taylor, H. (2009). Gilding the outcome by tarnishing the past: Inflationary biases in retrospective pretests. *American Journal of Evaluation*, 30(1), 31-43.

- Thomas, E. M. (2006). *Readiness to Learn at School Among Five-year-old Children in Canada*. Ottawa, ON: Statistics Canada. Available online at: <http://www.statcan.gc.ca/pub/89-599-m/89-599-m2006004-eng.pdf>.
- Thompson, G., Legault, L., Lalonde, P., & Bérubé, A. (2014). *Readiness to Learn in Minority Francophone Communities: First Cohort Findings Report*. Ottawa, ON: Social Research and Demonstration Corporation.
- Tremblay, R. E., Nagin, D. S., Séguin, P. D., Boivin, M., Pérusse, D., & Japel, C. (2004). Physical aggression during early childhood: Trajectories and predictors. *Pediatrics*, 114(1), e43-e50.
- van Steensel, R., McElvany, N., Kurvers, J., & Herppich, S. (2011). How effective are family literacy programs?: Results of a meta-analysis. *Review of Educational Research*, 81(1), 69-96.
- Vihman, M. M., Lum, J. A. G., Thierry, G., Nakai, S., & Keren-Portnoy, T. (2006). The onset of word form recognition in one language and in two. In P. McCardle & E. Hoff (Eds.), *Childhood bilingualism: Research on infancy through school age* (pp. 30–44). Clevedon, UK: Multilingual Matters.
- Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on Head Start children and teachers. *Journal of Educational Psychology*, 98(1), 63-74.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817-838.
- Williams, R. L. (2000). A note on robust variance estimation for cluster-correlated data. *Biometrics*, 56(2), 645-646.
- Willms, J. D. (2007). *Child Care Pilot Project: Final report on direct assessment measures and parent and community questionnaires*. Fredericton, NB: KSI Research International Inc.
- Willms, J. D. (2009). *Évaluation Petite Enfance: Appréciation Directe, version allongée*. Fredericton, NB: KSI Research International Inc.
- Woodbridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Appendix A: Evaluation Tools and Timetable

The timetable for the evaluation of children is presented in the table below. For each data collection wave, we report the construct measured (*in italics*) and the direct-assessment tool that was used to capture its development. We refer to the EYE-DA and the PPVT-R by their French appellations: the ÉPE-AD and ÉVIP-R respectively.

Table A.1: Timetable for Child Assessments

Baseline	+ 4 months (1 st post-test)	+8 months (2 nd post-test)	+ 12 months (3 rd post-test)	+16 months (4 th post-test)	+20 months (5 th post-test)	+24 months (6 th post-test)
<i>School Readiness</i>						
ÉPE-AD: ⁹³ Self-Awareness (A)						
ÉPE-AD: Cognitive Ability (B)						
ÉPE-AD: Language and Communication (C)						
ÉPE-AD: Physical Ability and Motor Skills (D)						
<i>Expressive Vocabulary</i>						
ÉPE-AD subscale						
					EOWPVT-F	
<i>Receptive Vocabulary</i>						
ÉPE-AD subscale						
						ÉVIP-R

¹⁰⁵ Évaluation de la petite enfance — Appréciation directe (Willms, 2007), which is the validated French translation of the Early Years Evaluation — Direct Assessment (EYE-DA).

Appendix B: Content of the Three Versions of the ÉPE-AD

This section concerns the content of three version of the ÉPE-AD (the EYE-DA in English) employed during the first two years of implementation of the Readiness to Learn project. New versions of the test were created in response to two issues noted in the fall of 2008 and following decisions taken in the wake of a meeting with the Consultative Committee of the HRSDC on December 17th, 2008. The first issue concerned the potential for a ceiling effect based on projected scores for the 5th evaluation (February 2009) derived from the performance of children on the 3rd and 4th evaluations. The test designer was hired by the HRSDC to create new items for the test that were more difficult, thereby allowing the developmental trajectory of children to be followed on Domains A, B, and C. The second issue concerned the need for a measure of language ability that makes finer discriminations among children. Though the ÉPE-AD measures school readiness well, it was not sensitive enough to capture the developmental dimension of children and only scratches the surface of their ability to communicate.

For the purpose of the discussion to follow, we give the name ‘initial ÉPE-AD’ to the version of the test administered in February of 2009, the name ‘extended ÉPE-AD’ to the version of the test administered in the winter of 2009 (containing the more difficult questions developed by Willms, and excluding the easiest questions), and finally the name ‘modified ÉPE-AD’ to the version of the test that was reworked by SRDC and used for the evaluations from June to October 2009.

Domains Measured by the Initial ÉPE-AD

The ÉPE-AD designed by Doug Willms measures four domains as well as a fifth domain designed specifically for the Readiness to Learn project:

- Domain A = Self-Awareness;
- Domain B = Cognitive Ability;
- Domain C = Language and Communication;
- Domain D = Physical and Motor Skills; and
- Domain E = Awareness and Engagement in Francophone Culture.

Domains Measured by the Modified ÉPE-AD

The modified ÉPE-AD measures seven dimensions, namely:

- Expressive Vocabulary;
- Self-Awareness;
- Phonological Awareness;
- Numeracy;
- Memory for Personal Information;

- Alphabetic Knowledge; and
- Oral Reading.

These scales allow the trajectories of children to be pursued for the domains A, B, and C. The modified ÉPE–AD contains all items from Domain A (minus question A18). As for Domain B, all items were kept, either in the Phonological Awareness subscale, the Numeracy subscale, or the Alphabetic Knowledge subscale. Finally, Domain C is partially preserved within the Expressive Vocabulary subscale, with which it correlates at .87, confirming that it will be possible to continue tracking the trajectory of this domain.

The tables below serve to compare the items used in the three versions of the ÉPE–AD. The modified version of the ÉPE–AD was conceived in French only; the items listed in the following tables are taken verbatim from that evaluation. Note that items which were not retained for the modified ÉPE–AD are not presented in the table. The list of excluded items includes those that were eliminated by Willms for the extended ÉPE–AD.

Table B.1: Expressive Vocabulary Subscale

Items	Initial ÉPE-AD	Extended ÉPE-AD	Modified ÉPE-AD
Expressive Vocabulary			
Comment s'appelle cet objet? <i>un seau/une chaudière, des boutons, des chandelles, une horloge/un réveille-matin</i>	A11	x	V1
Pointez chacune des images, une à la fois. Peux-tu me nommer quatre couleurs? Peux-tu me nommer quatre fruits? Peux-tu me nommer quatre animaux? Peux-tu me nommer quatre vêtements?	A12	x	V2
Comment se nomme cette partie du corps? <i>le menton, le coude, le poignet, l'épaule</i>	A14	x	V3
Qu'est-ce que cette personne fait comme métier? <i>le policier, l'enseignant(e), le fermier/jardinier, le médecin</i>	A15	x	V4
Qu'est-ce qu'on utilise pour écrire? pour s'asseoir? pour se promener? pour balayer?	C8	x	V5
Nomme le plus d'animaux possible.	C9 (nomme huit animaux)	x	V6
Quel temps fait-il dans chacune des images? <i>Il vente, il pleut, il neige, il fait soleil</i>		A21	V7
Comment s'appelle ceci? <i>thermomètre, tasse à mesurer, règle, balance</i>		A22	V8

Note: The Expressive Vocabulary scale (six items) is correlated .87 with the items from Domain C of the extended version (C7 à C14). We kept only items asking children to name an object, so as to be as consistent as possible with the methodology of conventional vocabulary tests such as the EOWPVT-F.

Table B.2: Self-Awareness Subscale

Items	Initial ÉPE-AD	Extended ÉPE-AD	Modified ÉPE-AD
Self-Awareness			
Que devrais-tu faire quand tu : es fatigué, as faim, as froid, as soif?	A7	x	CS1
La souris est petite; l'éléphant est ____ (gros ou grand). Cet oiseau est dehors : cet oiseau est ____ (en dedans ou à l'intérieur). Ce pot à biscuits est plein; ce pot à biscuits est ____ (vide). Cette échelle est courte; cette échelle est ____ (longue).	A8	x	CS2
Placer des objets pour démontrer sa compréhension de : premier, dernier, devant, derrière .	A10	x	CS3
Nommer le moment de la journée (le matin, l'après-midi, la soirée, la nuit) associé à deux de ces situations communes : des étoiles dans le ciel, prendre le petit déjeuner, retourner à la maison après l'école	A13	x	CS4
Quel jour est-ce aujourd'hui? Peux-tu me dire quel jour on sera demain? Hier, quel jour était-ce? Peux-tu me nommer un jour de la fin de semaine?		A17	CS5
Avant de traverser la rue, que dois-tu faire? Que dois-tu faire si un inconnu te demande d'aller avec lui? Dans la voiture, que portes-tu toujours pour être en sécurité? Si tu entends le détecteur de fumée, que dois-tu faire?		A20	CS6

Table B.3: Phonological Awareness Subscale

Items	Initial ÉPE-AD	Extended ÉPE-AD	Modified ÉPE-AD
Phonological Awareness			
Écoute-moi prononcer le mot jouet : jouet commence par le son /j/. Par quel son commence le mot piano? Par quel son commence le mot tomate? Par quel son commence le mot boîte? Par quel son commence le mot rouge?		C16	CP1
Dis-moi si les mots riment, s'ils finissent par le même son. pomme/homme; boîte/lune; chien/chat; pain/main	B11	x	CP2
Trouve deux mots qui commencent par /b/ comme dans « ballon ».	B13	x	CP3

Dis-moi si les mots suivants commencent par le même son : mère/lait, balle/beau, plat/clou, peau/pas	B14	x	CP4
Écoute pendant que je prononce le mot canif. Le mot canif se termine par le son /f/. Quel est le son qui termine le mot cheval? Quel est le son qui termine le mot album? Quel est le son qui termine le mot autobus? Quel est le son qui termine le mot neuf (9)?		C19	CP5
Voici un bateau (une pomme, une table, une fenêtre et du lait). Le mot bateau commence par un son /b/. Écoute /b/, bateau. Quelle image commence par le son t? Quelle image commence par le son f? Quelle image commence par le son l? Quelle image commence par le son p?		B18	CP6

Note: For item C16, we ask only for the sound at the start of a word and not the letter. The goal here was to measure phonological awareness specifically, and we wished to keep this measure pure. We added the words “piano” and “tomate” to keep the scale score at 4, the same as other version of the test.

Table B.4: Numeracy Subscale

Items	Initial ÉPE-AD	Extended ÉPE-AD	Modified ÉPE-AD
Numeracy			
Dire le nombre de parties : Un chat a combien de queues? Un chien a combien de pattes? Un oiseau a combien d'ailes? Ta main a combien de doigts?	A9	x	N1
Compte toutes les étoiles et dis-moi combien il y en a.	B9	x	N2
Regarde chaque chiffre et dis-moi lequel est <u>le plus</u> grand. 6 ou 8, 12 ou 10, 9 ou 7, 11 ou 12.	B10	x	N3
Mets ensemble des jetons pour faire un groupe de... 5, 7, 8, 9.	B12	x	N4
Montre-moi le chiffre... 13, 20, 45, 112.		B19	N5

Note: For item B19, the order of the numerals was changed on the picture so that they would not be presented in the same order as asked.

Table B.5: Memory of Personal Information Subscale

Items	Initial ÉPE-AD	Extended ÉPE-AD	Modified ÉPE-AD
Memory for Personal Information			
Peux-tu me dire quel âge tu as et quels sont le jour et le mois de ton anniversaire (de ta fête)?	A16	x	MP1
Savoir où l'on habite : Quel est le nom de la rue où tu habites?		A19	MP2

Quel est le nom de la ville où nous habitons? Quel est le nom de notre province? Quel est le pays où nous habitons?			
---	--	--	--

Table B.6: Alphabetic Knowledge Subscale

Items	Initial ÉPE-AD	Extended ÉPE-AD	Modified ÉPE-AD
Alphabetic Knowledge			
Dire le son de la lettre majuscule	B16	B17	Livret
Dire la lettre majuscule	B15		
Dire la lettre minuscule		B20	Livret

Note: The letters were presented in order of difficulty based on the performance of Canadian francophone children. A total of 26 letters were presented, including those with accents 'é, è, ê'. Each correct response contributed 1 to the total. The range of the scale is from 0 to 26 for B17 and B20..

Table B.7: Oral Reading Subscale

Items	Initial ÉPE-AD	Extended ÉPE-AD	Modified ÉPE-AD
Oral Reading			
Lire huit mots perçus de façon globale à fréquence élevée. Dis-moi quel mot est écrit ici.	B17	B16 Chien Oui Non Balle Chat Le La Nez	L1 Le La Oui Non Balle Maman Nez Papa

Note: We changed the order of the words so that they are in decreasing frequency of usage for French. We changed two of the words for alternatives that are more commonly encountered by French children. The words “dog” and “cat” are very common in English-language alphabet books, but the French equivalents are more orthographically complex (i.e., “ch” is a multi-letter symbol for one sound). We substituted these words with “maman” [mom] and “papa” [dad], which are slightly longer but also more frequent in French-language children books.

Appendix C: Summary of Topics Addressed in the Family Literacy Workshops

N° 1 – *Apprendre en français, c'est amusant!* [It's fun to learn in French!]

- Introduction to the Readiness to Learn project;
- Our children at 18

N° 2 – *Je suis le premier éducateur de mon enfant* [I am my child's first educator]

- Your role as your child's first educator
- Scaffolding
- Parenting styles
- Multiple intelligences (Gardner)

N° 3 – *L'éveil à l'écrit* [Early literacy]

- What children learn before they read and write
- Choosing a good book
- Shared reading

N° 4 – *L'apprentissage: Stimulation des sens de l'enfant* [Learning: Stimulating children's senses]

- Stimulating a child's senses
- Overall development
- How to stimulate your child's learning at home

N° 5 – *Mon enfant en quête d'autonomie* [My child's quest for independence]

- Self-discipline leads to self-esteem
- The role of emotions in learning and in life success
- Strategies to help children develop self-discipline, such as naming emotions

N° 6 – *Le développement langagier, culturel et identitaire de mon enfant* [My child's language, cultural and identity development]

- The importance of songs, stories and nursery rhymes
- Additive bilingualism
- Varieties of French
- How to stimulate language at home

N° 7 – *La communication* [Communication]

- Facilitating communication
- Verbal and non-verbal communication
- The role of physical position and attitude in communication

N° 8 – *Le bien-être de notre famille* [Family well-being]

- Family changes
- Children's needs (according to Maslow)
- Values worth passing on

N° 9 – *Notre vie au sein de la communauté francophone* [Living in a Francophone community]

- Activities, resources and services in the Francophone community
- Knowledge of La Francophonie
- Differences between French-language school and immersion

N° 10 – *Les célébrations* [Celebrations]

- Review of what learned in the family workshops
- Traditions
- Celebrations are an integral part of the family

Appendix D: Administration Procedure for the ÉPE-AD (Pre-Intervention Measure)

The evaluators who administered the *ÉPE-AD* (the EYE-DA in English) to children were recruited starting in late summer 2007. SRDC provided evaluators with theoretical and practical training lasting approximately six hours in August, September and, for the community of Orléans, in October. In addition to presenting the test administration protocol, training provided an introduction to the Readiness to Learn project and procedures relating to confidentiality. They signed a contract whereby they agreed to adhere to the administration and confidentiality protocol. The complete steps of the protocol for test administration are as follows:

1. The evaluators call parents to make an appointment for the home assessments or to notify them of the time of the daycare assessment. These calls are also intended to confirm the child's age in months as well as to answer questions about the languages spoken with parents and friends for Domain E (*Awareness and Engagement in Francophone Culture*).
2. The evaluators and community coordinator get in touch with the participating daycares to define the schedule and arrange a place in the classroom that is favourable for a good assessment.
3. The evaluator applies the "medical" method, which means she checks whether the child is the right one before filling out the identification information on the hardcopy questionnaire.
4. The evaluator addresses the child in his or her mother tongue first, then applies the protocol for determining the test language.
5. The evaluator follows the tool developer's scoring instructions, that is, she rounds the score to the lower whole number for the purpose of conducting a prudent assessment.
6. The evaluator encourages the child, but does not give any hints as to how to answer, unless the protocol indicates to do so.
7. If the child gets tired during the test, the evaluator stops and can start again later at the start of the domain where she left off.
8. At the end of the assessment, the evaluator gives the child a sticker to thank him or her for participating.
9. If the child really does not want to participate, the evaluator must try to assess the child at least one more time (another day).

Point 4 is crucial to the assessment because failure to comply with this rule can considerably affect the results. Table D.1 presents the administration protocol suggested by M. Willms to determine the language of testing (presented on December 8, 2006, and revised in July 2007).

Table D.1: Decisional Tree for Determining the Language of Test Administration

Steps	Protocol
1. Administration of Domain E	Administer the six questions of Domain E directly <u>to the child, and to the parent for questions E4 to E6</u> , at the start of the test.
2. Decisional tree for determining the language of testing (Domain E)	If the score is greater than 6 for Domain E, administer the remainder of the ÉPE-AD in French. If the score is less than or equal to 6 for Domain E, assess Domain C in French and in English.
3. Decisional tree for determining the test language (Domain C)	If the score in French for this domain is greater than 14, the remainder of the assessment must be administered in French. If the score in French for this domain is less than or equal to 14 and the score in English is less than or equal to 14, the remainder of the test must be administered in French as well. If the score in French for Domain C is less than or equal to 14, but the score in English is greater than 14, the remainder of the assessment must be administered in English.

In addition, there are two starting points for the test depending on the child's age. If the child is less than four years old, the evaluator starts with the first item for the domain. Otherwise she starts further on in the test with the option of going back to the starting point if the child is struggling. This decision rule was not mentioned in the *Reference report* (Legault et al., 2014) because at the time of the baseline measurement all children were less than four years of age.

Appendix E: Validation of ÉPE-AD Imputation

The imputation strategy adopted in this report is based on the child's prior and subsequent performance with the same scale. However, it is useful to quantify the accuracy of the algorithm that we used for a particular situation. Below we present the results of an analysis designed to validate the imputation strategy used to offset the missing values of the ÉPE-AD.

To do so, we used data from the second cohort for the case study, retaining only children with complete data. First, we eliminated the observed values of the second assessment. Next, the imputation algorithm was used to estimate these artificial missing values. The algorithm assessment was carried out by comparing the real observed values to the imputed values. If the imputation is satisfactory, the original scores and the imputed scores will be distributed similarly (i.e., have similar averages and standard deviations) and will be strongly correlated.

The results of this analysis are reported in Table E.1. For the Communication, Cognition, and Self-Awareness scales, the algorithm tends to overestimate the children's true performance. However, the difference between the mean original scores and the mean imputed scores is not significant. For these child outcomes, the association between the two distributions is very strong, with correlations ranging from 0.70 to 0.88. In total, the performance of the algorithm for these three scales confirms our expectations regarding the accuracy of the imputation method used.

Exceptionally, the algorithm tends to significantly underestimate children's performance in the Physical domain, that is, by about 0.30 standard deviations. The correlation between the original scores and the imputed scores is, however, relatively strong for this domain, with observed correlations of 0.50. In short, the imputation algorithm for this scale is less accurate than expected.

Note that whatever the observed magnitude of the bias, there are no serious consequences for the impact analyses. The missing values are distributed evenly across the treatment groups, except for the fourth period (see Table 7.2), thereby indicating that the internal validity of analyses for the Physical domain is not jeopardized. We deemed it preferable to keep the sample composition as constant as possible from one analysis to the next. Missing values are therefore imputed for all ÉPE-AD scales despite the potential bias for the Physical domain.

Table E.1: Means and Standard Deviations for Original and Imputed Scores as Well as Their Inter-correlation

ÉPE-AD Domain	Original Data		Imputed Data		Size of Difference (Cohen <i>d</i>)	Confidence Interval for <i>r</i> (95%)		
	M	SD	M	SD		<i>r</i>	Lower Bound	Upper Bound
Cognition	24.36	± 10.23	26.50	± 11.61	0.20	0.70	0.57	0.79
Communication	26.85	± 11.18	28.40	± 11.62	0.14	0.77	0.67	0.84
Self-Awareness	33.95	± 11.44	34.23	± 12.74	0.02	0.88	0.82	0.92
Physical	14.17	± 7.01	11.39	± 9.20	- 0.34**	0.50	0.32	0.64

Note: M = mean; SD = standard deviation; N=84; Significance levels set at *** $p < 1\%$; ** $p < 5\%$; * $p < 10\%$.

Appendix F: Result of the Confounding Variables Analysis

This appendix reports the results of a preliminary analysis designed to identify the relevant covariates for the impact analyses. The results concern a sample that combines both cohorts and excludes first-year withdrawals ($N = 342$). Unless indicated otherwise, the children's membership in treatment groups is that of the baseline for association tests involving this variable.

Variables that meet the two inclusion criteria are retained as covariates in the impact analysis. The two criteria are: i) a significant association with at least one dependant variable, and ii) a significant association with treatment group membership. The following tables report the significance tests for both criteria for each potential covariate.

The results are presented as follows. The examined variables are categorized in tables as follows: demographic and family composition variables, socioeconomic variables, linguistic variables, Francophonie engagement variables, parenting style variables, characteristics of the daycare environment, and a series of methodological factors. Lastly, a final table contains the results of an analysis verifying the stability of certain covariates' effects over time. In the tables that follow, variables that were selected as covariates are in boldface.

Table F.1: Relevance of Demographic and Family Composition Variables

Covariate	Significant Correlations ($p < .10$) with Dependent Variables at Baseline									Association Tests with Treatment Group Membership			
	Com	Awa	Cog	Phy	Expressive Voc.	Receptive Voc.	Freq. Lit.	Language Lit.	Sample parameter % or M (SD)	Group Association Test	Program Daycare Group % or M (SD)	Comparison Daycare Group % or M (SD)	Informal Care Group % or M (SD)
Gender (+Woman) ^a	0.11	0.13	0.19	0.17	0.12	0.13	-	-	50.9%	$\chi^2(2) = 1.37$	46.1%	53.4%	52.3%
Child's Age (in months)^b	0.30	0.35	0.44	0.47	0.29	0.25	-	-	38.42 (3.66)	F(2, 339) = 0.24	38.35 (3.78)	38.59 (3.52)	38.28 (3.75)
Twins (+Twins)	-	-	-	-	-	-	-	-	2.3%	N/A	3.9%	1.5%	1.9%
Older Siblings	-	-0.13	-	-	-0.13	-	-0.25	-0.11	52.3%	$\chi^2(2) = 2.74$	48.0%	57.9%	49.5%
Younger Siblings	-	-	0.10	-	-	-	0.20	-	34.5%	$\chi^2(2) = 22.19^{***}$	24.5%	27.8%	52.3%
Single-Parent Home (+single-parent) ^a	-	-	-	-	-	-	-	-	8.5%	$\chi^2(2) = 5.51^*$	6.9%	12.8%	4.7%
Household Size	-	-0.12	-	-	-0.10	-	-0.12	-0.13	4.04 (0.93)	F(2, 339) = 4.97^{***}	3.87 (0.83)	3.99 (1.01)	4.26 (0.88)

Note: Variables selected as covariates in the impact analyses are in boldface. ^aAll categorical variables are binary (codes 0, 1) and the descriptive statistics reported in the table are in category “1” indicated in parentheses and preceded by the (+) symbol. Com= Communication, Awa = Self-Awareness, Cog = Cognition, Phy = Physical, Voc. = Vocabulary, Freq Lit. =Frequency of Literacy Activities, Language Lit. = Language of Literacy Activities, M = mean; SD = standard deviation; N=342; Significance levels set at *** $p < 1\%$; ** $p < 5\%$; * $p < 10\%$.

^bExceptionally, this variable was retained as a covariate even though it is not associated with membership in the treatment conditions. This decision is justified by its very strong association with the dependent variables.

Table F.2: Relevance of Socioeconomic Variables

Covariate	Significant Correlations ($p < .10$) with Dependent Variables at Baseline									Association Tests with Treatment Group Membership			
	Com	Awa	Cog	Phy	Expressive Voc.	Receptive Voc.	Freq. Lit.	Language Lit.	Global parameters % or M (SD)	Group Association Test	Program Daycare Group % or M (SD)	Comparison Daycare Group % or M (SD)	Informal Care Group % or M (SD)
Age at First Birth (+Young mother < age 24) ^a	-0.17	-	-	-0.09	-0.13	-0.14	-	-0.12	16.1%	$\chi^2(2) = 0.77$	14.7%	15.2%	18.7%
Income (+\$60,000 and over) ^a	0.14	0.13	0.16	0.10	0.13	0.11	-	-	76.6%	$\chi^2(2) = 1.53$	76.5%	79.7%	72.9%
Mother's Education ^b (+At least a college diploma)	0.24	0.25	0.18	-	0.26	0.20	-	-	79.8%	$\chi^2(2) = 0.27$	81.4%	79.7%	78.5%
Father's Education ^b (+At least a college diploma) ^a	-	-	-	-	-	-	-	-	67.0%	$\chi^2(2) = 2.77$	70.6%	69.2%	60.7%
Immigrant Status	-	-	-	-	-	-	-	-0.15	4.7%	$\chi^2(2) = 0.55$	5.9%	4.5%	3.7%
Social Capital	-	-	0.11	-	-	-	-	-	15.43 (3.91)	$F(2, 338) = 0.26$	15.58 (3.91)	15.50 (4.02)	15.21 (3.78)
Social Support	-	-	-	-0.11	-	-	0.10	-	19.46 (1.69)	$F(2, 182) = 1.89^c$	19.14 (2.23)	19.59 (1.02)	19.59 (1.73)

Note: Variables retained as covariates for the impact analyses are in boldface. ^a All categorical variables are binary (codes 0, 1) and the reported percentages are for category "1" indicated between parentheses and preceded by the (+) symbol. Com= Communication, Awa = Self-Awareness, Cog = Cognition, Phy = Physical, Voc. = Vocabulary, Freq Lit. = Frequency of Literacy Activities, Language Lit. = Language of Literacy Activities, M = mean; SD = standard deviation; N=342; Significance levels set at *** $p < 1\%$; ** $p < 5\%$; * $p < 10\%$.

^b Exceptionally, this variable was retained as a covariate only for the analyses by linguistic profile. ^c Welch's F-test is robust to heterogeneous variance.

Table F.3: Relevance of Linguistic Variables

Covariate	Significant Correlations ($p < .10$) with Dependent Variables at Baseline								Association Tests with Treatment Group Membership				
	Com	Awa	Cog	Phy	Expressive Voc.	Receptive Voc.	Freq. Lit.	Language Lit.	Global parameters % or M (SD)	Group Association Test	Program Daycare Group % or M (SD)	Comparison Daycare Group % or M (SD)	Informal Care Group % or M (SD)
Frequency of Literacy Activities (baseline)	-	0.10	0.10	-	-	0.10	N/A	-0.17	15.12 (2.94)	F(2, 339) = 4.63**	15.37 (2.56)	14.53 (3.00)	15.62 (3.11)
Language of Literacy Activities (baseline)	0.45	0.38	-	-	0.36	0.34	-0.17	N/A	19.96 (5.97)	F(2, 339) = 3.14**	18.95 (6.47)	19.89 (5.74)	21.00 (5.63)
Language Spoken to Child (mother)	0.45	0.37	-	-	0.37	0.31	-0.15	0.81	4.06 (1.36)	F(2, 336) = 4.57**	3.83 (1.45)	3.99 (1.39)	4.38 (1.17)
Language Spoken to Child (father)	0.37	0.29	-	-	0.30	0.24	-0.09	0.70	3.69 (1.65)	F(2, 324) = 5.23***	3.36 (1.75)	3.62 (1.66)	4.09 (1.46)
Continuum of French Spoken by Child	0.52	0.41	-	-	0.42	0.37	-0.10	0.84	3.79 (1.47)	F(2, 339) = 5.02***	3.46 (1.58)	3.79 (1.45)	4.10 (1.34)
Language of Care (0-12)	0.18	0.11	-	-	-	0.11	-	0.32	2.25 (0.87)	F(2, 337) = 1.21	2.21 (0.86)	2.19 (0.89)	2.36 (0.86)
Language of Care (13-24)	0.12	-	-	-	-	-	-	0.22	2.25 (0.86)	F(2, 337) = 1.17	2.17 (0.84)	2.23 (0.89)	2.35 (0.84)
Language of Care (25-36)	-	-	-0.12	-	-	-	-	0.17	2.33 (0.85)	F(2, 337) = 0.27	2.29 (0.85)	2.33 (0.88)	2.37 (0.82)
Household Type Based on FOLS (+Endog-French)	0.30	0.26	-	-	0.26	0.21	-0.13	0.69	50.1%	$\chi^2(2) = 3.81$	46.1%	47.0%	57.9%
Household Type Based on Language Spoken to the Child (+Endog-French) ^a	0.30	0.26	-	-	0.26	0.21	-0.13	0.60	52.2%	$\chi^2(2) = 4.57$	45.1%	51.5%	59.8%

Note: Variables retained as covariates for the impact analyses are in boldface. ^aAll categorical variables are binary (codes 0, 1) and the reported percentages are for category “1” indicated between parentheses and preceded by the (+) symbol. Com= Communication, Awa = Self-Awareness, Cog = Cognition, Phy = Physical, Voc. = Vocabulary, Freq Lit. =Frequency of Literacy Activities, Language Lit. = Language of Literacy Activities, M = mean; SD = standard deviation; N=342; Significance levels set at *** $p < 1\%$; ** $p < 5\%$; * $p < 10\%$.

Table F.4: Relevance of Sociolinguistic Variables

Covariate	Significant Correlations ($p < .10$) with Dependent Variables at Baseline									Association Tests with Treatment Group Membership			
	Com	Awa	Cog	Phy	Expressive Voc.	Receptive Voc.	Freq. Lit.	Language Lit.	Global parameters % or M (SD)	Group Association Test	Program Daycare Group % or M (SD)	Comparison Daycare Group % or M (SD)	Informal Care Group % or M (SD)
Vitality	0.26	0.21	-	-	0.20	0.22	-0.17	0.53	15.61 (5.34)	F(2, 288) = 5.01***	15.18 (5.05)	14.88 (5.32)	17.16 (5.41)
Engagement in Francophone Culture	-	-	-	-	-	-	-	0.24	8.16 (3.04)	F(2, 308) = 0.06	8.22 (3.00)	8.08 (2.99)	8.20 (3.16)
Sense of Belonging	0.41	0.35	-	-	0.32	0.35	-0.17	0.62	2.42 (0.75)	F(2, 337) = 2.79*	2.30 (0.82)	2.41 (0.75)	2.55 (0.66)

Note: Variables retained as covariates for the impact analyses are in boldface. Com= Communication, Awa = Self-Awareness, Cog = Cognition, Phy = Physical, Voc. = Vocabulary, Freq Lit. =Frequency of Literacy Activities, Language Lit. = Language of Literacy Activities, M = mean; SD = standard deviation; N=342; Significance levels set at *** $p < 1\%$; ** $p < 5\%$; * $p < 10\%$.

Table F.5: Relevance of Parenting Variables

Covariate	Significant Correlations ($p < .10$) with Dependent Variables at Baseline									Association Tests with Treatment Group Membership			
	Com	Awa	Cog	Phy	Expressive Voc.	Receptive Voc.	Freq. Lit.	Language Lit.	Global parameters % or M (SD)	Group Association Test	Program Daycare Group % or M (SD)	Comparison Daycare Group % or M (SD)	Informal Care Group % or M (SD)
Family Functioning	0.16	0.15	-	-	0.14	0.14	-	-	29.75 (3.16)	$F(2, 204) = 1.78^a$	29.17 (4.12)	30.02 (2.57)	29.98 (2.69)
Depression (mother)	0.15	0.15	-0.10	-0.12	-0.12	-0.12	-	-0.13	10.41 (3.22)	$F(2, 331) = 2.23$	10.71 (3.42)	9.95 (2.70)	10.72 (3.57)
Authoritarian Parenting Style	0.15	0.15	0.15	0.12	0.14	0.15	0.12	-	12.97 (2.10)	$F(2, 336) = 9.39^{***}$	12.25 (2.27)	13.37 (2.05)	13.16 (1.80)
Positive Parenting Style	-	-	-	-	-	-	0.30	-	23.09 (1.90)	$F(2, 336) = 2.35^*$	22.95 (1.98)	22.94 (2.00)	23.42 (1.65)
Empowerment	-	-	-	-	-	-	0.18	-	16.64 (2.31)	$F(2, 289) = 1.41$	16.35 (2.36)	16.63 (2.23)	16.93 (2.36)

Note: Variables retained as covariates for the impact analyses are in boldface. Com= Communication, Awa = Self-Awareness, Cog = Cognition, Phy = Physical, Voc. = Vocabulary, Freq Lit. =Frequency of Literacy Activities, Language Lit. = Language of Literacy Activities, M = mean; SD = standard deviation; N=342; Significance levels set at *** $p < 1\%$; ** $p < 5\%$; * $p < 10\%$.

^a Welch's F-test is robust to heterogeneous variance.

Table F.6: Relevance of Daycare Environment Characteristics

Covariate	Significant Correlations ($p < .10$) with Dependent Variables at Baseline									Association Tests with Treatment Group Membership			
	Com	Awa	Cog	Phy	Expressive Voc.	Receptive Voc.	Freq. Lit.	Language Lit.	Global parameters % or M (SD)	Group Association Test	Program Daycare Group % or M (SD)	Comparison Daycare Group % or M (SD)	Informal Care Group % or M (SD)
Dosage – hrs/week (baseline)	0.21	0.12	-	-	0.10	0.12	-	0.29	30.77 (10.93)	$F(1, 233) = 0.00$	30.81 (10.31)	30.73 (11.42)	N/A
Dosage – hrs/week (at 4 months)	0.26	0.13	-	-	0.12	0.19	-0.15	0.41	27.51 (9.15)	$F(1, 230) = 0.48$	27.03 (8.63)	27.87 (9.55)	N/A
Dosage – hrs/week (at 8 and 12 months)	0.18	-	-	-0.12	-	0.14	-	0.26	27.49 (9.43)	$F(1, 224) = 0.00$	27.48 (9.27)	27.49 (9.58)	N/A
Educators:													
Years of Experience	-	-	0.12	-	-	-	-	-	6.15 (5.40)	$F(1, 38) = 0.12$	5.69 (3.25)	6.35 (6.14)	N/A
Special Training	0.21	0.12	-	-	0.12	0.24	-	0.20	80.5%	$\chi^2(1) = 0.09$	83.3%	79.3%	N/A
Education (+At least a college diploma) ^a	-	-	-	-0.13	-	-0.14	-	-0.12	51.2%	$\chi^2(1) = 0.01$	50.0%	51.7%	N/A
Mother Tongue (+French) ^a	-	-	-	-	-	-	-	-	90.2%	$\chi^2(1) = 0.92$	83.3%	93.1%	N/A

Note: Variables retained as covariates for the impact analyses are in boldface. ^aAll categorical variables are binary (codes 0, 1) and the reported percentages are for category “1” indicated between parentheses and preceded by the (+) symbol. Com= Communication, Awa = Self-Awareness, Cog = Cognition, Phy = Physical, Voc. = Vocabulary, Freq Lit. = Frequency of Literacy Activities, Language Lit. = Language of Literacy Activities, M = mean; SD = standard deviation; N=342; Significance levels set at *** $p < 1\%$; ** $p < 5\%$; * $p < 10\%$.

Table F.7: Relevance of Methodological Factors

Covariate	Significant Correlations ($p < .10$) with Dependent Variables at Baseline									Association Tests with Treatment Group Membership			
	Com	Awa	Cog	Phy	Expressive Voc.	Receptive Voc.	Freq. Lit.	Language Lit.	Global parameters % or M (SD)	Group Association Test	Program Daycare Group % or M (SD)	Comparison Daycare Group % or M (SD)	Informal Care Group % or M (SD)
School Enrolment (Year 2) ^b	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	$\chi^2(4) = 24.90^{***}$	N/A	N/A	N/A
Full-time (+ Full-time) ^a	0.27	0.21	0.28	N/A	0.18	N/A	-	0.14	36.1%	$\chi^2(2) = 3.63$	38.9%	31.0%	42.5%
Part-time (+Part-time) ^a	-0.37	-0.27	-0.15	N/A	-0.28	N/A	-	-	29.0%	$\chi^2(2) = 23.23^{***}$	32.2%	38.0%	9.2%
Language of Administration (+French) ^a :													
Survey – baseline	0.25	0.26	-	-	0.24	0.19	-0.11	0.47	92.1%	N/A	91.2%	87.2%	99.1%
at 4 months	0.36	0.31	0.16	-	0.27	0.29	-	0.54	82.7%	$\chi^2(2) = 1.83$	80.8%	80.8%	86.8%
at 8 months	0.36	0.23	-	-	0.21	0.25	-	0.57	87.6%	$\chi^2(2) = 0.60$	85.4%	88.5%	88.4%
at 12 months	0.37	0.30	-	-	0.30	0.28	-0.12	0.54	86.6%	$\chi^2(2) = 1.38$	83.9%	86.0%	89.4%
ÉPE-AD - baseline	0.44	0.20	-	-	0.20	0.13	-	0.59	87.2%	$\chi^2(2) = 2.32$	83.0%	88.5%	89.6%
at 4 months	0.57	0.30	0.24	0.16	0.26	0.27	-0.13	0.62	86.5%	$\chi^2(2) = 1.61$	86.1%	89.2%	83.6%
at 8 months	0.65	0.32	0.23	-	0.30	0.33	-	0.59	88.2%	$\chi^2(2) = 2.66$	87.5%	91.5%	84.8%
at 12 months	0.63	0.30	0.22	-	0.26	0.44	-0.14	0.49	92.3%	$\chi^2(2) = 7.84^{**}$	93.7%	96.1%	86.7%
Change in Group (+change) ^a	-0.11	-0.14	-	-	-0.10	-	-	-	4.1%	N/A	5.9%	3.8%	2.8%
Cohort ^a (+ 1 ^{iere}) ^c	-	-	-	-	-	-	-	-	71.3%	$\chi^2(2) = 0.08$	70.6%	72.2%	71.0%
Community	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	$\chi^2(6) = 13.30^{**}$	N/A	N/A	N/A
Orleans (+Orleans) ^a	0.26	0.19	-	-	0.19	0.21	.010	0.21	31.0%	$\chi^2(2) = 0.94$	29.4%	29.3%	34.6%
Cornwall (+Cornwall) ^a	-0.37	-0.30	-0.10	-0.12	-0.27	-0.37	0.10	-0.39	33.6%	$\chi^2(2) = 1.65$	38.2%	33.1%	29.9%
Durham (+Durham) ^a	-	-	0.12	0.30	-	-	-	-0.22	10.8%	$\chi^2(2) = 8.17^{**}$	14.7%	13.5%	3.7%
Edmunston (+Edmunston) ^a	0.13	0.13	-	-	0.11	0.13	-0.23	0.36	24.6%	$\chi^2(2) = 5.66^*$	17.6%	24.1%	31.8%

Note: Variables retained as covariates for the impact analyses are in boldface. ^aAll categorical variables are binary (codes 0, 1) and the reported percentages are for category “1” indicated between parentheses and preceded by the (+) symbol. Com= Communication, Awa = Self-Awareness, Cog = Cognition, Phy = Physical, Voc. = Vocabulary, Freq Lit. = Frequency of Literacy Activities, Language Lit. = Language of Literacy Activities, M = mean; SD = standard deviation; N=342; Significance levels set at *** $p < 1\%$; ** $p < 5\%$; * $p < 10\%$. ^b Children group membership at the fifth assessment period. ^cExceptionally, cohort was retained as a covariate in the impact analyses.

Table F.8: Tests Robust to the Instability of the Confounding Variables' Effects over Time

Covariate	Dependent Variables							
	Com	Awa	Cog	Phy	Expressive Voc.	Receptive Voc.	Freq. Lit.	Language Lit.
Community - Wald F(9,14)	2.88**	0.57	0.56	1.16	0.37	3.28**	2.26*	2.03
Cohort - Wald F(3, 20)	2.61*	1.25	0.70	15.79***	0.62	3.22**	0.30	8.87***
Younger Siblings - Wald F(3, 20)	0.52	0.18	3.36**	4.24**	0.16	0.77	0.06	3.56**
Household Size - Wald F(3, 20)	1.45	1.98	1.04	0.55	1.22	0.93	2.31	0.64
Frequency of Literacy Activities (baseline) - Wald F(3, 20)	0.86	2.01	3.12**	0.26	0.69	0.42	40.92***	0.67
Language of Literacy Activities (baseline) - Wald F(3, 20)	5.25***	2.41*	0.56	1.01	1.42	0.30	1.24	15.96***
Language Spoken to Child (mother) - Wald F(3, 20)	8.32***	6.62***	1.45	1.34	1.94	0.89	4.39**	2.72*
Language Spoken to Child (father) - Wald F(3, 20)	6.73***	1.19	0.60	0.37	0.32	2.73*	4.81**	5.84***
Continuum of French Spoken by Child - Wald F(3, 20)	1.76	0.29	0.93	0.42	0.45	1.02	4.73**	7.74***
Child's Age (in months) - Wald F(3, 20)	2.94*	3.01*	0.51	0.61	0.88	1.74	2.29	2.29
Vitality - Wald F(3, 20)	0.05	2.95*	1.10	0.78	0.49	0.37	0.60	2.57*
Authoritarian Parenting - Wald F(3, 20)	1.36	1.46	3.55**	0.18	0.31	2.85*	1.64	1.50

Note: The tests reported in the table stem from a specification, including: “time” dichotomies, the main effect of the covariates and the covariates’ interaction with the time dichotomies. Error terms are clustered by daycare and the data are for the first four assessments. Com= Communication, Awa = Self-Awareness, Cog = Cognition, Phy = Physical, Voc. = Vocabulary, Freq Lit. =Frequency of Literacy Activities, Language Lit. = Language of Literacy Activities. Significance levels set at *** $p < 1\%$; ** $p < 5\%$; * $p < 10\%$.

Appendix G: Analysis Results for the ÉVIP-R and EOWPVT-F Scales

Table G.1: Program Impact on French Vocabulary: Standardized Receptive Vocabulary (ÉVIP-R) and Expressive Vocabulary (EOWPVT-F) Scales

Type of Difference	Receptive Vocabulary (ÉVIP-R)				Expressive Vocabulary (EOWPVT-F)			
	Unadjusted (N = 333)		Adjusted (N = 328)		Unadjusted (N = 328)		Adjusted (N = 321)	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
G1 vs. G2								
Raw score	-3.64	2.43	-3.74**	1.33	-3.09**	1.48	-2.66	1.64
Standardized score	-0.20	0.13	-0.20**	0.07	-0.23**	0.11	-0.19	0.12
G1 vs. G3								
Raw score	-2.10	3.63	-3.59**	1.64	-1.42	2.91	-3.37	2.48
Standardized score	-0.11	0.20	-0.20**	0, 09	-0.10	0.21	-0.25	0.18

*Note: The negative values of DinD estimates for all group comparisons are a positive treatment effect (i.e., an advantage for the Program Daycare group). Standardized scores are raw data converted to a Z-score with a mean of zero and a standard deviation of 1. Assignment to treatment groups at the time of scale administrations is used for the comparisons reported in this table. The ÉVIP-R was administered at 20 months and the EOWPVT-F, at 24 months. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.*

Appendix H: Analysis Results for the ÉPE-AD Scales (Second Year)

Table H.1: Unadjusted Program Impact on School Readiness in the Second Year – Standardized Scores

Type of Difference	School Readiness Indicator							
	Communication		Self-Awareness		Cognition		Expressive Vocabulary	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
G1 vs. G2								
Baseline	0.44	0.31	0.35	0.25	0.26**	0.12	0.42	0.26
12 months (DinD)	-0.34***	0.11	-0.22*	0.11	-0.16	0.14	-0.26**	0.11
16 months (DinD)	-0.23**	0.09	-0.15*	0.07	-0.15	0.19	-0.12	0.12
20 months (DinD)	-0.32	0.20	-0.35***	0.09	-0.18*	0.09	-0.38***	0.11
24 months (DinD)	-0.38***	0.12	-0.36***	0.10	-0.28*	0.14	-0.40***	0.12
G1 vs. G3								
Baseline	0.24	0.32	0.21	0.28	0.07	0.10	0.30	0.26
12 months (DinD)	-0.38**	0.17	-0.07	0.07	-0.07	0.13	-0.08	0.13
16 months (DinD)	-0.21	0.16	0.20*	0.11	0.12	0.22	0.20	0.19
20 months (DinD)	-0.31	0.20	-0.32*	0.18	-0.26*	0.15	-0.38	0.26
24 months (DinD)	-0.25*	0.13	-0.49***	0.16	-0.25	0.25	-0.50**	0.21
Group x time Wald F	3.02**		5.69***		2.99**		2.28*	

Note: Contrary to the figures, the negative values of the DinD estimates presented in this table are a positive treatment effect (i.e., an advantage for the Program Daycare group). The degrees of freedom for the Wald F-test are 8 and 15. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Table H.2: Adjusted Program Impact on School Readiness in the Second Year – Standardized Scores

Type of Difference	School Readiness Indicator							
	Communication		Self-Awareness		Cognition		Expressive Vocabulary	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
G1 vs. G2								
Baseline	0.30***	0.08	0.20**	0.08	0.09	0.09	0.28***	0.09
12 months (DinD)	-0.41***	0.11	-0.19*	0.10	-0.09	0.11	-0.29**	0.11
16 months (DinD)	-0.17	0.11	-0.18**	0.08	0.01	0.14	-0.11	0.11
20 months (DinD)	-0.20*	0.11	-0.31***	0.08	-0.01	0.11	-0.35***	0.08
24 months (DinD)	-0.27**	0.13	-0.31***	0.08	-0.13	0.10	-0.37***	0.11
G1 vs. G3								
Baseline	-0.01	0.05	-0.04	0.10	-0.06	0.09	0.01	0.08
12 months (DinD)	-0.38***	0.09	-0.02	0.12	-0.09	0.08	-0.07	0.11
16 months (DinD)	-0.23	0.15	0.16	0.11	0.22	0.14	0.17	0.17
20 months (DinD)	-0.18	0.15	-0.27	0.16	-0.06	0.09	-0.35	0.20
24 months (DinD)	-0.07	0.15	-0.38**	0.15	-0.02	0.06	-0.40**	0.16
Group x time Wald F	3.18**		14.30***		2.65**		5.19***	

Note: Contrary to the figures, the negative values of the DinD estimates presented in this table are a positive treatment effect (i.e., an advantage for the Program Daycare group). The degrees of freedom for the Wald F-test are 8 and 15. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Appendix I: Results of Impact Analyses by Linguistic Profile

Table I.1: Tests of the Hypothesis that the Household Linguistic Profile Serves as the Moderator of Program Effect in the Second Year

Moderation Test	School Readiness Indicator							
	Communication		Self-Awareness		Cognition		Expressive Vocabulary	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
G1 vs. G2								
Time 4 (DinDinD)	0.22	0.15	0.00	0.19	-0.55***	0.19	0.12	0.24
Time 5 (DinDinD)	0.52***	0.18	0.17	0.20	-0.30	0.22	0.31	0.23
Time 6 (DinDinD)	0.26	0.22	0.03	0.21	-0.31	0.32	0.06	0.22
Time 7 (DinDinD)	-0.20	0.25	-0.15	0.21	-0.25	0.29	-0.14	0.24
G1 vs. G3								
Time 4 (DinDinD)	0.43**	0.19	-0.40**	0.16	-0.67***	0.21	-0.09	0.28
Time 5 (DinDinD)	0.56*	0.31	-0.21	0.18	-0.56	0.35	0.17	0.35
Time 6 (DinDinD)	0.64**	0.30	0.43	0.38	0.13	0.35	0.65	0.55
Time 7 (DinDinD)	0.00	0.36	0.26	0.43	-0.17	0.28	0.46	0.60

Note: The DinDinD estimates represent the difference in program impact for children from households with high and low exposure. The benchmark category is “low exposure.” *The DinDinD effects are interpreted as follows: (a) the positive estimates indicate that the tested program’s effect decreases with high exposure to French; and (b) the negative estimates indicate that the tested program’s effect is enhanced with higher exposure to French.* The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.

Table I.2: Impact of the Tested Program by Household Linguistic Profile – (Year Two) – Adjusted Scores

School Readiness Indicator								
Type of Difference	<i>Household with Low Exposure to French</i>							
	Communication		Self-Awareness		Cognition		Expressive Vocabulary	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
G1 vs. G2								
Baseline	0.40***	0.10	0.27**	0.11	0.11	0.14	0.34***	0.10
Time 4 (DinD)	-0.50***	0.11	-0.21	0.14	0.15	0.15	-0.35**	0.13
Time 5 (DinD)	-0.39***	0.11	-0.24*	0.13	0.12	0.19	-0.31**	0.13
Time 6 (DinD)	-0.29*	0.15	-0.34**	0.15	0.13	0.20	-0.39***	0.13
Time 7 (DinD)	-0.14	0.19	-0.25	0.16	-0.04	0.22	-0.30*	0.15
G1 vs. G3								
Baseline	-0.23**	0.10	-0.29**	0.12	-0.36**	0.16	-0.10	0.18
Time 4 (DinD)	-0.59***	0.09	0.25*	0.12	0.28	0.18	-0.00	0.12
Time 5 (DinD)	-0.48**	0.20	0.37**	0.14	0.59**	0.25	0.06	0.15
Time 6 (DinD)	-0.46*	0.26	-0.50*	0.28	-0.11	0.23	-0.70*	0.36
Time 7 (DinD)	0.00	0.34	-0.48	0.30	0.11	0.18	-0.59	0.40
Type of Difference	<i>Household with High Exposure to French</i>							
	Communication		Self-Awareness		Cognition		Expressive Vocabulary	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
G1 vs. G2								
Baseline	0.19	0.11	0.15	0.12	0.15	0.11	0.24	0.15
Time 4 (DinD)	-0.30**	0.14	-0.21	0.14	-0.40***	0.12	-0.23	0.19
Time 5 (DinD)	0.09	0.15	-0.07	0.12	-0.17	0.16	0.00	0.17
Time 6 (DinD)	-0.07	0.16	-0.31**	0.12	-0.19	0.17	-0.33**	0.14
Time 7 (DinD)	-0.37**	0.16	-0.40***	0.11	-0.29**	0.13	-0.44**	0.17
G1 vs. G3								
Baseline	0.09	0.08	0.05	0.14	0.12	0.13	0.07	0.17
Time 4 (DinD)	-0.19	0.13	-0.15	0.18	-0.40***	0.10	-0.09	0.22
Time 5 (DinD)	0.03	0.18	0.16	0.16	0.03	0.20	0.23	0.29
Time 6 (DinD)	0.12	0.12	-0.07	0.18	0.02	0.16	-0.04	0.27
Time 7 (DinD)	-0.05	0.09	-0.22	0.19	-0.06	0.13	-0.13	0.25

*Note: Contrary to the figures, the negative values of the DinD effects reported in this table comparing the DinD groups are a positive treatment effect (i.e., an advantage for the Program Daycare group). Household Type consists of the language (i.e., French, English, etc.) the child usually speaks to the mother crossed with the language the child usually speaks to the father: household with high exposure to French, other. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.*

Table I.3: Impact of the Tested Program on Vocabulary by Linguistic Profile: Standardized Receptive Vocabulary (ÉVIP-R) and Expressive Vocabulary (EOWPVT-F) Scales

School Readiness Indicator								
<i>Household with Low Exposure to French</i>								
Type of Difference	Receptive Vocabulary (ÉVIP-R)				Expressive Vocabulary (EOWPVT-F)			
	Initial Model		Adjusted Model		Initial Model		Adjusted Model	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
G1 vs. G2								
Raw score	-5.96***	1.58	-6.40***	1.39	-2.44	1.62	-2.64*	1.42
Standardized score	-0.33***	0.09	-0.36***	0.08	-0.18	0.12	-0.20*	0.11
G1 vs. G3								
Raw score	-2.82**	1.31	-3.41**	1.53	0.17	2.01	-0.42	1.44
Standardized score	-0.16**	0.07	-0.19***	0.09	0.01	0.15	-0.03	0.11
<i>Household with High Exposure to French</i>								
Type of Difference	Receptive Vocabulary (ÉVIP-R)				Expressive Vocabulary (EOWPVT-F)			
	Initial Model		Adjusted Model		Initial Model		Adjusted Model	
	Diff.	SE	Diff.	SE	Diff.	SE	Diff.	SE
G1 vs. G2								
Raw score	-0.58	2.98	-0.79	2.23	-2.52	1.49	-2.33	1.94
Standardized score	-0.03	0.17	-0.04	0.13	-0.19	0.11	-0.17	0.14
G1 vs. G3								
Raw score	-1.95	5.92	-3.10	3.80	-5.89	4.98	-6.37	3.89
Standardized score	-0.11	0.33	-0.17	0.21	-0.44	0.37	-0.47	0.29

*Note: Contrary to the figures, the negative values of the effects presented in this table are a positive treatment effect (i.e., an advantage for the Program Daycare group). N = 321 for Expressive Vocabulary (EOWPVT-F) and N = 328 for Receptive Vocabulary (ÉVIP-R). Standardized scores are raw data converted to a Z-score with a mean of zero and a standard deviation of 1. Assignment to study groups for scale administration is used for the comparisons reported in this table. The ÉVIP-R was administered at 20 months and the EOWPVT-F, at 24 months. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance levels set at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SE = standard error.*

Appendix J: Analysis Results for Frequency and Language of Literacy Activities

Table J.1: Impact of the Tested Program on Frequency of Literacy Activities as well as Language of Literacy Activities (Unadjusted)

Type of Difference	Parents' Behaviour at Home							
	Frequency of Activities				Language of Activities			
	Raw Score		Standardized Score		Raw Score		Standardized Score	
	Diff.	SD	Diff.	SD	Diff.	SD	Diff.	SD
G1 vs. G2								
Baseline	-.918*	.488	-.312*	.166	.933	2.151	.156	.360
4 months (DinD)	-.005	.673	.001	.228	-.987	.616	-.166*	.093
8 months (DinD)	-.007	.438	.001	.148	-.580	.713	-.089	.099
12 months (DinD)	-.374	.633	-.121	.214	-.775	.976	-.123	.125
16 months (DinD)	.106	.750	.047	.248	-.724	1.677	-.131	.271
20 months (DinD)	.290	.704	.108	.232	-1.230	1.398	-.198	.249
24 months (DinD)	.289	.759	.096	.259	-1.313	1.325	-.214	.230
G1 vs. G3								
Baseline	0.244	.674	.083	.229	2.051	2.244	.344	.376
4 months (DinD)	-1.262	.895	-.426	.303	-1.662***	.577	-.274***	.092
8 months (DinD)	-.915	.624	-.308	.212	-2.000**	.782	-.334**	.125
12 months (DinD)	-1.290*	.743	-.433*	.252	-1.322	.859	-.191	.116
16 months (DinD)	-.771	1.038	-.255	.346	-3.381*	1.741	-.506*	.282
20 months (DinD)	-1.001	.924	-.329	.308	-3.079**	1.331	-.487*	.242
24 months (DinD)	-.884	.961	-.302	.327	-2.893**	1.344	-.472*	.235
Group x Time Wald F	3.04**		2.92**		3.407**		4.78***	

Note: The degrees of freedom for the Wald F-test are 12 and 11. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance level at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SD = standard error.

Table J.2: Adjusted Impact of the Tested Program on Frequency of Literacy Activities and Language of Literacy Activities

Type of Difference	Parents' Behaviour at Home							
	Frequency of Activities				Language of Activities			
	Raw Score		Standardized Score		Raw Score		Standardized Score	
	Diff.	SD	Diff.	SD	Diff.	SD	Diff.	SD
G1 vs. G2								
Baseline	-0.009	.016	-0.003	.005	-0.016	.028	-0.002	.005
4 months (DinD)	-0.301	.298	-0.102	.101	-0.895**	.401	-0.161**	.072
8 months (DinD)	-0.569**	.267	-0.191**	.090	-0.562	.363	-0.092	.067
12 months (DinD)	-0.909**	.353	-0.304**	.118	-0.563	.401	-0.118	.084
16 months (DinD)	-0.492	.373	-0.161	.122	-1.361	.933	-0.093	.123
20 months (DinD)	-0.383	.513	-0.124	.167	-1.162**	.477	-0.185**	.069
24 months (DinD)	-0.441	.342	-0.151	.117	-0.234	.429	-0.035	.064
G1 vs. G3								
Baseline	0.006	.035	0.002	.012	0.027	.043	0.007	.007
4 months (DinD)	-1.041***	.249	-0.351***	.084	-1.309**	.511	-0.235**	.092
8 months (DinD)	-0.683***	.197	-0.229***	.066	-1.38***	.418	-0.249***	.074
12 months (DinD)	-1.093***	.254	-0.366***	.085	-0.49	.517	-0.103	.108
16 months (DinD)	-0.63	.397	-0.206	.130	-2.83***	.628	-0.273***	.086
20 months (DinD)	-0.848*	.488	-0.276*	.159	-1.021	.673	-0.207**	.089
24 months (DinD)	-0.723*	.364	-0.248*	.125	-0.396	.628	0.016	.101
Group x Time Wald F	4.382**		4.383**		7.429***		4.084**	

Note: The degrees of freedom for the Wald F-test are 12 and 11. The Huber-White robust standard errors are reported with error terms clustered by daycare. Significance level at *** $p < 1\%$, ** $p < 5\%$; * $p < 10\%$. Diff = difference, SD = standard error.

SOCIAL RESEARCH AND DEMONSTRATION CORPORATION

OTTAWA OFFICE

55 Murray Street, Suite 400
Ottawa, Ontario K1N 5M3
Tel.: 613.237.4311
Fax: 613.237.5045

TORONTO OFFICE

481 University Avenue, Suite 705
Toronto, Ontario M5G 2E9
Tel.: 416.593.0445
Fax: 647.725.6293

VANCOUVER OFFICE

128 West Pender Street, Suite 301
Vancouver, British Columbia V6B 1R8
Tel.: 604.601.4070
Fax: 604.601.4080

www.srdc.org