



# **CAPACITÉ D'APPRENTISSAGE DANS LES COMMUNAUTÉS FRANCOPHONES EN SITUATION MINORITAIRE**

**RAPPORT DES RÉSULTATS DE LA PREMIÈRE COHORTE**



## Conseil d'administration de la SRSA

Richard A. Wagner  
Associé principal, Norton Rose Fulbright S.E.N.C.R.L., s.r.l.

Maria David-Evans  
Présidente sortante de l'IAPC et ancienne sous-ministre,  
Gouvernement de l'Alberta

Robert Flynn  
Professeur émérite, École de psychologie, Université d'Ottawa

John Helliwell  
Codirecteur, programme Interactions sociales, identité et  
mieux-être, Institut canadien de recherches avancées

Suzanne Herbert  
Ancienne sous-ministre, Gouvernement de l'Ontario

Guy Lacroix, Ph. D.  
Professeur d'économie, Université Laval

Renée F. Lyons, Ph. D.  
Présidente de la recherche sur les maladies chroniques  
complexes et directrice scientifique de TD du Bridgepoint  
Collaboratory for Research and Innovation, Université de Toronto

Sharon Manson Singer  
Ancienne présidente des Réseaux canadiens de recherche en  
politiques publiques

Jim Mitchell  
Partenaire fondateur de la firme d'experts-conseils en politiques  
Sussex Circle

## Président et chef de la direction de la SRSA

Jean-Pierre Voyer

## Auteurs

Glenn Thompson, Ph. D.

Louise Legault, Ph. D.

Paul Lalonde

Annie Bérubé, Ph. D.

La Société de recherche sociale appliquée (SRSA) est un organisme de recherche sans but lucratif, créé dans le but précis d'élaborer, de mettre à l'essai sur le terrain et d'évaluer rigoureusement de nouveaux programmes. Notre mission, qui comporte deux volets, consiste à aider les décideurs et les intervenants à déterminer les politiques et programmes qui améliorent le bien-être de tous les Canadiens, en se penchant particulièrement sur les effets qu'ils auront sur les personnes défavorisées, et à améliorer les normes relatives aux éléments probants utilisées pour évaluer ces politiques.

Depuis sa création en décembre 1991, la SRSA a réalisé plus de 200 projets et études pour différents ministères fédéraux et provinciaux, des municipalités ainsi que d'autres organismes publics et sans but lucratif. La SRSA a des bureaux à Ottawa, à Toronto et à Vancouver.

## Promoteur du projet Capacité d'apprentissage

Le projet Capacité d'apprentissage a été financé par Emploi et Développement social Canada.



Le rapport d'analyse contenu dans la présente publication est la responsabilité des auteurs et ne représente pas le point de vue du Gouvernement du Canada.

# Table des matières

<b>Sommaire</b> .....	<b>1</b>
<b>1. Présentation du projet Capacité d'apprentissage</b> .....	<b>4</b>
1.1    Genèse du projet .....	4
1.2    Le projet Capacité d'apprentissage.....	7
1.3    L'apport du projet .....	9
1.4    Le programme testé.....	9
1.4.1    Le volet en garderie .....	10
1.4.2    Le volet Ateliers familles.....	10
1.4.3    Arrimage entre le volet garderie et le volet Ateliers familles.....	11
1.4.4    Déroulement du programme durant la première année du projet Capacité d'apprentissage.....	12
1.4.5    Déroulement du programme durant la deuxième année du projet Capacité d'apprentissage.....	13
1.5    Le modelage du développement des enfants en milieu minoritaire.....	13
1.5.1    Le microsystème : environnement familial (variables contextuelles) .....	15
1.5.2    Le microsystème : environnement familial (processus familiaux).....	17
1.5.3    Le microsystème : milieu de garde .....	18
1.5.4    Le mésosystème : liens entre la famille et le milieu de garde .....	19
1.5.5    Le macrosystème : variables communautaires .....	20
1.6    Une définition de la préparation à la scolarisation.....	21
1.7    Le présent rapport .....	21
<b>2. Méthodologie</b> .....	<b>23</b>
2.1    Population ciblée.....	23
2.2    Devis expérimental .....	23
2.3    Validité interne.....	24
2.3.1    La taille de l'échantillon .....	24
2.3.2    Stratégie d'échantillonnage ciblé.....	25
2.3.3    Les mesures pré-intervention.....	25
2.3.4    Attrition.....	26
2.3.5    Prévention de la contamination des groupes témoins.....	26
2.4    Échantillon par communautés et par groupes expérimentaux .....	27
2.4.1    Taux de rétention .....	29

2.5	Mesures .....	29
2.5.1	Plan des collectes de données .....	30
2.5.2	Mesures auprès des enfants.....	32
2.5.3	Mesures auprès des parents.....	38
2.5.4	Connaissances, attitudes et croyances des parents.....	42
2.5.5	Dosage.....	44
2.5.6	Indices de la mise en œuvre du volet en garderie .....	44
2.5.7	Étude de la mise en œuvre des Ateliers familles .....	46
2.6	Hypothèses à l'étude .....	46
2.6.1	L'impact du volet en garderie sur les dimensions linguistiques des enfants .....	46
2.6.2	L'impact du volet en garderie sur les dimensions identitaire et culturelle des enfants .....	48
2.6.3	L'impact du volet Ateliers familles sur les attitudes et comportements des parents .....	48
<b>3.</b>	<b>Stratégie d'analyse .....</b>	<b>50</b>
3.1	Conceptualisation du traitement et du dosage .....	50
3.1.1	Exposition hétérogène au traitement.....	50
3.1.2	Évaluation de l'effet du dosage ou de l'exposition.....	51
3.1.3	Évaluation de l'effet de la fidélité/qualité du programme en garderie .....	53
3.2	Approche statistique.....	54
3.2.1	Modélisation linéaire hiérarchique (MLH).....	55
3.2.2	Estimateur par différence de différences .....	56
3.2.3	Contrôle statistique dans le cadre d'une étude longitudinale.....	57
3.2.4	Ateliers d'alphabétisation familles : Un cas spécial .....	59
<b>4.</b>	<b>Analyses préliminaires.....</b>	<b>62</b>
4.1	Processus de contrôle de la qualité .....	62
4.2	Analyse des valeurs manquantes .....	64
4.2.1	Répartition générale des valeurs manquantes .....	65
4.2.2	Patron des données manquantes.....	68
4.2.3	Stratégie d'imputation des données .....	74
4.3	Spécification des termes d'erreurs .....	77
4.4	Liste des variables de contrôle.....	79
4.5	Représentativité de l'échantillon : projet Capacité d'apprentissage vs. EVMLO (Enquête sur la vitalité des minorités de langue officielle).....	84
4.5.1.	Statut d'immigrant et profil linguistique.....	85

4.5.2. Caractéristiques sociodémographiques .....	88
4.6. Sommaire .....	92
<b>5. Analyse d'impact : Première année.....</b>	<b>95</b>
5.1    Impacts sur les enfants .....	96
5.1.1    Variables retenues pour les analyses.....	96
5.1.2    Résultats de la comparaison des groupes.....	98
5.1.3    Modérateurs et médiateurs de l'effet d'intervention.....	109
5.2    Impacts sur les parents .....	150
5.2.1    Fréquence et langue des activités de littératie.....	150
5.2.2    Effets pré-post: Connaissances, auto-efficacité et comportements de modélisation déclarés par les répondants .....	156
5.2.3    Modérateurs des effets sur les parents .....	161
5.3    Sommaire .....	164
<b>6. Analyses d'impact : Suivi de la deuxième année.....</b>	<b>170</b>
6.1    Variables retenues pour les analyses.....	170
6.1.1    Prédicteurs et covariables significatifs.....	170
6.1.2    Mesure de résultantes.....	171
6.1.3    Résultats de la comparaison des groupes.....	172
6.2    Modérateurs et médiateurs de l'effet d'intervention.....	180
6.2.1    Définitions précises du dosage: Moyenne d'heures passées en garderie durant la première année .....	180
6.2.2    Définitions précises de l'intégrité du programme: Fidélité et qualité du programme en garderie.....	185
6.2.3    Caractéristiques linguistiques de l'échantillon .....	190
6.3    Sommaire .....	199
<b>7. Discussion et conclusion .....</b>	<b>201</b>
7.1.    Préparation à fréquenter l'école en français?.....	205
7.1.1    Effets moyens du programme .....	205
7.1.2.    Effets du programme en fonction du profil linguistique.....	208
7.2    Résultantes des parents .....	210
7.3    Effets différenciés du dosage .....	212
7.4.    Fidélité et qualité des garderies.....	212
7.5    Limites et travaux futurs .....	214
7.6    Conclusion .....	216
<b>Références .....</b>	<b>217</b>

<b>Annexe A: Calendrier et outils d'évaluation .....</b>	<b>229</b>
<b>Annexe B : Procédure pour l'administration de l'ÉPE-AD (Mesure pré-intervention) ...</b>	<b>230</b>
<b>Annexe C : Une comparaison des différentes versions de l'ÉPE-AD.....</b>	<b>232</b>

## Tableaux et figures

Figure 1.1 : Localisation des six communautés participantes en fonction des limites géographiques de Statistique Canada, 2006 .....	8
Figure 1.2 : Modèle écologique complet de Bronfenbrenner (1979) .....	15
Tableau 2.1 : Répartition des participants par communauté depuis l'inscription.....	28
Tableau 2.2 : Répartition des participants par groupe expérimental .....	29
Tableau 2.3 : Raisons justifiant le retrait de l'enfant du projet Capacité d'apprentissage.....	29
Tableau 2.4 : Taux de réponses des sondages auprès des parents .....	30
Tableau 2.5 : Taux de réponses des sondages liés aux Ateliers familles.....	31
Tableau 2.6 : Taux de réponses des évaluations auprès des enfants.....	32
Tableau 2.7 : Alpha de Cronbach des domaines de l'ÉPE-AD par vague d'évaluation .....	34
Tableau 2.8 : Alpha de Cronbach des sous-échelles de vocabulaire de l'ÉPE-AD par vague d'évaluation .....	36
Tableau 2.9 : Coefficients de corrélation Pearson des échelles de l'ÉPE-AD et des échelles de vocabulaire expressif et réceptif .....	38
Tableau 2.10 : Alpha de Cronbach de l'échelle des activités de littératie et de l'échelle des langues utilisées lors des activités de littératie à chaque période de sondage .....	41
Tableau 4.1 : Statistiques descriptives des cas d'attrition des variables corrélées avec l'attrition .....	70
Tableau 4.2 : Nombre d'évaluations manquantes durant la première année du projet en fonction de l'appartenance au groupe expérimental.....	72
Tableau 4.3 : Valeurs françaises imputées de l'ÉPE-AD en fonction de l'évaluation et du groupe expérimental .....	76
Tableau 4.4 : Statistiques descriptives, différences observées entre les groupes expérimentaux quant aux caractéristiques observées et leur association avec les résultantes des enfants ....	81
Tableau 4.5 : Comparaison entre le projet Capacité d'apprentissage et l'EVMLO .....	86
Tableau 4.6 : Comparaison entre le projet Capacité d'apprentissage et l'EVMLO – Enfants groupés par langue maternelle.....	87

Tableau 4.7 : Comparaison entre les mères du projet Capacité d'apprentissage et de l'EVML0 – Mères groupées par langue maternelle .....	87
Tableau 4.8 : Comparaison entre les pères du projet Capacité d'apprentissage et de l'EVML0 – Pères groupés par langue maternelle .....	88
Tableau 4.9 : Comparaison entre le projet Capacité d'apprentissage et l'EVML0 – Familles par classe de revenu .....	89
Tableau 4.10 : Comparaison des degrés de scolarité des mères du projet Capacité d'apprentissage et de l'EVML0.....	90
Tableau 4.11 : Comparaison entre les degrés de scolarité des pères du projet Capacité d'apprentissage et de l'EVML0.....	90
Tableau 4.12 : Comparaison de la taille des familles <sup>1</sup> dans le projet Capacité d'apprentissage et l'EVML0 .....	91
Tableau 4.13 : Comparaison entre le projet Capacité d'apprentissage et l'EVML0 – Nombre d'enfants par répondant .....	91
Tableau 4.14 : Comparaison entre le projet Capacité d'apprentissage et l'EVML0 – Nombre de familles monoparentales et biparentales.....	92
Tableau 5.1 : Première année – Différence des différences (DD) : Effets programme pour les scores de communication standardisés de l'ÉPE–AD (version française, échantillon total).....	103
Tableau 5.3 : Première année – Effets programme basés sur l'estimateur DDD (temps x groupe x dosage) et sur les scores de Communication de l'ÉPE–AD (version française, échantillon total) .....	113
Tableau 5.4 : Première année – Effets programme basés sur l'estimateur DDD des effets programme (temps x groupe x dosage) pour les scores des sous-échelles standardisées de l'ÉPE–AD (participants ayant répondu aux échelles en français seulement) .....	115
Tableau 5.5 : Corrélations entre les indices de fidélité et de qualité .....	119
Tableau 5.6 : Fidélité du programme, comparaison entre les groupes et à travers le temps .....	121
Tableau 5.7 : Comparaison de la qualité entre les garderies programme et les garderies témoins .....	123
Tableau 5.8 : Estimation DD de l'effet de la fidélité sur les scores de Communication standardisés (temps x indices de fidélité) – Version française.....	124



Tableau 5.9 : Estimation DD de l'effet de la fidélité sur les scores de Conscience de soi standardisés (temps x indices de fidélité) – (participants ayant répondu au test en français seulement).....	125
Tableau 5.10 : Estimation DD de l'effet de la fidélité sur les scores d'Habilités cognitives standardisés (temps x indices de fidélité) – (participants ayant répondu au test en français seulement).....	126
Tableau 5.11 : Estimation DD de l'effet de la fidélité sur les scores d'Habilité physique standardisés (temps x indices de fidélité) – (participants ayant répondu au test en français seulement).....	127
Tableau 5.12 : Estimation DD de l'effet de la fidélité sur les scores de Vocabulaire réceptif standardisés (temps x indices de fidélité) – (participants ayant répondu au test en français seulement).....	127
Tableau 5.13 : Estimation DD de l'effet de la fidélité sur les scores de Vocabulaire expressif standardisés (temps x indices de fidélité) – (participants ayant répondu au test en français seulement).....	128
Tableau 5.14 : Estimation DD de l'effet de la qualité sur les scores de Communication standardisés (temps x indices de qualité) – (version française) .....	129
Tableau 5.15 : Estimation DD de l'effet de la qualité sur les scores de Conscience de soi standardisés (temps x indices de qualité) – (participants ayant répondu au test en français seulement).....	130
Tableau 5.16 : Estimation DD de l'effet de la qualité sur les scores d'Habilités cognitives standardisés (temps x indices de qualité) – (participants ayant répondu au test en français seulement).....	130
Tableau 5.17 : Estimation DD de l'effet de la qualité sur les scores d'Habilité physique standardisés (temps x indices de qualité) – (participants ayant répondu au test en français seulement).....	131
Tableau 5.18 : Estimation DD de l'effet de la qualité sur les scores de Vocabulaire réceptif standardisés (temps x indices de qualité) – (participants ayant répondu au test en français seulement).....	131
Tableau 5.19 : Estimation DD de l'effet de la qualité sur les scores de Vocabulaire expressif standardisés (temps x indices de qualité) – (participants ayant répondu au test en français seulement).....	132

Table 5.20 : Scores de Communication, d'Habilités cognitives et de Vocabulaire expressif en fonction de la qualité de lecture (temps x qualité) – (participants ayant répondu au test en français seulement) .....	133
Table 5.21 : Corrélations d'ordre zéro entre les indicateurs de profils linguistiques .....	136
Tableau 5.22 : Effets estimés du programme sur les scores de Communication (version française) conditionnels au profil linguistique (échantillon total).....	139
Tableau 5.23 – Effets estimés du programme sur les scores standardisés d'Habilités cognitives (version française) conditionnels au profil linguistique.....	142
Tableau 5.24 : Effets estimés du programme sur les scores standardisés d'Habilité physique (version française) conditionnels au profil linguistique.....	146
Tableau 5.25 : Effets estimés du programme sur les scores standardisés de Vocabulaire expressif (version française) conditionnels au profil linguistique.....	148
Tableau 5.26 : Effets différence des différences (DD) du groupe pour la fréquence déclarée des activités de littératie au foyer (scores standardisés) .....	153
Tableau 5.27 : Effets différence des différences (DD) du groupe pour la langue déclarée des activités de littératie au foyer (scores standardisés).....	155
Tableau 5.28 : Comparaison des mesures déclarées par les parents pré/post-ateliers et participants v. non-participants .....	157
Tableau 5.29 : Changement des Connaissances, de l'Auto-efficacité et de la Modélisation déclarés par les répondants en fonction de la qualité des ateliers .....	163
Tableau 6.1 : Deuxième année – Différence des différences (DD) : Effets programme pour les scores de Communication standardisés de l'ÉPE-AD (version française, échantillon total).....	174
Tableau 6.2 : Deuxième année – Différence des différences (DD) : Effets programme pour les scores standardisés des sous-échelles de l'ÉPE-AD (participants ayant répondu au test en français seulement).....	175
Tableau 6.3 : Scores de l'EOWPVT (recueillis à la sixième évaluation, juin 2009) en fonction de l'appartenance au groupe aux quatre périodes d'évaluation de la deuxième année (non normalisés).....	178
Tableau 6.4 : Scores de l'ÉVIP-R (recueillis à la septième évaluation, octobre 2009) en fonction de l'appartenance au groupe aux quatre périodes d'évaluation de la deuxième année (non normalisés).....	179

Tableau 6.5 : Groupes expérimentaux et dosage de la première année comme prédicteurs des résultantes de la deuxième année (participants ayant répondu au test en français seulement) .....	182
Tableau 6.6 : Effets de l'exposition à la garderie durant la première année en fonction du type de garderie – Scores de l'ÉVIP-R et de l'EOWPVT (non normalisés).....	184
Tableau 6.7 : Qualité et fidélité globales du programme comme prédicteurs des résultantes de la deuxième année .....	187
Tableau 6.8 : Scores de Vocabulaire de la deuxième année en fonction des indices globaux de fidélité et de qualité du programme – Scores de l'ÉVIP-R et de l'EOWPVT (non normalisés) .....	189
Tableau 6.9 : Effets estimés du programme pour les scores de Communication standardisés (français) conditionnels à la faible exposition au français – Échantillon total.....	192
Tableau 6.11 : Effets estimés du programme pour les scores standardisés de Vocabulaire expressif (français) conditionnels à la faible exposition au français.....	196
Tableau 6.12 : Effets estimés du programme sur l'EOWPVT pour les enfants à «faible exposition». ....	198
Tableau 7.1 : Sommaire des principales conclusions du Rapport des résultats de la première cohorte du projet Capacité d'apprentissage : – Résultantes des enfants .....	202
Tableau 7.2 : Sommaire des principales conclusions du Rapport des résultats de la première cohorte du projet Capacité d'apprentissage – Résultantes des parents .....	204
Table A1: Calendriers des évaluations des enfants.....	229
Tableau B1 : Arbre décisionnel pour déterminer la langue d'administration du test .....	231
Tableau C1 : Sous-échelle de Vocabulaire expressif.....	233
Tableau C2 : Sous-échelle de Conscience de soi.....	235
Tableau C3 : Sous-échelle de Conscience phonologique .....	236
Tableau C4 : Sous-échelle de Numératie.....	237
Tableau C5 : Mémoire pour les informations personnelles .....	237
Tableau C6 : Sous-échelle de connaissances alphabétiques.....	238
Tableau C7 : Lecture orale.....	238

# Sommaire

Ce rapport vise à documenter les impacts du nouveau programme préscolaire sur les enfants et leur famille participant au projet Capacité d'apprentissage dans les communautés francophones en situation minoritaire (titre abrégé : projet Capacité d'apprentissage; auparavant le Projet pilote de garde d'enfant), un projet de démonstration financé par Ressources humaines et Développement des compétences Canada (RHDCC). Le projet met à l'essai un programme préscolaire<sup>1</sup> qui jumelle un volet de service de garde spécifiquement développé pour satisfaire aux besoins des enfants francophones en milieu minoritaire à un volet d'alphabétisation familiale ciblant les parents de ces enfants. Ce programme a pour objectif de développer les compétences langagières de l'enfant, sa connaissance et son utilisation du français, sa connaissance de la culture francophone et son appartenance à celle-ci, ainsi que de favoriser sa préparation à l'école et son développement en général. L'évaluation de ce programme se fait au moyen d'une méthodologie quasi expérimentale avec groupes témoins non équivalents. La méthodologie prévoit trois groupes expérimentaux : le groupe programme composé d'enfants inscrits dans la garderie francophone offrant le nouveau programme préscolaire; le groupe témoin en garderie comprenant les enfants inscrits dans une garderie francophone n'offrant pas le nouveau programme; et le groupe témoin hors garderie regroupant les enfants dont la garde de jour a lieu à la maison ou en garderie familiale non réglementée. Le groupe témoin en garderie vise à tenir compte de l'influence d'une garderie en installation sur le développement des enfants, ce qui constitue en soit un traitement. Le groupe témoin hors garderie vise à tenir compte de l'influence d'un milieu de garde informel sur le développement de l'enfant. Le projet compte deux cohortes de participants, la première recrutée en 2007 et la deuxième recrutée en 2008.

Le présent rapport porte sur les données collectées de mai 2007 à octobre 2009 auprès de la première cohorte de participants. Cette période correspond au moment où les enfants sont âgés, en moyenne, entre trois et cinq ans. Une méthodologie mixte de recherche a été utilisée dans le cadre du projet Capacité d'apprentissage. Cette approche préconise l'utilisation d'une diversité d'outils, de nature autant quantitative que qualitative, provenant de plusieurs sources d'informations toutes choisies en fonction des objectifs de la recherche. La richesse des informations collectées facilite la triangulation des résultats de recherche permettant ainsi d'assurer la rigueur des conclusions tirées. De plus, la complémentarité des données collectées permet d'obtenir un portrait plus complet voire plus nuancé du phénomène à l'étude. Ainsi, les analyses ont été effectuées à partir des données provenant, entre autres, des évaluations des enfants, des sondages auprès des parents, des observations en garderie et des Ateliers familles ainsi que des données administratives (par ex. le registre des présences et absences des enfants à la garderie).

Le plan des analyses d'impact prévoit deux principales séries d'analyses. L'une cherche à établir les retombées du nouveau programme préscolaire en garderie sur le développement des enfants. L'autre examine les effets du volet Ateliers familles sur les attitudes et les comportements des parents. Les résultats de l'ensemble des analyses suggèrent que le programme a un impact positif sur les résultantes des enfants et des parents. Spécifiquement, on

---

<sup>1</sup> Officiellement connu sous le nom de services de garde enrichis dans les documents de RHDCC, la Société de recherche sociale appliquée, ou SRSA, en accord avec RHDCC, nomme dorénavant le programme « programme préscolaire de service de garde ».

observe des gains significatifs chez les enfants du groupe programme sur toutes les dimensions de préparation à la scolarisation comparativement à ceux observés chez le groupe témoin en garderie<sup>2</sup>. Des gains significatifs sont observés au premier posttest et à la mi-parcours de la seconde année du projet. Quant au programme du volet Ateliers familles, les résultats des analyses démontrent que le programme a un impact quant à certaines dimensions attitudinales. Ainsi, les parents rapportent des gains significatifs au niveau des connaissances sur le développement des enfants et au niveau de leur sentiment d'efficacité suite aux ateliers. Ces résultats suggèrent qu'une partie des impacts observés chez les enfants proviendrait des changements survenus chez les parents et seraient, par conséquent, attribuables au volet Ateliers familles de l'intervention. L'image dégagée de l'ensemble des analyses permet de conclure que le programme testé a un impact modeste sur le niveau de préparation à la scolarisation ainsi que sur le développement langagier des enfants du groupe programme.

Deux limites de l'étude permettent de mettre en perspective ces résultats. Dans un premier temps, l'outil principal pour mesurer le niveau de préparation à la scolarisation des enfants s'est avéré être un enjeu. Dans sa forme originelle, l'ÉPE-AD remplit bien l'objectif d'identifier les enfants accusant un retard développemental ou ceux pouvant bénéficier d'une aide ciblée afin qu'ils puissent acquérir les habiletés ou les compétences nécessaires pour effectuer certaines tâches habituellement observées chez leurs pairs. Toutefois, dans le contexte d'une évaluation de programme, ce genre d'outil n'est pas assez sensible pour mesurer de fines différences dans le développement d'habiletés et de compétences chez les enfants. Spécifiquement, les résultantes mesurées par l'ÉPE-AD ne fournissent pas d'information sur les *processus développementaux* qui sont considérés dans la littérature scientifique comme étant de bons prédicteurs de résultantes éducatives à plus long terme (Hirsh-Pasek, Kochanoff, Newcombe et de Villiers, 2005). Il s'en suit que les impacts du programme sont en fait sous-estimés dans la présente étude sur le niveau de préparation à la scolarisation et les habiletés langagières.

Dans un deuxième temps, l'échantillon relativement petit du projet Capacité d'apprentissage tant au niveau des participants (environ 250 enfants) qu'au niveau des groupes expérimentaux (moins de 100 enfants par groupe expérimental) entraîne un manque de puissance statistique et de stabilité, empêchant la conduite de certaines analyses plus poussées qui auraient été très intéressantes à mener. De plus, l'adoption d'un devis quasi expérimental comporte en soit le risque que les résultats observés soient causés par l'effet d'une variable non contrôlée dans les analyses. La conduite d'analyses incorporant les données recueillies auprès des deux cohortes est recommandée afin de confirmer les résultats observés auprès de la première cohorte et d'augmenter la puissance statistique nécessaire pour dégager des impacts plus subtils du programme sur les résultantes des enfants et des parents.

Le lecteur est prié de noter que ce rapport s'insère dans une série de rapports préparés par la Société de recherche sociale appliquée (SRSA). Il fait suite au rapport *Capacité d'apprentissage dans les communautés francophones en situation minoritaire : Rapport de référence* (Legault, Mák, Verstraete et Bérubé, 2014) dont la version finale a été remise à RHDCC le 13 octobre 2009. Ce premier rapport dressait le profil des enfants, de leur famille et des communautés participant au projet Capacité d'apprentissage. De plus, le présent rapport se veut complémentaire au rapport *Capacité d'apprentissage dans les communautés francophones en*

---

<sup>2</sup> La seule exception était pour le domaine du développement physique, pour lequel aucun effet du programme testé n'était attendu.

*situation minoritaire : Rapport de mise en œuvre du projet* (Bérubé, Legault, Janisse, Carson, Saucier et Lefebvre, 2014), soumis à RHDCC le 31 mai 2010. Trois autres rapports sont prévus, chacun portant respectivement sur : les analyses d'impact et l'étude de la mise en œuvre du programme préscolaire de service de garde auprès des deux cohortes de participants; les impacts à un an suivant la fin de l'intervention; et les impacts deux ans après la fin de l'intervention.

# **1. Présentation du projet Capacité d'apprentissage dans les communautés francophones en situation minoritaire (projet Capacité d'apprentissage)**

## **1.1 GENÈSE DU PROJET**

Le projet Capacité d'apprentissage dans les communautés francophones en situation minoritaire (projet Capacité d'apprentissage) fait partie du Plan d'action pour les langues officielles 2003–2008 du gouvernement du Canada et se poursuit sous la bannière de la Feuille de route pour la dualité linguistique canadienne de 2008–2013. Les principes directeurs qui orientent le projet comprennent la volonté de fournir un appui aux communautés francophones en situation minoritaire pour assurer aux enfants un bon départ dans la vie et d'encourager les parents à participer activement aux apprentissages de leur enfant (Ressources humaines et Développement des compétences Canada, 2006). Au niveau communautaire, le projet Capacité d'apprentissage se veut une évaluation rigoureuse d'une intervention prometteuse ayant comme objectif le maintien, voire le renouvellement de la vitalité ethno-linguistique de la communauté francophone minoritaire<sup>3</sup>.

Ce projet reconnaît que le développement des dimensions linguistique et identitaire prend une place plus importante en contexte linguistique minoritaire que dans un contexte majoritaire. Alors que l'appartenance à un groupe linguistique peut être prise pour acquis par les membres de la majorité linguistique, ce processus développemental atteint toute sa complexité en situation linguistique minoritaire. Être francophone se conçoit comme un processus continu qui englobe l'apprentissage de la langue française, la construction d'une identité et d'une culture francophones, et l'intégration à une communauté francophone.

Le développement des dimensions linguistique et identitaire est l'aboutissement du processus de socialisation vécu dans le milieu familial, le milieu scolaire ou préscolaire et le milieu socio-institutionnel (Landry & Allard, 1997). Les études pionnières quant à l'importance de la culture sur le développement de l'enfant nous viennent de Vygotsky (1978). La culture dans laquelle un enfant grandit a une influence sur le développement de ses habiletés langagières et sur ses apprentissages en général via l'intégration des symboles sociaux auxquels il est exposé. Ainsi, l'environnement social dans lequel baigne l'enfant est indissociable de la construction de son identité culturelle et linguistique, de même que de son développement global.

Dans la sphère publique, la réalité d'un contexte minoritaire fait en sorte que les petits francophones sont exposés à deux cultures différentes à un moment où leur identité se définit. Selon Gilbert (2003), l'exposition au français dans l'ensemble des contextes sociaux est d'autant plus importante pour l'enfant grandissant dans un milieu francophone très minoritaire où, par le simple poids démographique, l'anglais prédomine dans tous les aspects de la vie quotidienne. Des études empiriques démontrent que l'accès à des ressources et des services offerts en français seulement permet de contrebalancer la forte influence du poids démographique et social de l'anglais sur le vécu quotidien des membres de la communauté francophone. En outre, la

---

<sup>3</sup> Voir Guimond (2003) pour un survol des études portant sur la vitalité ethno-linguistique en milieu minoritaire.

présence de multiples milieux francophones favorise le maintien et l'épanouissement de l'identité ethnolinguistique et de la langue (Landry, Allard, & Deveau, 2007a). Au niveau de l'enfance, plusieurs auteurs préconisent la disponibilité de services de garde et de scolarisation en français comme principaux vecteurs de la vitalité communautaire (Commission nationale des parents francophones, 2005; Landry & Allard, 1997; Gilbert, 2003). L'idéal, selon les défenseurs de la vitalité de la communauté francophone, serait donc que les parents ayants droit inscrivent leurs enfants dans des services de garde de qualité et dans des écoles de langue française. La réalité est toutefois toute autre.

Les enfants de parents ayants droit sont souvent inscrits dans des écoles anglophones. Au Nouveau-Brunswick, par exemple, 83 % des enfants dont l'un des deux parents est francophone, ou dont les deux parents le sont, fréquentent l'école en français. En Ontario, seulement 51% des enfants dont au moins un des parents est de langue française fréquentent l'école en français (Corbeil, Grenier, & Lafrenière, 2007). Ce n'est le cas que de 26 % des jeunes franco-manitobains (Revue trimestrielle de l'éducation Canada, 2004). Les résultats de l'Enquête sur la vitalité des minorités de langue officielle de 2006 (EVMLO; Corbeil et al., 2007) mettent en lumière le fait que seulement 56 % des enfants de parents ayants droit fréquentent des écoles primaires de langue française. Cette proportion diminue à 44 % à l'adolescence. Ces données soulignent qu'une proportion non négligeable de parents ayants droit choisissent d'inscrire leurs enfants dans des écoles d'immersion ou dans des écoles de langue anglaise. Les parents citent plusieurs raisons sous-jacentes au choix d'un programme d'immersion notamment la proximité de l'école, la non-disponibilité d'une école de langue française et la qualité du programme ou de l'école. Les principales raisons sous-jacentes au choix d'un programme scolaire régulier de langue anglaise sont, en ordre décroissant : la proximité des écoles (27 %); le fait que l'anglais est la langue maternelle ou la langue la mieux connue de l'enfant (18 %) ou du parent (17 %); et la qualité de l'école ou du programme (17 %) (Corbeil et al., 2007).

Pour ce qui est des jeunes enfants francophones inscrits dans les écoles de langue française, plusieurs études démontrent que ces derniers obtiennent des résultats en littératie et en numératie inférieurs aux enfants des groupes linguistiques majoritaires. Cette disparité entre les deux groupes est observée, entre autres, dans les résultats des tests internationaux tels que le Programme international pour le suivi des acquis des élèves (PISA) où les enfants francophones inscrits à l'école de langue française en milieu linguistique minoritaire obtiennent des résultats inférieurs en lecture par rapport à leurs pairs anglophones canadiens (Bussière, et al., 2001; Conseil canadien sur l'apprentissage, 2008). Les quelques études portant sur les petits francophones minoritaires font ressortir que cet écart apparaît dès le jeune âge. Par exemple, des enseignants de maternelle ont classé la moitié des élèves échantillonnés comme ayant une compétence globale du français inférieure à la norme provinciale en utilisant la grille de rendement établie par le gouvernement de l'Ontario (Masny, 2006). Ces résultats trouvent un écho dans une étude plus récente auprès d'enfants franco-manitobains âgés de quatre à six ans. Ces derniers obtiennent des scores plus faibles aux tests de vocabulaire (Échelle de vocabulaire en images de Peabody, ou ÉVIP, et échelles de communication et de connaissances générales de l'Instrument de mesure du développement de la petite enfance, ou IMDPE). De façon non surprenante, les enfants grandissant dans un environnement quotidien majoritairement anglophone ont obtenu les scores les plus faibles. Cette tendance se poursuit lorsque ces enfants atteignent la troisième année du primaire. Les enfants ayant grandi dans un environnement familio-préscolaire francophone obtiennent de meilleurs résultats en lecture en troisième année



que les enfants francophones ayant vécu dans un environnement linguistique majoritairement anglophone (Chartier, Dumaine, Daudet-Mitchell, Gosselin, & Vielfaure, 2008).

Ces résultats doivent cependant être nuancés, puisque les dernières données montrent que les francophones hors Québec ont maintenant surpassé leurs camarades canadiens quant au taux d'obtention de diplômes postsecondaires (D'Amours, 2010). En Ontario, les jeunes franco-ontariens montrent une amélioration constante, au point où les enfants de troisième et de sixième année obtiennent maintenant des résultats comparables, voire supérieurs, aux enfants anglophones. Ces tendances suggèrent que l'introduction récente de certaines actions semble efficace pour améliorer la situation des francophones en milieu minoritaire (Office de la qualité et de la responsabilité en éducation, 2009). Toutefois, les communautés francophones se doivent de rester vigilantes afin de maintenir les acquis au niveau de la performance scolaire et d'étendre ces accomplissements aux milieux très minoritaires là où les défis sont les plus grands.

L'exposition restreinte de ces enfants à la langue française à la maison, limitant ainsi le développement de leurs habiletés langagières en français, serait vraisemblablement à l'origine de ces difficultés. De fait, près des deux tiers des petits francophones minoritaires sont issus de foyers exogames (67%) et la plupart adoptent l'anglais comme langue du foyer (Landry, 2010).<sup>4</sup> Seulement 20 % des couples exogames choisissent le français pour élever leurs enfants de 0 à 4 ans (Martel, 2001). En outre, les dernières données tirées du Recensement de 2006 indiquent que près de 39 % des francophones vivant hors Québec parlent plutôt l'anglais à la maison, bien que le français demeure une langue utilisée (Corbeil & Blaser, 2007).

Cette plus grande utilisation de la langue anglaise au quotidien explique en partie pourquoi 62 % des adultes francophones hors Québec passant un test de littératie en français (plutôt qu'anglais) démontrent un niveau de littératie inférieur à ce qui est jugé nécessaire pour bien fonctionner en société (c.-à-d. qu'ils obtiennent un niveau de littératie de 2 ou moins sur une échelle de 5; RHDCC et Statistique Canada, 2005, Tableau 3.24). Cette proportion serait sans aucun doute plus élevée si tous les adultes francophones hors Québec avaient fait le test en français (65% de ces derniers choisissent de remplir le test en anglais bien qu'ils identifient le français comme langue maternelle; RHDCC et Statistique Canada, 2005, p.54). Bien que cette statistique de 62% soit vraisemblablement une sous-estimation de l'ampleur réelle du problème, il n'en demeure pas moins que les adultes hors Québec testés en français ont un niveau de littératie plus faible que ceux testés en anglais (tout le Canada incluant le Québec) ou que ceux testés en français au Québec, qui ne démontrent pas le niveau de littératie fonctionnelle dans 44% et 55% des cas respectivement (RHDCC et Statistique Canada, 2005, Tableau 3.24). En d'autres mots, le taux de littératie fonctionnelle des francophones hors Québec est 7% plus bas que celui des personnes testées en français au Québec et il est 18% inférieur à celui des personnes testées en anglais (incluant le Québec). Selon ces résultats, il y a tout lieu de consolider l'acquisition de la langue auprès des populations en contexte minoritaire et d'émousser l'engagement des parents envers la préservation de la communauté francophone.

Pour les petits francophones minoritaires, l'ensemble de ces constats sont inquiétants lorsque l'on considère que les compétences langagières demeurent le *pivot central* de la réussite scolaire, qui elle influence la réussite professionnelle et sociale. Une intervention précoce auprès des

---

<sup>4</sup> Landry, R. (2010). Petite enfance et autonomie culturelle : Là où le nombre le justifie...V. Rapport de recherche réalisé pour la Commission nationale des parents francophones. Institut canadien de recherche sur les minorités linguistiques. Moncton, Nouveau-Brunswick.

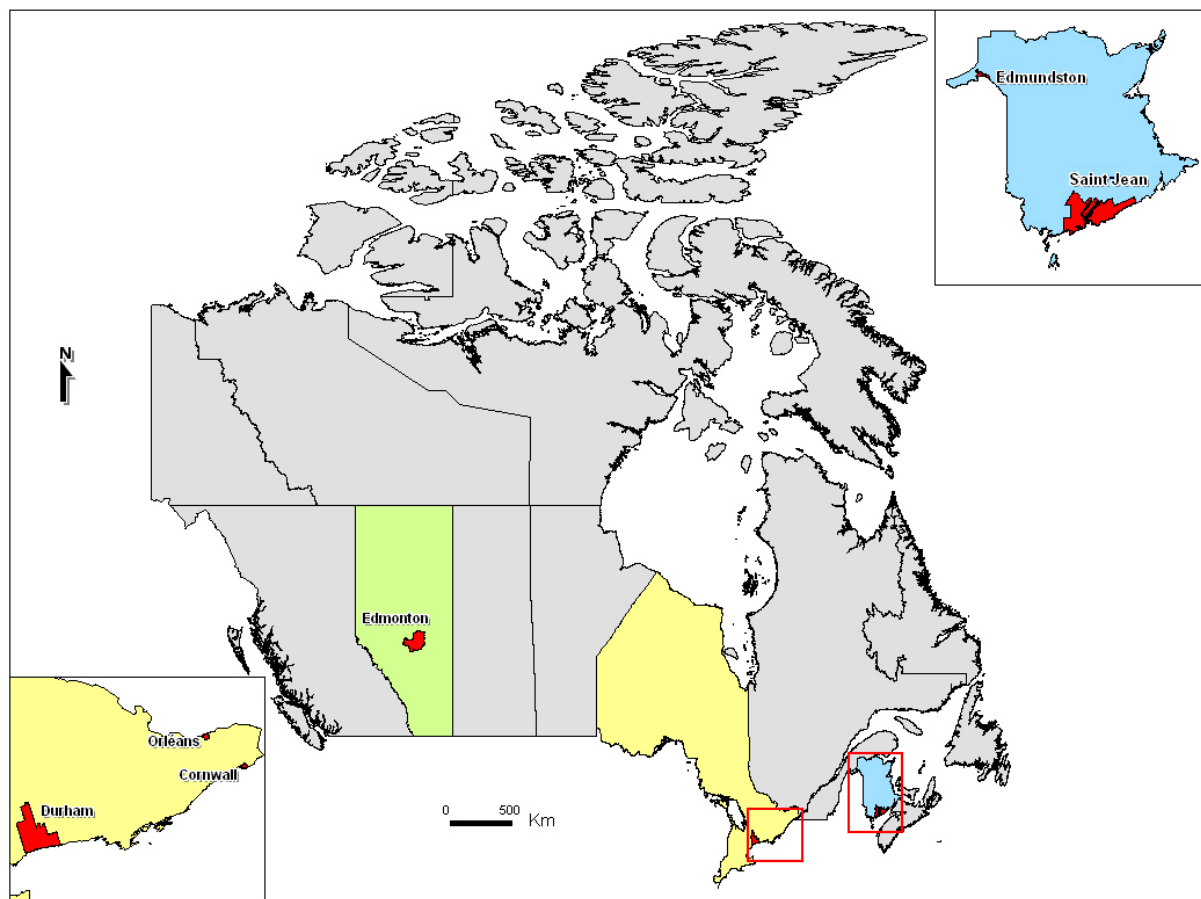
petits francophones minoritaires ciblant autant le milieu familial que préscolaire pourrait affecter positivement leur trajectoire linguistique et, par extension, favoriser leur succès scolaire et leur intégration au sein de la communauté francophone. Cette intervention devrait, entre autres, comprendre une composante ciblant les parents afin de les conscientiser aux défis de vivre en milieu minoritaire et aux actions qu'ils peuvent prendre pour transmettre ce riche héritage culturel à leurs enfants. Le projet Capacité d'apprentissage vise à évaluer l'impact d'une telle intervention.

## 1.2 LE PROJET CAPACITÉ D'APPRENTISSAGE

Le projet Capacité d'apprentissage met à l'essai un programme préscolaire de service de garde et vise à mettre en évidence les avantages qu'apporte ce programme préscolaire aux enfants vivant en milieu minoritaire francophone par rapport à d'autres enfants n'y étant pas exposés. L'effet du programme sur le développement des enfants est identifié en comparant un groupe de participants ayant été exposé au nouveau programme (connu sous le nom de *groupe programme*) à des groupes témoins comprenant des participants n'ayant *pas été* exposé au nouveau programme (nommés *groupe témoin en garderie* et *groupe témoin hors garderie*). Un comité consultatif de recherche, composé d'universitaires spécialisés en petite enfance francophone et de représentants des communautés francophones, a été créé afin d'aider à concevoir, mettre en œuvre, faire le suivi et évaluer le projet pilote. La Société de recherche sociale appliquée gère, met en œuvre et évalue le programme, en plus de participer activement au développement du projet.

La prestation du programme s'est faite pour deux cohortes de participants. La première cohorte a débuté le programme à l'automne 2007 dans six communautés francophones en milieu minoritaire (Saint-Jean et Edmundston au Nouveau-Brunswick; Orléans, Cornwall et Durham en Ontario; et Edmonton en Alberta). La prestation du programme à la deuxième cohorte a commencé à l'automne 2008 dans deux communautés (Orléans et Cornwall en Ontario) (la figure 1.1 permet de localiser les six communautés au niveau national). L'étude se déroule sur une période de quatre ans. Les enfants sont suivis dès l'âge de trois ans jusqu'à sept ans, alors qu'ils seront au début de la deuxième année du primaire. La durée de l'étude nous permet de suivre le développement des petits francophones minoritaires du préscolaire au début de la scolarisation. Les dernières collectes de données se feront à l'automne 2011 pour les familles de la première cohorte et à l'automne 2012 pour les familles de la deuxième cohorte.

Figure 1.1 Localisation des six communautés participantes en fonction des limites géographiques de Statistique Canada, 2006



Dans une première phase, le projet Capacité d'apprentissage vise d'abord à répondre à une première question de recherche à savoir : *Le nouveau programme préscolaire, comprenant un volet en garderie et un volet d'ateliers parents/enfants, a-t-il un impact significatif sur les habiletés langagières, l'identité culturelle francophone et la préparation à la scolarisation des enfants au-delà du développement qui prendrait place en l'absence de ce programme, et indépendamment d'autres facteurs externes qui pourraient entrer en jeu?* Des questions connexes sont également investiguées dont : Pour qui ce programme est-il le plus bénéfique? Ce programme est-il rentable? Le nouveau programme peut-il être répété? Dans une deuxième phase, le projet Capacité d'apprentissage aborde une nouvelle question de recherche: *Le nouveau programme préscolaire permet-il aux enfants francophones grandissant en milieu minoritaire d'être mieux outillés pour réussir dans les tâches essentielles à la réussite scolaire que sont la lecture et les mathématiques?* Alors que la première question de recherche se concentre sur la période préscolaire, la deuxième question de recherche se penche sur la période de scolarisation englobant la première et la deuxième année scolaire, alors que les enfants sont âgés de 6 et 7 ans.

### **1.3 L'APPORT DU PROJET**

L'apport de ce projet est qu'il tient compte de l'environnement particulier des communautés minoritaires francophones en ayant une composante de francisation, une composante d'identité culturelle et une composante de sensibilisation et d'implication des parents. Le projet s'articule autour des jeunes enfants francophones et s'inspire du modèle écologique de Bronfenbrenner (1979) dont l'une des prémisses est que les enfants sont influencés par l'ensemble des environnements dans lesquels ils baignent. Le programme testé vise à renforcer les liens entre les principaux milieux dans lesquels l'enfant grandit — les actions du milieu en garderie appuyant ceux du milieu familial et vice versa — de façon à optimiser les apprentissages de l'enfant tant au niveau de son développement général et de sa préparation à la scolarisation qu'au niveau de son développement langagier et culturel francophones.

Les multiples bienfaits des programmes qui modifient à la fois l'environnement de l'enfant à la garderie et à la maison ont été établis dans le cadre de plusieurs études portant sur d'autres populations dites vulnérables (voir la revue de littérature de Reese, Sparks, & Leyva, 2010; Engle, Black, Behrman et al., 2007). Tel que l'ont démontré Pelletier et Corter (2005), les bienfaits de ces programmes sont accrus lorsque le parent et l'éducateur adoptent les mêmes approches avec l'enfant. Ces auteurs ont évalué un programme de préparation à l'école impliquant des enfants de quatre ans, leurs parents et leurs éducatrices.<sup>5</sup> Les résultats de leur étude ont démontré qu'un programme à deux volets (parent et éducatrice), axé sur les activités de littératie précoce, favorisait davantage les acquisitions des enfants dans ce domaine, qu'un programme semblable comprenant seulement un des deux volets. Les enfants dont l'environnement à la garderie et celui à la maison avaient changé, étaient plus avancés sur le plan du développement du vocabulaire, de la pré-lecture et de la numératie. Les résultats d'autres études sont repris à la Section 1.5.4.

Si les bienfaits d'un programme impliquant les parents et les éducatrices semblent être relativement bien établis, il n'y va pas de même pour les retombées d'un programme ciblant à la fois les activités de littératie précoce et le développement d'une identité culturelle francophone chez l'enfant.

### **1.4 LE PROGRAMME TESTÉ**

Le programme préscolaire testé jumelle un volet de service de garde spécifiquement développé pour satisfaire aux besoins des enfants francophones en milieu minoritaire à un volet d'alphabétisation familiale ciblant les parents de ces enfants. Ce dernier volet cherche à stimuler la participation active des parents dans le développement de leur enfant et dans sa préparation à l'école, ainsi que la transmission de la langue et de la culture françaises. Le projet comme tel s'inscrit dans la foulée des nombreuses études portant sur le développement des enfants d'âge préscolaire et sur la vitalité de la langue française dans des milieux minoritaires. Les sections suivantes décrivent chacun des volets du programme testé et leur arrimage.

---

<sup>5</sup> Le féminin est privilégié en reconnaissance de la très forte proportion de femmes dans cette profession.

### **1.4.1 Le volet en garderie**

Le programme adopte une approche préscolaire, ou de préparation à l'école. Dans ce type de programme, les enfants sont amenés à atteindre des objectifs de développement précis directement liés à la préparation à l'école et ayant pour but de faciliter leur réussite scolaire. Cette approche s'oppose à une pédagogie dite « sociale » où le programme prévoit des orientations générales et où chaque milieu adopte des éléments en fonction des spécificités de leur communauté.

Par ailleurs, le programme en garderie préconise la pédagogie par le jeu. Selon cette approche :

« [...] les enfants sont vus comme des êtres indépendants qui peuvent contribuer activement à façonner leurs environnements d'apprentissage. L'objectif est d'enrichir et d'élargir les nouvelles occasions d'apprentissage, en se fondant sur les connaissances de l'éducatrice sur le développement de l'enfant, sur l'observation et la documentation des activités de l'enfant, ainsi que sur le contexte familial et communautaire de l'enfant. » (Bertrand, 2007, p. 4)

Le programme préscolaire offert en garderie s'inspire du Programme des prématernelles fransaskoises (pour les enfants de quatre ans) développé par la Direction de l'éducation française du ministère de l'Apprentissage de la Saskatchewan et il est adapté pour les enfants de trois ans. Les éducatrices sont appuyées, dans la prestation du programme, au moyen d'une formation de base et de formations de suivi régulières. Du matériel pédagogique et d'animation en français est également fourni aux éducatrices afin d'encourager chez les enfants le développement de leurs aptitudes linguistiques françaises, leur préparation à la scolarisation à l'école de langue française, leur connaissance de la culture francophone, et leur sentiment d'appartenance à la communauté francophone.

### **1.4.2 Le volet Ateliers familles**

À la différence de certains programmes en alphabétisation familiale, les Ateliers familles du projet Capacité d'apprentissage visent à outiller le parent afin que ce dernier puisse appuyer le développement de son enfant et non pour améliorer son propre niveau d'alphabétisme. Adoptant une approche andragogique, le programme des Ateliers familles compte plusieurs objectifs dont le plus important est de sensibiliser le parent à son rôle de premier éducateur de son enfant. Parmi les autres objectifs du programme, on retrouve : outiller les parents afin qu'ils puissent appuyer le développement langagier, culturel et identitaire de leur enfant en français, et ce, en contexte familial unilingue, bilingue, trilingue ou multiculturel; et conscientiser les parents au travail de l'éducatrice ainsi qu'à l'importance des rôles complémentaires parent-éducatrice pour soutenir les apprentissages de leur enfant.

Les ateliers sont des occasions d'échanger en français avec d'autres parents et des personnes-ressources. L'accent est mis sur le plaisir d'apprendre ensemble. Les activités sont conçues pour appuyer les divers vécus et périodes de socialisation des enfants, soit avec leurs parents, frères et sœurs, ami(e)s de la garderie ou autres membres de la communauté. Les parents reçoivent des conseils, des ressources (du matériel prêté aux parents et à l'enfant) et de la formation sur différents aspects du développement de leur enfant (p.ex., intelligences multiples, discipline, nutrition, etc.). Les participants sont appelés à échanger des informations sur les ressources et services offerts aux familles francophones de leur communauté.

Le volet Ateliers familles prend la forme d'une série de dix ateliers parents/enfants offerts sur une base hebdomadaire et animés par des intervenantes en alphabétisation. Il a été développé spécifiquement pour le projet Capacité d'apprentissage par Éduk. Il est fondé sur les pratiques exemplaires en littératie familiale émises par le « Centre for Family Literacy » (2002). Son contenu s'inspire des programmes reconnus dont :

- *Grandir avec mon enfant* (2002) et ses adaptations, dont *J'apprends en famille* de la Nouvelle Écosse, en particulier en ce qui concerne les activités en lien avec les compétences parentales et les besoins des enfants;
- *Chansons, contes et comptines*, de même que *Grandir avec des livres*, compte tenu de l'accent qui est mis sur l'éveil à l'écrit et la composante culturelle francophone;
- le programme anglophone *Learning Together* qui a fait l'objet d'une étude longitudinale en Alberta (de 2001 à 2005) et dont les résultats auprès des enfants et des familles sont bien documentés (Phillips, Hayden, & Norris, 2006); et
- le *Programme des prématernelles fransaskoises* (2001).

### **1.4.3 Arrimage entre le volet garderie et le volet Ateliers familles**

Pour maximiser l'impact du projet Capacité d'apprentissage pour les enfants, les deux volets principaux (le programme préscolaire et le programme d'Ateliers familles) sont arrimés. Les deux volets partagent les objectifs d'optimisation du développement des aptitudes linguistiques en français chez les enfants, leur préparation à la scolarisation à l'école de langue française, le développement de leur connaissance de la culture francophone et celui de leur sentiment d'appartenance à la communauté francophone. Le programme préscolaire met l'accent sur la francisation et les activités de pré-littératie chez l'enfant d'âge préscolaire (en incluant des sous-thèmes de lecture, d'écriture et de numératie). Les Ateliers familles complètent le programme préscolaire par des activités qui informent les parents sur une variété de sujets tels que leur rôle comme premier éducateur de leur enfant (appui au développement de l'enfant, la stimulation des apprentissages), les ressources communautaires disponibles en français et les méthodes pour transmettre la langue et la culture à leur enfant.

Pour lier ces volets, les conceptrices des Ateliers familles ont travaillé en étroite collaboration avec la consultante en petite enfance spécialiste du programme en garderie dans le but d'assurer une cohérence entre les différents aspects du projet Capacité d'apprentissage, et une continuité au niveau des apprentissages de l'enfant et de son parent. Les conceptrices ont tenu compte :

- de l'approche et des valeurs qui sous-tendent le programme mis en place dans la garderie;
- des stratégies de francisation à mettre en pratique auprès de l'enfant;
- des thèmes traités à chaque mois dans la garderie;
- de la liste des ressources — jeux, jouets, et livres achetés pour le volet en garderie (les ressources proposées pour les Ateliers familles complètent celles du volet en garderie); et
- du stade développemental des enfants d'âge préscolaire.

Les deux volets sont offerts dans les « garderies programme » (c.-à-d. les garderies livrant le programme du projet Capacité d'apprentissage). Leur arrimage favorise un partenariat entre les différentes personnes intervenant auprès de l'enfant (éducatrice, parent et autres) et il renforce les apprentissages de l'enfant vécus dans les différentes sphères de sa vie telles que la garderie, le foyer et la communauté. L'importance d'un lien entre le milieu préscolaire et la famille découle des résultats d'études soutenant l'idée selon laquelle ce lien contribue aux habiletés de pré-littératie des enfants. Les parents qui s'impliquent en discutant avec l'éducatrice, en lui posant des questions sur le déroulement de la journée de leur enfant et en participant aux activités en garderie ont des enfants qui démontrent un vocabulaire plus étendu, une meilleure conscience phonologique et de meilleures habiletés de pré-écriture (Arnold, Zeljo, Doctoroff, & Ortiz, 2008).

#### **1.4.4 Déroulement du programme durant la première année du projet Capacité d'apprentissage**

La première année du programme regroupe les mois de septembre 2007 à août 2008, période au cours de laquelle la prestation du programme testé a pris place pour les participants de la première cohorte. Le programme du volet en garderie a été inauguré le 1<sup>er</sup> septembre 2007 dans les communautés de Cornwall, Edmonton, Edmundston et Saint-Jean. Il a été lancé en octobre 2007 dans les communautés de Durham et d'Orléans. Toutes les éducatrices impliquées dans la prestation du programme en garderie ont suivi une formation au cours des mois d'avril ou de mai 2007. Par la suite, des formations de suivi, ayant lieu aux deux mois, ont été offertes dans les six communautés, débutant en août 2007 et se terminant en juin 2008. Lors de chaque suivi, la formatrice observait les activités de la classe, offrait une rétroaction aux éducatrices et donnait des formations ciblées. Parmi les thèmes abordés durant les suivis, on note la planification thématique, l'art créatif, les portfolios des enfants, le développement d'un Centre de littératie, l'infusion des activités de littératie et de numératie dans les activités journalières et le développement des capacités de dessiner de l'enfant. Notons que le programme en garderie n'était pas fidèlement appliqué lors de la saison estivale, et ce, autant dans les garderies livrant le programme testé que les garderies témoins livrant un autre programme.

Seules les familles du groupe programme étaient appelées à participer au volet d'alphabétisation familiale. Les ateliers étaient présentés en deux séries — une première série de quatre ateliers en novembre 2007 et une deuxième série de six ateliers aux mois de janvier et février 2008. À la fin de chaque atelier, les familles recevaient une trousse-famille. Chaque trousse comprenait un livre pour le parent, et une ressource audio-visuelle ou un jeu. Les familles rapportaient la trousse à l'atelier suivant, pour en recevoir une nouvelle. La trousse contenait également les matériaux et la démarche écrite pour faire une activité créatrice à la maison avec l'enfant. De plus, les familles étaient invitées à visiter le Centre de ressources établi spécifiquement pour les ateliers et à choisir parmi les 300 ressources de langue française à leur disposition, incluant des livres pour parents. À partir du mois de mars 2008, une fois les Ateliers familles terminées, les coordonnatrices communautaires du projet Capacité d'apprentissage ont visité les enfants du groupe programme à la garderie une fois par semaine pour leur permettre d'emprunter un livre au Centre de ressources. Une fois par semaine également, la coordonnatrice se présentait à la garderie programme en fin de journée pour permettre au parent de visiter le centre avec son enfant et de choisir parmi la gamme complète des ressources incluant les livres pour parents, les disques compacts, les DVD et les jeux pour la famille.

### **1.4.5 Déroulement du programme durant la deuxième année du projet Capacité d'apprentissage**

La deuxième année du programme regroupe les mois de septembre 2008 à août 2009. La prestation du programme en garderie a eu lieu dans quatre des six communautés. Les enfants des communautés d'Edmonton, d'Edmundston et de Saint-Jean ont été exposés au programme en garderie à temps plein alors que les enfants de la communauté de Cornwall y ont été exposés à temps partiel. Les enfants des communautés de Durham et d'Orléans n'ont pas été exposés au programme en garderie durant la deuxième année. Dans ces communautés, les parents ont inscrit leur enfant à la maternelle de quatre ans qui était offerte gratuitement à temps plein.

Dès septembre ou octobre 2008, la formatrice a recommencé les formations de suivi dans les communautés de Cornwall, d'Edmundston et de Saint-Jean pour offrir aux éducatrices son appui à la mise en œuvre du nouveau programme en garderie. Les nouvelles éducatrices impliquées dans la prestation du programme en garderie ont suivi une mini-formation de base en septembre ou octobre 2008. La formatrice a donné sa dernière formation de suivi en juin 2009. Comme pour la première année, nous avons observé que le programme testé n'était pas fidèlement appliqué lors de la saison estivale, et ce, autant dans les garderies livrant le programme du PGGE que les garderies témoins livrant un autre programme. Précisons qu'il y a eu un échec à la mise en œuvre du programme en garderie dans la communauté d'Edmonton.<sup>6</sup> La dernière visite de la formatrice dans cette communauté a eu lieu en septembre 2008.

Aucune nouvelle série d'Ateliers familles n'a été offerte lors de la deuxième année du projet Capacité d'apprentissage. Toutefois, les activités du Centre de ressources se sont poursuivies au cours de la deuxième année dans chacune des six communautés du projet, que les enfants soient inscrits ou non à l'école. Les Centres de ressources étaient également ouverts une fois par semaine pour permettre aux enfants et aux parents de faire l'emprunt de ressources francophones variées.

En somme, l'ensemble du programme préscolaire s'appuie sur une perspective intégrative et tente, par différentes approches, d'influer sur les principaux vecteurs qui façonnent le développement de l'enfant. Les sections qui suivent présentent d'abord les différentes sources d'influence du développement des tout-petits en tenant compte des particularités de grandir en milieu minoritaire francophone pour ensuite se concentrer sur une définition de la préparation à scolarisation. Le chapitre se termine avec une section détaillant les objectifs et le contenu du présent rapport.

## **1.5 LE MODELAGE DU DÉVELOPPEMENT DES ENFANTS EN MILIEU MINORITAIRE**

Le projet Capacité d'apprentissage est basé sur une vision écologique selon laquelle des facteurs à divers niveaux influencent le développement de l'enfant. Cette vision est clairement représentée dans le modèle écologique de Bronfenbrenner (1979). Ce chercheur fut le premier à mettre en mots et en images tout le système d'influences qui agit sur le développement de l'enfant. Son modèle s'articule autour de trois prémisses:

---

<sup>6</sup> L'analyse de cet échec se retrouve dans le rapport *Capacité d'apprentissage dans les communautés francophones en situation minoritaire : Rapport de mise en œuvre du projet* (Bérubé, Legault, Janisse, Carson, Saucier & Lefebvre, 2014).



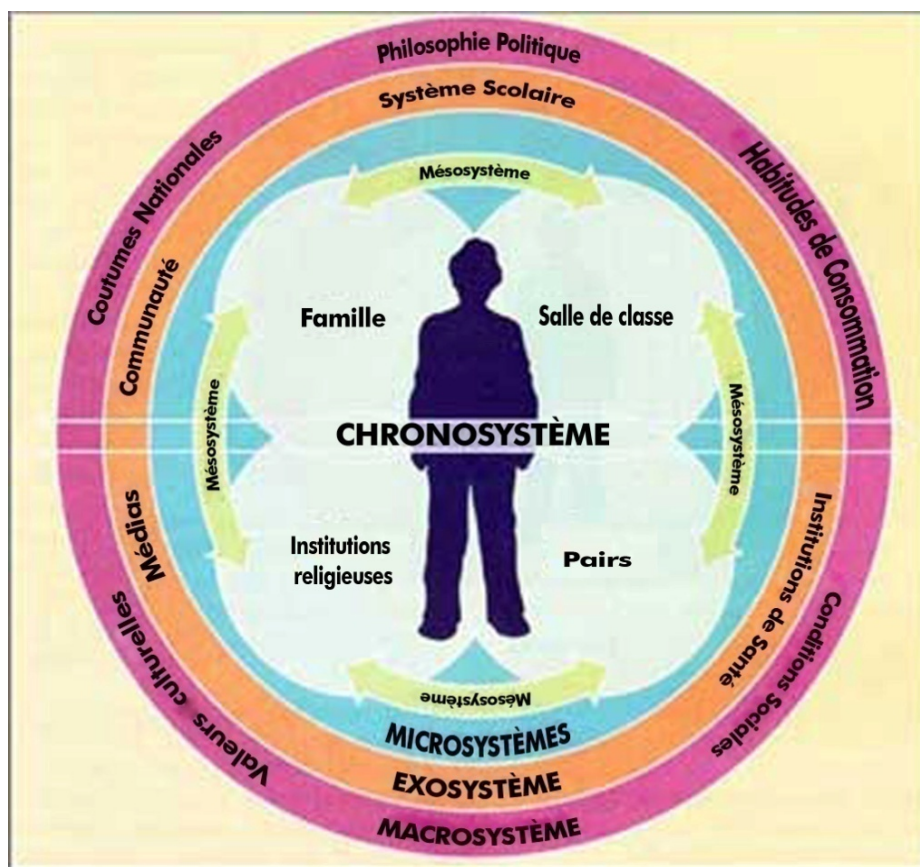
- a) l'enfant est au centre du modèle;
- b) la place centrale qu'occupent les expériences vécues par l'enfant (ces expériences sont considérées comme des « moteurs » du développement); et
- c) la nature des relations entre les différents milieux de l'enfant.

Le modèle de Bronfenbrenner comprend cinq systèmes (figure 1.2) :

- **Microsystème** : environnement immédiat (famille, école, mode de garde, pairs, voisinage);
- **Mésosystème** : interactions entre les environnements immédiats (p. ex., entre la maison et l'école);
- **Exosystème** : environnement externe qui affecte indirectement l'enfant (tel que le travail des parents);
- **Macrosystème** : contexte culturel plus large (culture occidentale par rapport à culture orientale, économie nationale, culture politique, sous-culture);
- **Chronosystème** : structure des événements affectant l'environnement et les transitions au cours de l'existence.

Dans le contexte du projet Capacité d'apprentissage où le centre d'intérêt est les jeunes enfants en milieu linguistique minoritaire, trois systèmes du modèle de Bronfenbrenner prennent une place particulièrement importante. Tout d'abord, le microsystème, via les caractéristiques familiales et les caractéristiques du milieu de garde de l'enfant, influe sur le développement des enfants. Au niveau des caractéristiques familiales, nous distinguons les variables contextuelles des processus familiaux (cette division se base sur l'Enquête longitudinale nationale sur les enfants et les jeunes, ou ELNEJ; Statistique Canada, 2006). Les variables contextuelles font référence à des données « factuelles » reconnues comme étant importantes pour le développement des enfants (p. ex., la composition de la famille). Ensuite le mésosystème, tel que les liens qu'entretiennent la famille et le milieu de garde, joue aussi un rôle dans le développement des tout-petits. Enfin, le macrosystème, soit la communauté dans laquelle l'enfant grandit, et tout particulièrement ses caractéristiques linguistiques, fait aussi partie des influences d'intérêt pour le projet.

Figure 1.2 Modèle écologique complet de Bronfenbrenner (1979)



Source : Dans UW-Extension ABC Project, Annexe B (novembre 2004), traduction libre.

### 1.5.1 Le microsystème : environnement familial (variables contextuelles)

Le développement des enfants subit l'influence de plusieurs facteurs intrinsèques à l'enfant comme son historique fœtal, son poids à la naissance ou la durée de sa gestation. Outre les caractéristiques inhérentes à l'enfant, les recherches ont identifié au cours des années une série de facteurs environnementaux pouvant affecter son développement. Les variables contextuelles de l'environnement familial font partie des facteurs contribuant le plus au développement de l'enfant (Sanders & Morawska, 2006). Ces dernières comprennent la composition de la famille, le revenu, le niveau de scolarité des parents, ainsi que les langues utilisées à la maison.

#### **Composition de la famille**

Dans le contexte du projet Capacité d'apprentissage, nous nous intéressons aux variables familiales influençant le développement de l'enfant, plus particulièrement à celles qui influencent son parcours langagier. Ainsi, le rang de naissance s'avère être une variable d'importance. En effet, les études démontrent que les aînés des familles ont un vocabulaire plus étendu en moyenne que leur fratrie (Tamis-LeMonda & Rodriguez, 2008).

Le type de famille (biparentale intacte, monoparentale, recomposée, etc.) est également une variable à considérer. Une étude effectuée avec les données longitudinales de l'ELNEJ a montré

que le type de famille (biparentale ou monoparentale) avait une influence concrète sur plusieurs aspects du développement de l'enfant. En effet, pour les enfants de six à onze ans, les auteurs concluent que « les enfants des familles biparentales étaient moins hyperactifs, avaient de meilleures aptitudes scolaires, étaient moins anxieux ou déprimés et ont reçu de leurs enseignants une bonne évaluation quant à leur niveau de connaissances » (Adams & Ryan, 2000, p. iii). Cependant, ce n'est pas tant la composition de la famille que les facteurs associés à la monoparentalité, comme le stress maternel ou la diminution du revenu familial, qui affectent l'enfant *a priori*, d'où l'importance d'interpréter les résultats avec précaution.

### **Revenu familial brut**

Des centaines d'études ont déjà démontré une association entre la pauvreté familiale et la santé de l'enfant, sa préparation à l'école, sa réussite scolaire et son comportement. Toutefois, peu de ces études se sont penchées sur l'effet du moment, de la durée et de l'intensité de la pauvreté. Un revenu familial insuffisant affecte le développement de l'enfant via de nombreux canaux : nutrition inadéquate, diminution des situations d'apprentissage, instabilité du lieu de résidence, inscription à des écoles ayant moins de ressources, violence familiale, etc. (voir Duncan & Brooks-Gunn, 1997). Certaines études, comme celle de Berger et de ses collègues (2005), ont confirmé l'hypothèse qu'un faible revenu influence le développement de l'enfant par le biais de ses effets multiples sur l'environnement à la maison. Cependant, bien que la relation causale entre un faible revenu et les résultantes développementales chez les enfants (sur le plan intellectuel et comportemental) soit bien claire, l'interprétation de ces associations reste discutable, de même que les implications en termes de politiques.<sup>7</sup> Les études récentes sur le sujet tendent à démontrer deux « routes » principales par lesquelles un faible revenu affecte potentiellement l'enfant : l'environnement physique et la qualité des pratiques parentales (Berger, Paxson & Waldfogel, 2005).<sup>8</sup> Autrement dit, la pauvreté affectera l'achat de ressources matérielles destinées à la famille, en plus d'affecter le niveau de stress familial, lesquels à leur tour influencent le développement de l'enfant. Plusieurs études basées sur des données nationales ont démontré que le niveau de stimulation cognitive amené par l'environnement à la maison (mesuré par le matériel d'apprentissage et les pratiques parentales relatives à l'apprentissage) explique de 33 à 50 % l'association entre le revenu et les différentes résultantes sur le développement cognitif et langagier de l'enfant (Dearing, Berry & Zaslow, 2006).

Les études tendent aussi à démontrer que les effets néfastes de la pauvreté sur le développement cognitif et la réussite scolaire de l'enfant sont plus grands lors de la période préscolaire qu'à toute autre période (Dearing et al, 2006). La prise en compte de cette information devient donc incontournable pour évaluer l'effet du programme dans le cadre du projet Capacité d'apprentissage.

### **Niveau de scolarité des parents**

Tout comme le revenu familial, le niveau de scolarité des parents est un facteur important de la réussite d'un enfant (voir, entre autres, Haveman & Wolfe, 1995). Toutefois, le mécanisme par lequel l'éducation des parents influence le développement de l'enfant a été moins bien étudié.

---

<sup>7</sup> Par exemple, les transferts monétaires directs vers les familles seraient adéquats si une relation causale claire existait entre le revenu, la qualité de l'environnement et les résultantes des enfants (Berger et al., 2005).

<sup>8</sup> La première théorie a été développée initialement dans la littérature économique (voir Becker, 1993) et la seconde dans la littérature en psychologie du développement (p.ex. Dearing et al. (2006)).

Klebanov et ses collègues (1994) ont démontré que l'éducation de la mère et le revenu familial sont d'importants facteurs reliés à l'aménagement d'un environnement physique propice à l'apprentissage, mais que seule l'éducation s'avère un facteur d'importance dans les pratiques parentales « chaleureuses ». Un ensemble d'études de Davis-Kean (2005) conclut que l'éducation des parents influence le développement de l'enfant non seulement via la réussite sociale des parents, mais aussi via leurs croyances et leurs comportements envers leur enfant. On peut donc voir que le niveau d'éducation est important pour le développement des enfants, au-delà du contexte socio-économique auquel il est souvent associé.

### ***Âge de la mère à la naissance du premier enfant***

L'âge de la mère à la naissance du premier enfant est un autre facteur qui a de multiples effets sur le développement des enfants. Les études montrent que l'âge de la mère à la naissance de son premier enfant a un impact sur le développement de l'enfant et des enfants subséquents, même une fois les pratiques parentales et le fonctionnement de la famille pris en compte (Tremblay, et al., 2004). Dans le cas qui nous préoccupe, retenons que les très jeunes mères offrent une stimulation langagière moins complexe à leur enfant. Elles utilisent un éventail de vocabulaire moins riche et elles stimulent moins verbalement leurs enfants (Tamis-LeMonda & Rodriguez, 2008). Ce comportement se répercute chez l'enfant qui est alors à risque de présenter un développement langagier insuffisant pour lui assurer une entrée à l'école réussie.

### ***Langue(s) parlée(s) à la maison***

En contexte minoritaire, les langues les plus souvent parlées à la maison par les parents de l'enfant cible sont reliées à la transmission de la langue et à la vitalité du français à la maison (Forgues & Landry, 2006). Ainsi, l'environnement linguistique familial est associé au niveau de préparation scolaire et à la réussite scolaire de l'enfant (Chartier et al., 2008). Afin de trouver cette information, il faut savoir quelles sont les langues habituellement parlées dans l'environnement de l'enfant, en particulier par ses parents, ses frères et sœurs aînés à la maison et à l'extérieur de la maison. Il est aussi intéressant de connaître les langues habituellement parlées par l'enfant. Ce comportement linguistique est une indication concrète des langues avec lesquelles l'enfant se sent plus à l'aise de s'exprimer. L'ensemble de ces informations permettra d'établir un portrait global des influences linguistiques auxquelles l'enfant est exposé.

## **1.5.2 Le microsystème : environnement familial (processus familiaux)**

Outre les caractéristiques descriptives des familles, les processus familiaux ont une influence certaine sur le développement des enfants. Parmi les principaux éléments affectant le développement de l'enfant, notons le fonctionnement familial, les pratiques parentales et les activités d'alphabétisation de la famille.

### ***Fonctionnement familial***

Au-delà des difficultés psychosociales de la mère, l'environnement global dans le ménage est aussi un élément influençant le développement de l'enfant. Le fonctionnement de la famille se définit comme la qualité des liens dans la famille, tant au niveau de la qualité de la communication, de l'entente entre les membres que du soutien disponible au sein de la famille. D'un point de vue social, un fonctionnement familial déficient est largement associé à des trajectoires de développement agressives chez les enfants (Tremblay et al., 2004). Du côté du

développement langagier, l'acquisition du vocabulaire chez l'enfant est positivement liée au niveau de fonctionnement de sa famille, soit le niveau d'entente et de communication entre les membres (Desrosiers & Ducharme, 2006). Le fonctionnement familial a donc des répercussions importantes et multiples sur le parcours de vie des enfants.

### **Pratiques parentales**

En tant que premiers éducateurs de l'enfant, les parents jouent un rôle clé dans le développement et le fonctionnement de leur enfant. Le style parental aura une influence sur le développement social, intellectuel, moral et affectif de l'enfant (Bornstein & Bornstein, 2007). Les styles parentaux sont composés de deux dimensions : la *sensibilité*, qui mesure à quel point le parent est à l'écoute de l'enfant et est capable de répondre à ses besoins et intérêts; et le *contrôle* (ou l'*exigence*) qui réfère au degré de supervision et de discipline, ainsi qu'à quel point le parent exige de son enfant qu'il fasse preuve d'obéissance et de maîtrise de soi (Conseil canadien sur l'apprentissage, 2007). Les études sur le sujet indiquent que les enfants démontrent de meilleures habiletés langagières et ont des scores plus élevés aux tests d'intelligence lorsque leurs parents sont plus encourageants et moins contrôlants (Sanders & Morawska, 2006). Il est donc important de mesurer ces deux dimensions lorsque vient le temps de suivre le développement langagier et cognitif de l'enfant.

### **Activités d'alphabétisation**

Les expériences vécues durant la petite enfance se répercutent sur les habiletés langagières de l'enfant dès l'entrée à l'école (Doherty, 1997). Selon Desrosiers et Ducharme (2006), les enfants à qui les parents ont fait la lecture régulièrement sont moins susceptibles de présenter un retard sur le plan du vocabulaire. La lecture à la maison en bas âge (avant trois ans et demi) serait même associée à une amélioration des capacités verbales chez les enfants qui montraient des difficultés langagières. De la même façon, la participation des parents à des activités d'apprentissage à la maison permet de prédire l'adaptation sociale et scolaire des enfants à long terme (Izzo, Weissberg, Kasprow, & Fendrich, 1999). D'intérêt particulier pour le projet Capacité d'apprentissage, l'exposition au français oral et écrit durant le préscolaire est cruciale pour que l'enfant développe une forte compétence linguistique. Les activités telles que l'heure du conte et l'emprunt de livres à la bibliothèque permettent aux éléments uniques de la culture francophone d'être transmis aux enfants (Salerno dans Lafrance, 1993).

### **1.5.3 Le microsystème : milieu de garde**

Les caractéristiques du milieu de garde de l'enfant influent sur de nombreuses facettes du développement de l'enfant. Depuis plusieurs années, psychologues et éducateurs ont conclu, sur la base d'observations et d'expériences, que les soins non parentaux affectent aussi bien le développement cognitif que langagier de l'enfant (*National Institute of Child Health and Human Development Early Child Care Research Network & Duncan*, 2003). De façon plus précise, la qualité d'un milieu de garde se répercute sur le développement cognitif et langagier des enfants, sur leur préparation à l'école, ainsi que sur leurs comportements (Cleveland et al., 2006). Par ailleurs, cet impact varie grandement selon certains facteurs familiaux. Par exemple, les effets d'un milieu de garde de qualité sont plus importants auprès des enfants grandissant dans un contexte socio-économique défavorisé (Burchinal et al., 2000) ou dans un milieu familial utilisant peu le français comme c'est le cas en contexte de minorité linguistique (Maltais, 2007).

La garderie peut être un lieu d'apprentissage menant à de meilleures habiletés scolaires. Par exemple, dans des études réalisées aux États-Unis, le fait d'avoir fréquenté un service de garde a été associé à des scores plus élevés en lecture et en mathématiques à l'entrée à la maternelle à cinq ans (voir, entre autres, Howes et al., 2008). Un matériel adéquat comprenant des jeux et des livres de qualité, un milieu physique approprié et des éducatrices affectueuses soutenant le développement des enfants peuvent s'avérer très bénéfiques pour tous les enfants, et ce, particulièrement pour des enfants vivant en situation familiale difficile. La qualité dans les services de garde peut prendre deux formes : la qualité structurelle, qui prend en compte les facteurs pouvant être modifiés par le biais de la législation (diplôme de l'éducatrice, heures d'ouverture, taille du groupe, etc.), et la qualité des processus qui réfère aux expériences de l'enfant par rapport à son environnement en garderie (qualité des activités offertes et du lien avec l'éducatrice) (Burchinal et al., 2000). Dans le cadre du projet Capacité d'apprentissage, ces informations sont prises en compte par le biais d'observations. Ces informations permettront de comparer le programme testé offert dans le volet garderie aux activités offertes dans les garderies témoins. C'est à ce niveau que nous observerons si le programme testé a véritablement changé l'environnement de la garderie.

Dans un contexte minoritaire francophone, l'aspect linguistique prend une place particulièrement importante dans l'étude de l'impact d'un programme en garderie. En effet, le fait d'exposer l'enfant à un milieu de garde en français au préscolaire a un impact positif sur sa réussite à l'école. Chartier et ses collègues (2008) ont démontré cette affirmation dans une étude utilisant des données longitudinales portant sur 217 enfants de la communauté francophone du Manitoba.<sup>9</sup> Les enfants exposés au français dans leur environnement familial et leur milieu de garde ont obtenu des scores supérieurs à l'ÉVIP-R (qui mesure l'étendue du vocabulaire réceptif) et aux tests de communication et de connaissances générales de l'IMDPE comparativement aux enfants ayant été exposés au français à la maison seulement.

#### **1.5.4 Le mésosystème : liens entre la famille et le milieu de garde**

Le mésosystème est composé des liens entre les différents systèmes que l'enfant fréquente. Ces liens contribuent de façon unique au développement de l'enfant. Plusieurs études appuient l'idée selon laquelle le lien entre le milieu scolaire et familial contribue au développement de l'enfant. En effet, les enfants dont les parents s'impliquent dans le cheminement scolaire présentent une meilleure adaptation sociale et scolaire que les autres enfants, en plus de posséder des attitudes plus positives envers l'école et de démontrer de meilleures aspirations quant à leur futur, et ce, indépendamment du revenu familial et du niveau d'éducation des parents (Connors & Epstein, 1995). Le même parallèle a été établi entre l'implication des parents en milieu préscolaire et les habiletés de pré-littératie des enfants. Les parents qui s'impliquent en discutant avec l'éducatrice, en lui posant des questions sur le déroulement de la journée de leur enfant et en participant aux activités en garderie, ont des enfants qui démontrent un vocabulaire plus étendu, une meilleure conscience phonologique et de meilleures habiletés de pré-écriture (Arnold, et al., 2008). La valeur ajoutée d'un programme combinant les services de garderie et des Ateliers familles est la mise en relief de l'importance d'une bonne collaboration entre les parents et les éducatrices et de l'adoption d'approches communes et complémentaires (Corter & Pelletier, 2005).

---

<sup>9</sup> L'étude se nomme Parlons petite enfance, appelée aussi Étude manitobaine de la cohorte des naissances de 1997.

### 1.5.5 Le macrosystème : variables communautaires

De plus en plus de chercheurs reconnaissent l'influence importante des caractéristiques de la communauté sur le développement de l'enfant (Hertzman & Kohen, 2003; Moore, 2005). L'environnement communautaire favorise autant le développement cognitif des enfants que leur état de santé physique et émotionnel (Willms, 2007). Un des facteurs clés influençant le développement des enfants est la disponibilité des ressources et des services offerts seulement en langue française aux familles.

#### ***Disponibilité et utilisation des ressources***

La disponibilité et l'utilisation des ressources communautaires ciblant les jeunes familles d'une communauté s'avèrent des facteurs importants pour le développement des enfants. Selon Connor et Brink (1999), certaines catégories de ressources communautaires sont particulièrement importantes pour le développement des enfants, notamment le système d'éducation et de santé, les divertissements et la culture, les programmes sociétaux, les programmes axés sur les besoins spéciaux, et les sports et les loisirs. Dans le contexte de la minorité francophone, les services et ressources en petite enfance offerts en français se définissent davantage comme un élément protecteur de l'identité francophone et du maintien de la langue française (Commission nationale des parents francophones, 2005). Cette notion d'élément protecteur s'inspire du construit d'*infrastructure institutionnelle complète* proposée par Breton (1964). À son extrême, une communauté ayant une infrastructure institutionnelle complète offrirait à sa population francophone la possibilité de mener toutes ses activités quotidiennes en français. Il s'ensuit que la présence d'institutions de langue française au sein d'une communauté favorise la création de réseaux sociaux et l'accroissement de la cohésion sociale au sein de la communauté. Landry (1994) a repris ce concept d'infrastructure institutionnelle complète dans ses écrits, qui portent sur les quatre formes de capital — démographique, politique, économique et culturel — ayant une énorme influence sur le développement, le maintien (voire même le renouvellement) de la vitalité ethnolinguistique<sup>10</sup> d'une communauté. C'est par l'entremise de cette vitalité ethnolinguistique que les communautés minoritaires préservent leur sentiment d'appartenance et de fierté, appuyant ainsi l'intégration, et non l'assimilation, de la langue et de la culture francophones au sein de la communauté majoritaire. L'une de ces formes de capital, le capital culturel, « ...se réfère aux ressources et à l'information qui agissent comme agents de transmission de la culture » (Landry, 1994, p. 18, cité dans Guimond, 2003). L'auteur avance le postulat qu'une évaluation de ce capital culturel se voit possible en partie par la diversité des institutions éducatives et l'accès aux ressources culturelles dans la communauté. Landry, Allard, et Deveau (2007b) avancent que les écoles francophones — et particulièrement la présence d'un système scolaire permettant aux jeunes francophones de poursuivre leurs études du préscolaire au postsecondaire en français — ont un rôle crucial à jouer dans l'élaboration, le maintien et la croissance d'une infrastructure institutionnelle complète. Au palier préscolaire, se sont les garderies, les prématernelles, les services de garde en milieu familial, les programmes parascolaires, les centres de ressources et les groupes de jeu qui agissent à titre de porte d'entrée à l'école en français (Gilbert, 2003).

---

<sup>10</sup> Voir Guimond (2003) pour un survol des études portant sur la vitalité ethnolinguistique en milieu minoritaire.

## 1.6 UNE DÉFINITION DE LA PRÉPARATION À LA SCOLARISATION

Comme tous les enfants d'âge scolaire, les petits francophones en milieu minoritaire doivent se préparer à leur entrée à l'école. Cette transition est importante, comme l'indiquent les études dans le domaine. En effet, il existe une forte association entre le niveau de préparation des enfants à leur entrée à l'école et leur réussite scolaire (Lemelin & Boivin, 2007). Certains indicateurs permettent de prédire les dispositions d'un enfant à apprendre à l'école, et ce, dès l'âge de trois ans (Thomas, 2006).

Le terme « préparation à l'école » (« school readiness ») est un concept multidimensionnel, mais qui, selon Doherty (1997, p. 25), réfère surtout à la capacité de l'enfant à faire face aux tâches typiques demandées à l'école, telles que de rester assis et d'assimiler la matière. Il englobe aussi les compétences devant être mises en place de la naissance à l'âge de six ans afin d'assurer non seulement la réussite scolaire mais aussi « le succès dans tous les aspects de la vie adulte, notamment sur le marché du travail... ». Cinq aspects de la préparation de l'enfant à l'école sont communs à un ensemble de recherches (Meisels & Atkins-Burnett, 2006) :

- 1) le bien-être physique et le développement moteur approprié;
- 2) la santé émotive et une approche positive face à des expériences nouvelles;
- 3) la connaissance et la compétence sociale appropriées selon l'âge;
- 4) les habiletés linguistiques appropriées selon l'âge; et
- 5) les connaissances générales et aptitudes cognitives appropriées selon l'âge.

Bien que les facteurs 1, 4 et 5 soient les aspects qui sont généralement associés à la réussite de tâches précises nécessaires au succès scolaire, les points 2 et 3 permettent aussi de prédire ce succès. En effet, la gestion des émotions, l'attitude générale de l'enfant en classe (l'acte de rester assis toute la journée ou la curiosité face aux apprentissages, par exemple) et les compétences sociales de ce dernier sont autant d'éléments essentiels à la réussite scolaire. Le *National Education Goals Panel* (Meisels & Atkins-Burnett, 2006) reconnaît que ces cinq facteurs ne peuvent être dissociés de la famille, de l'école et de la communauté, et que la préparation à la scolarisation doit être mise en relation avec ces divers niveaux d'influence.

## 1.7 LE PRÉSENT RAPPORT

Le présent rapport porte sur les résultats d'analyses visant à établir les impacts à court terme du nouveau programme préscolaire du volet en garderie sur le développement des enfants. Il vise également à étayer les résultats d'analyses d'impact à court terme du volet Ateliers familles sur les attitudes et comportements des parents. Précisons que ces analyses se fondent sur les données collectées de mai 2007 à octobre 2009 auprès de la première cohorte de participants, et ce, dans quatre des six communautés à l'étude.<sup>11</sup> Le chapitre 2 détaille les aspects méthodologiques du

---

<sup>11</sup> Tôt dans le projet, la décision a été prise d'exclure les communautés d'Edmonton et de Saint-Jean des analyses d'impact. Ces communautés n'avaient pas de garderies francophones pouvant servir de contrefactuel. Plutôt que d'inclure ces communautés dans certaines analyses et de les exclure d'autres analyses, nous avons choisi de rapporter un ensemble cohérent d'estimés des impacts du programme dérivés d'un même échantillon pour toutes les mesures directes du développement des enfants. En conséquence, nous pouvons plus facilement interpréter certains aspects des résultats qui dépendent du groupe témoin utilisé pour estimer l'effet du programme.



projet notamment le devis expérimental adopté, l'échantillon de l'étude, les outils de mesure et les hypothèses de l'étude. Le chapitre 3 porte plutôt sur la stratégie analytique adoptée pour effectuer les analyses d'impact. Le chapitre 4 concerne les stratégies mises en place pour assurer la qualité et la fiabilité des données collectées ainsi que le résultat d'analyses préliminaires portant sur l'imputation des données manquantes, la spécification des termes d'erreurs, et l'identification des variables contrôles. Les chapitres 5 et 6 couvrent, respectivement, les résultats des analyses d'impact sur les résultantes développementales des enfants de la première et deuxième année de la mise en œuvre du projet. Le chapitre 5 présente également les résultats d'analyses d'impact du volet Ateliers familles sur différentes dimensions des parents participant à ces derniers. Nous concluons, au chapitre 7, avec un retour sur les principaux constats et une discussion sur l'ensemble des résultats rapportés.

Le lecteur est prié de noter que ce rapport s'insère dans une série de rapports. Il fait suite au rapport *Capacité d'apprentissage dans les communautés francophones en situation minoritaire : Rapport de référence* (Legault, Mák, Verstraete et Bérubé, 2014; titre abrégé : *Rapport de référence*) dont la version finale a été remise à RHDCC le 13 octobre 2009. Ce premier rapport établissait le profil des enfants, de leur famille et des communautés participant au projet. Ajoutons que le présent rapport a été rédigé en parallèle avec et se veut complémentaire au rapport *Capacité d'apprentissage dans les communautés francophones en situation minoritaire : Rapport de mise en œuvre du projet* (Bérubé, Legault, Janisse, Carson, Saucier et Lefebvre, 2014; titre abrégé : *Rapport de mise en œuvre du projet*)<sup>12</sup>. Trois autres rapports sont prévus, chacun portant respectivement sur : les analyses d'impact et l'étude de la mise en œuvre du programme préscolaire de service de garde auprès des deux cohortes de participants; les impacts un an suivant la fin de l'intervention; et les impacts deux ans après la fin de l'intervention.

---

<sup>12</sup> Nommé auparavant *Rapport intégratif, version finale: L'évaluation de la mise en œuvre du Projet pilote de garde d'enfants*.

## 2. Méthodologie

Ce chapitre porte sur les aspects méthodologiques utilisés dans le cadre du projet Capacité d'apprentissage. Une première section décrit les critères d'admissibilité et la population ciblée tandis que la seconde présente le devis expérimental du projet. Dans une troisième section, on explore certaines des menaces à la validité interne et les stratégies mises en place pour les contrer. La quatrième section détaille les mesures utilisées dans les analyses d'impact du programme testé. La dernière section concerne la série d'hypothèses testées dans le cadre des analyses d'impact.

### 2.1 POPULATION CIBLÉE

Les participants devaient satisfaire aux critères d'admissibilité suivants :

- Un des parents des enfants devait être un « ayant droit » selon la Charte canadienne des droits et libertés (article 23), c'est-à-dire que les enfants étaient éligibles à l'école de langue française. De plus, les enfants de nouveaux arrivants non ayants droit dont la première langue acquise est le français étaient aussi éligibles à participer.
- Les enfants (de la première cohorte) devaient être nés en 2004 ou en janvier 2005.
- Les parents devaient avoir l'intention d'inscrire leur enfant à l'école en français.

Le programme testé visait à augmenter les compétences langagières et le niveau de préparation à la scolarisation des enfants devant être inscrits dans des écoles de langue française. Le premier critère a été établi afin de rejoindre la population cible, c'est-à-dire les enfants ayant le droit de fréquenter l'école en français. Le second critère correspond au critère d'admissibilité à l'inscription à la maternelle en septembre 2009 des ministères de l'Éducation de l'Ontario et du Nouveau-Brunswick. Le troisième critère avait pour but d'exclure de l'étude les enfants dont les parents avaient la ferme intention de les inscrire dans une école de langue anglaise. Ce critère a été en réalité peu appliqué puisque les parents d'enfants aussi jeunes n'ont généralement pas pris leur décision quant au choix de l'école. Par contre, si les parents répondaient que l'école anglaise était leur choix final, la SRSA prenait la décision de ne pas poursuivre avec l'obtention du consentement éclairé étant donné que le nouveau programme préscolaire vise en partie à mieux préparer les enfants à l'école en français.

### 2.2 DEVIS EXPÉRIMENTAL

L'évaluation du programme se fait au moyen d'un devis quasi expérimental avec groupes témoins non équivalents. Un devis quasi expérimental partage avec les études expérimentales sur le terrain l'objectif de tester — au moyen de groupes témoins et de mesures prises avant l'intervention — l'hypothèse causale selon laquelle une intervention a un effet significatif sur les variables à l'étude au-delà de ce qui prend place en l'absence de ce programme et indépendamment d'autres facteurs externes pouvant entrer en jeu.

La méthodologie prévoit trois groupes expérimentaux : un groupe programme composé d'enfants inscrits dans des garderies francophones offrant le nouveau programme préscolaire; un groupe témoin en garderie comprenant des enfants inscrits dans des garderies francophones n'offrant pas le nouveau programme; et un groupe témoin hors garderie regroupant des enfants dont la garde de jour a lieu à la maison ou en garderie familiale non réglementée. Le premier groupe témoin vise à tenir compte de l'influence d'une garderie en installation sur le développement des enfants, ce qui constitue en soi un traitement. Le deuxième groupe témoin vise à tenir compte de l'influence d'un milieu de garde informel sur le développement de l'enfant, surtout au niveau de la langue française. Les enfants inscrits à une garderie de langue anglaise ont été ajoutés au groupe témoin hors garderie pour les analyses d'impact. Bien que ces enfants soient exposés à un environnement de garde formel, ils ne sont pas exposés au français. En ne classifiant pas ces enfants dans le groupe témoin en garderie, nous préservons l'homogénéité de la langue dans le groupe témoin en garderie et l'exposition à un programme en garderie offert en français.

## **2.3 VALIDITÉ INTERNE**

En l'absence d'une assignation aléatoire, il est probable qu'il existe au préalable des différences inhérentes entre le groupe programme et les groupes témoins. La mise en place de conditions visant à assurer la validité interne de l'étude devient encore plus importante afin d'éliminer d'emblée les explications alternatives *plausibles* des résultats. En effet, l'utilisation de techniques de contrôle, souvent statistiques, une fois l'étude terminée est moins avantageuse.

Dans le cadre du projet Capacité d'apprentissage, les conditions nécessaires pour assurer la validité interne sont:

- la taille de l'échantillon;
- une stratégie d'échantillonnage assurant une distribution équitable entre les groupes;
- des mesures pré-intervention prises sur les résultantes et les facteurs qui leur sont associés;
- une vérification de l'effet d'attrition de l'échantillon sur la distribution des groupes;
- une attention aux situations pouvant mener à la contamination des groupes témoins.

### **2.3.1 La taille de l'échantillon**

La validité interne de l'étude d'impact avec trois groupes exige un échantillon minimal de 156 enfants répartis également entre le groupe programme ( $n = 52$ ), le groupe témoin en garderie ( $n = 52$ ) et le groupe témoin hors garderie ( $n = 52$ ). Ce nombre d'enfants par groupe produit la puissance statistique nécessaire pour détecter un impact moyen, avec un très bon niveau de confiance que la valeur réelle de l'impact estimé se situe à l'intérieur de la valeur estimative trouvée, plus ou moins un certain montant (c.-à-d. 19 fois sur 20, nous obtiendrions les mêmes résultats auprès d'autres échantillons).

### **2.3.2 Stratégie d'échantillonnage ciblé**

Une stratégie d'échantillonnage ciblé a été choisie dans le but de créer des groupes témoins relativement homogènes au groupe programme de façon à éliminer l'influence, sur les résultantes à l'étude, de facteurs autres que ceux déjà mesurés. Des études antérieures ont démontré que certaines caractéristiques, comme le revenu familial et le niveau d'éducation des parents, expliquaient une partie de la trajectoire développementale de l'enfant. Il s'ensuit que si le groupe programme diverge considérablement du groupe témoin sur l'une de ces caractéristiques, les différences observées entre les deux groupes pourraient bien être expliquées par ces dernières et non par le programme. Ainsi, une attention particulière a été apportée afin de cibler des participants des groupes témoins présentant un profil sociodémographique similaire à celui du groupe programme (p. ex., le niveau socio-économique) et vivant dans le même arrondissement, assurant ainsi qu'ils ont accès aux mêmes ressources et services francophones que le groupe programme.<sup>13</sup> Un premier contrôle sur la composition des groupes a été fait lors de l'inscription des familles au projet, notamment au niveau de l'emplacement du domicile des membres potentiels des groupes témoins. Un deuxième contrôle fut l'utilisation des mesures pré-intervention, le sujet de la prochaine section.

### **2.3.3 Les mesures pré-intervention**

L'inférence causale découlant d'une méthodologie quasi expérimentale est facilitée par l'utilisation de mesures pré-intervention prises sur les résultantes et les facteurs qui leur sont associés. Dans le cas du projet Capacité d'apprentissage, la résultante à l'étude est le niveau de préparation à la scolarisation des enfants mesuré au moyen de l'Évaluation de la petite enfance — Appréciation directe (ÉPE-AD) et de ses sous-échelles de vocabulaire expressif et réceptif.

Des mesures pré-intervention permettent de mieux saisir l'impact du programme sur le développement de l'enfant puisqu'on pourra effectuer une comparaison du niveau de préparation à la scolarisation de tous les enfants au moment précédent la mise en œuvre du programme avec le niveau observé suite à cette mise en œuvre. Ces mesures permettent également de vérifier si les enfants du groupe programme et ceux des groupes témoins présentent un profil développemental similaire avant l'intervention. Cette information est utile puisque les estimés de l'impact du programme se trouveraient biaisés si l'un des groupes expérimentaux se composait d'enfants démontrant un développement plus avancé comparativement à un autre groupe avant que ne débute l'intervention.

Des mesures pré-intervention des variables sociodémographiques et socio-économiques visent également à établir si les enfants des trois groupes expérimentaux vivent des conditions similaires avant l'intervention. Le choix des variables repose sur les résultats d'études antérieures démontrant que ces dernières influençaient significativement le niveau de préparation à la scolarisation des enfants. Le contrôle statistique de ces variables dans les analyses permettra de distinguer les impacts du nouveau programme préscolaire en garderie sur le développement des enfants des impacts d'autres variables pour lesquels les évidences empiriques ont déjà été établies.

---

<sup>13</sup> Le lecteur intéressé à en apprendre plus est invité à lire le Rapport sur les plans de travail et de méthodologie révisés remis à RHDCC le 30 mars 2007.

Le profil initial des enfants, de leur famille et des communautés participant au projet Capacité d'apprentissage a été rapporté dans le *Rapport de référence* (Legault et al., 2014). Ce rapport contient également le résultat d'analyses établissant l'homogénéité entre les groupes expérimentaux avant que l'intervention n'ait lieu. Notons que les résultats rapportés dans le rapport découlaient d'analyses portant sur les six communautés. Ces analyses ont dû être refaites pour le présent rapport pour inclure uniquement les quatre communautés incluses dans les analyses d'impact. La liste des variables identifiées se retrouve à la Section 4.5 de ce rapport alors que les résultats des analyses comparatives sont compris dans la Section 4.4 de ce rapport.

### **2.3.4 Attrition**

La perte de participants au cours de l'étude entraîne une menace à la validité externe de l'étude (c.-à-d. jusqu'à quel point l'échantillon représente la population qu'il est censé représenter) ainsi qu'à la validité interne (c.-à-d. en changeant la composition des groupes expérimentaux). Elle peut être due à plusieurs raisons dont un déménagement ou encore une perte d'intérêt des participants. Peu importe la raison du retrait, il importe d'établir si le groupe de participants qui se sont retirés de l'étude se distingue du groupe de participants demeurant dans l'étude quant à certaines dimensions sociodémographiques. Des différences entre les deux groupes pourraient signaler la présence d'un sous-groupe de la population cible possédant des caractéristiques particulières (ce qui constitue une menace à la validité externe). En outre, un changement important dans la composition des groupes expérimentaux pourrait affecter la validité des estimés d'impact (c.-à-d. affecter la validité interne). Un contrôle des raisons du retrait au projet a été donc été instauré dès le début du projet afin de mieux évaluer la menace à la validité interne de l'étude découlant des circonstances rattachées aux retraits.

### **2.3.5 Prévention de la contamination des groupes témoins**

Il y a contamination lorsque des changements dans les scores obtenus sur les résultantes des groupes témoins sont dus à la mise en application des conditions et modalités du programme testé. En d'autres mots, il y a contamination lorsque les membres des groupes témoins sont exposés aux composantes du programme en garderie et/ou ceux du programme Ateliers familles.

La contamination du groupe témoin en garderie est plus probable dans les communautés où plus d'une des garderies francophones de la région participent au projet, rendant ainsi possible la circulation des informations entre la garderie programme et la garderie témoin. Afin de minimiser les occasions de transfert d'information, plusieurs stratégies ont été mises en place (une liste est fournie ci-dessous). En outre, les pourvoyeurs des services de garde ont été sensibilisés aux situations propices à une contamination des groupes témoins et à l'importance pour l'étude d'éviter ou de minimiser ces occasions. Cette sensibilisation était surtout ciblée dans les communautés comptant plusieurs garderies francophones sous l'égide d'un même pourvoyeur de services de garde. Ces discussions ont résulté en une collaboration étroite entre la SRSA et les pourvoyeurs de services de garde.

Les stratégies mises en place pour minimiser les occasions de contamination étaient :

Au niveau du volet en garderie :

- Des formations de base et de suivi étaient offertes seulement aux éducatrices et aide-éducatrices impliquées dans la livraison du programme en garderie.

- Le mouvement des éducatrices formées au programme en garderie à une garderie témoin était limité. Dans les faits, le mouvement du personnel d'une garderie programme à une garderie témoin a eu lieu seulement dans une communauté. Dans un cas, l'éducatrice a été affectée à un groupe d'enfants plus jeunes que les enfants du projet Capacité d'apprentissage. Dans les autres cas, les aides-éducatrices ont été affectées à des garderies ne participant pas au projet Capacité d'apprentissage. Tout personnel quittant la garderie programme était sensibilisé à l'importance de ne pas partager les connaissances et les façons de faire avec les autres éducatrices et était tenu de remettre à la coordonnatrice le manuel de formation. Enfin, la coordonnatrice surveillait de près la situation dans la garderie témoin.
- Les classes du groupe témoin en garderie se situaient dans un édifice distinct de celui dans lequel se trouvaient les classes du groupe programme.
- Les ressources et le matériel du programme en garderie devaient rester dans la garderie programme jusqu'à trois mois suivant la fin de la prestation du programme testé.

Au niveau du volet Ateliers familles :

- La formation de base et le suivi étaient offerts seulement aux intervenantes impliquées dans la livraison des Ateliers familles.
- À la fin de la prestation du volet Ateliers familles, les intervenantes devaient remettre à la coordonnatrice le manuel de formation. Elles étaient également sensibilisées à l'importance de ne pas partager les connaissances et les façons de faire avec d'autres personnes.
- Les pourvoyeurs de services de garde et les conseils scolaires étaient tenus de ne pas livrer le programme Ateliers familles ou un programme similaire dans la communauté à l'étude.
- Les ressources et le matériel du Centre de ressources devaient rester dans la garderie programme jusqu'à trois mois suivant la fin de la prestation du programme testé.

## 2.4 ÉCHANTILLON PAR COMMUNAUTÉS ET PAR GROUPES EXPÉRIMENTAUX

Les participants du projet faisant partie des analyses d'impact sont issus des communautés minoritaires francophones de Cornwall, Orléans et Durham en Ontario, et d'Edmundston au Nouveau-Brunswick.<sup>14</sup> Le recrutement des participants a débuté le 1er mai 2007 et s'est terminé le 31 octobre 2007.

*Échantillon global* : Au moment de l'inscription, le projet comptait 254 enfants provenant de 250 familles. Tel qu'observé au tableau 2.1, les communautés d'Edmundston et de Cornwall comptaient les plus grandes proportions de participants (33,5 % et 28,3 %, respectivement), suivies par les communautés d'Orléans (21,7 %) et de Durham (16,5 %). L'échantillon

---

<sup>14</sup> L'absence de garderies francophones pouvant servir de contrefactuel rendait impossible de bien mesurer l'impact de ce programme dans les communautés d'Edmonton et de Saint-Jean.

comprenait 244 enfants à la fin de la première année du projet, dont 116 garçons et 128 filles. La taille de l'échantillon a diminué légèrement à 240 enfants à la fin de la deuxième année du projet, dont 114 garçons et 126 filles.

L'âge moyen des enfants à l'inscription était de 3.0 ans. À ce moment, le projet Capacité d'apprentissage comptait un nombre plus élevé de filles (52,5 %) que de garçons (47,5 %). La langue maternelle des enfants de l'échantillon (selon le répondant qui connaît le mieux l'enfant) était en majorité le français (72,4 %) suivi par l'anglais (16,5 %).

D'après le sondage de base, l'âge de la mère à la naissance de l'enfant ciblé était en moyenne de 27 ans. De plus, les trois quarts de ces mères ont obtenu au moins un diplôme collégial, et la moitié de ces dernières ont un diplôme universitaire. La taille du ménage se chiffrait à quatre membres en moyenne. On observait 9,4 % de familles monoparentales. Plus de la moitié des familles participantes (60,7 %) avaient un revenu annuel de plus de 70,000\$; le revenu annuel médian se retrouvait entre 80,000\$ et 99,999\$.

Au niveau du profil linguistique des familles participantes, environ le deux tiers des mères (63,1 %) et des pères (60,9 %) parlaient seulement français à leur enfant. La majorité des enfants provenaient de foyers endogames francophones (58 %) alors qu'un quart provenaient de foyers exogames (23,5 %).<sup>15</sup>

**Tableau 2.1 : Répartition des participants par communauté depuis l'inscription**

	Inscription	1 <sup>ère</sup> année	2 <sup>e</sup> année
<b>Communauté</b>			
<b>Cornwall</b>	72 (28,3 %)	69 (28,3 %)	68 (28,3 %)
<b>Durham</b>	42 (16,5 %)	36 (14,8 %)	35 (14,6 %)
<b>Edmundston</b>	85 (33,5 %)	84 (34,4 %)	83 (34,6 %)
<b>Orléans</b>	55 (21,7 %)	55 (22,5 %)	54 (22,5 %)
<b>Total</b>	254 (100 %)	244 (100 %)	240 (100 %)

*Échantillon par groupes expérimentaux* : En octobre 2007, le projet comptait 77 enfants inscrits au groupe programme (G1), 99 enfants au groupe témoin en garderie (G2) et 78 enfants au groupe témoin hors garderie (G3). L'examen du tableau 2.2 révèle une augmentation du G2 accompagnée d'une diminution du G1 et du G3 en février 2009. Ces changements dans la composition des groupes expérimentaux s'expliquent par l'inscription à une école des petits ontariens en septembre 2008, soit le début de la deuxième année du projet. En outre, ces fluctuations sont plus importantes à Cornwall où plusieurs enfants du G1 et du G3 ont été inscrits à une école abritant une garderie témoin.<sup>16</sup> Les changements dans la composition des groupes

<sup>15</sup> La catégorisation des foyers s'est faite à partir de la combinaison de la Première langue officielle parlée (PLOP) de la mère et de la PLOP du père.

<sup>16</sup> Bien que les enfants soient inscrits à l'école dès septembre 2008, les changements dans la composition des groupes expérimentaux ne paraissent pas immédiatement dans la collecte de données en raison des règles d'attribution. Ces dernières sont détaillées à la Section 3.1.2.

expérimentaux ont donc été pris en considération dans les analyses de la deuxième année du programme.

**Tableau 2.2 : Répartition des participants par groupe expérimental**

Groupe expérimental	oct. 2007	fév. 2008	juin 2008	oct. 2008	fév. 2009	juin 2009	oct. 2009
<b>G1</b>	77 (30,3 %)	74 (29,7 %)	70 (28,2 %)	70 (28,7 %)	66 (27,2 %)	63 (26,3 %)	63 (26,3 %)
<b>G2</b>	99 (39,0 %)	95 (38,2 %)	96 (38,7 %)	95 (38,9 %)	112 (46,1 %)	113 (47,1 %)	113 (47,1 %)
<b>G3</b>	78 (30,7 %)	80 (32,1 %)	82 (33,1 %)	79 (32,4 %)	65 (26,7 %)	64 (26,3 %)	64 (26,7 %)
<b>Total</b>	254 (100 %)	249 (100 %)	248 (100 %)	244 (100 %)	243 (100 %)	240 (100 %)	240 (100 %)

### 2.4.1 Taux de rétention

Le projet Capacité d'apprentissage a un excellent taux de rétention, comptant seulement 14 retraits (5,5 %) depuis le début du projet en 2007. La raison principale pour le retrait d'un enfant du projet est le déménagement de la famille à l'extérieur de la communauté. Notons que certains enfants ont continué à être suivis dans les cas où le déménagement s'est fait vers une autre communauté participant au projet. Ces enfants étaient désormais considérés comme des participants de la communauté d'accueil. Le tableau 2.3 rapporte le décompte des raisons de retrait du projet.

**Tableau 2.3 : Raisons justifiant le retrait de l'enfant du projet Capacité d'apprentissage**

Raison pour le retrait	1ère année	2e année	Total
Déménagement à l'extérieur d'une communauté participante	4	2	6
L'enfant passe à une garderie anglophone/pas assez d'anglais à la garderie programme	3	0	3
Manque de disponibilité des familles	1	2	3
Dérangé par des questions du sondage de base	1	0	1
Perte de contact avec la famille	1	0	1
<b>Total des retraits</b>	<b>10</b>	<b>4</b>	<b>14</b>

## 2.5 MESURES

Cette section présente le plan des collectes de données auprès des parents et des enfants, ainsi que pour les Ateliers familles. Elle se poursuit avec une description des outils de mesures quantitatifs utilisés depuis le début du projet, soit l'été 2007, jusqu'à la fin de la deuxième année, soit octobre 2009.



## 2.5.1 Plan des collectes de données

Les collectes de données auprès des parents et des enfants se sont déroulées sur une base trimestrielle. La collecte des données de base a été effectuée de mai à octobre 2007. Les collectes suivantes ont eu lieu en février, juin et octobre 2008, ainsi qu'en février, juin et octobre 2009. Cette dernière collecte de données a été effectuée post-programme, c'est-à-dire au moment où tous les enfants étaient inscrits à temps plein dans une école.

### *Auprès des parents*

Au total, on compte huit sondages auprès des parents depuis le début du projet Capacité d'apprentissage. Le sondage de base a été réalisé lors de l'inscription de l'enfant au projet, soit du 1<sup>er</sup> mai au 31 octobre 2007. D'une durée d'environ une heure, le sondage s'est fait face-à-face par la coordonnatrice communautaire avec le parent connaissant le mieux l'enfant (PCM). Le premier sondage de suivi fut administré seulement aux parents s'ayant inscrit au projet avant le 1er septembre 2007. Ce sondage servait à actualiser l'information relative au type et à la fréquence des activités de littératie et des activités hors garderie faites avec l'enfant à l'automne 2007.<sup>17</sup> Puisque cette mise-à-jour n'a pas été nécessaire pour les parents ayant répondu au sondage de base à l'automne, nous comptons un nombre relativement faible de répondants pour ce deuxième sondage. Les sept sondages suivants étaient de courte durée, variant entre 10 et 30 minutes. Ils ont été administrés par téléphone par les coordonnatrices ou par un membre de la SRSA. Le taux de réponse moyen des sondages de suivi était de 97,9 % (voir le tableau 2.4).

**Tableau 2.4 : Taux de réponses des sondages auprès des parents**

	<b>Sondage de base</b>				<b>Sondages de suivi</b>			
	<i>mai à oct. 2007</i>	<i>nov. 2007*</i>	<i>fév. 2008</i>	<i>juin 2008</i>	<i>oct. 2008</i>	<i>fév. 2009</i>	<i>juin 2009</i>	<i>oct. 2009</i>
Sondages administrés	254	175	241	244	238	240	237	235
Absents	0	8	8	5	6	3	3	5
Retraits	0	5	5	5	10	11	14	14
Taux de réponse	100 %	95,6 %	96,8 %	98,0 %	97,5 %	98,8 %	98,8 %	97,9 %

*Note* : \* Ces données exclues les participants inscrits après le 1er septembre 2007.

### *Pour les Ateliers familles*

Puisque le nouveau programme testé comprenait un volet Ateliers familles ciblant les parents du groupe programme, nous avons prévu une évaluation de l'impact de ce volet sur les attitudes et comportements des parents. Deux sondages ont été développés afin de recueillir des informations avant et après la mise en œuvre des Ateliers familles. Le sondage pré-intervention a été administré par téléphone auprès de tous les parents du groupe programme au même moment

<sup>17</sup> Pour les analyses d'impact, les données du premier sondage de suivi ont remplacé les données du sondage de base lorsqu'elles étaient présentes. Ces deux sondages sont donc considérés comme étant un seul sondage dans le cadre des analyses.

que le premier sondage de suivi en novembre 2007. Ce sondage a permis de mesurer les attentes, les opinions et certains comportements des parents du groupe programme au début de l'intervention des Ateliers familles. Un sondage post-atelier fut administré trois semaines après le dernier atelier afin de mesurer l'évolution de ces résultantes. Son administration fut également faite par téléphone.

Un troisième sondage anonyme a été administré lors du dernier atelier pour solliciter l'opinion des parents sur les aspects logistiques des ateliers, notamment au niveau de l'environnement physique, de la qualité des échanges, et de la qualité de l'animation des intervenantes. L'ensemble de ces données a servi à évaluer la qualité de la mise en œuvre du volet Ateliers familles dont les résultats se retrouvent dans le *Rapport de mise en œuvre du projet* (Bérubé et al., 2014). Le taux de réponse était excellent pour les sondages pré- et post-atelier, mais relativement bas pour le sondage de logistique en raison de la faible assiduité à cet atelier (voir le tableau 2.5).

**Tableau 2.5 : Taux de réponses des sondages liés aux Ateliers familles**

	<b>Pré-intervention</b>	<b>Sondage de logistique</b>	<b>Post-intervention</b>
	<i>nov. 2007</i>	<i>Atelier 10</i>	<i>mars 2008</i>
Sondages administrés	107	69	105
Absents	6	40	4
Taux de réponse	94,7 %	63,3 %	96,3 %

### ***Auprès des enfants***

À partir d'octobre 2007, chaque période de sondage auprès des parents était jumelée avec une évaluation auprès des enfants. En tout, sept évaluations ont été administrées à tous les quatre mois auprès des enfants pendant les deux premières années du projet Capacité d'apprentissage avec un taux de réponse moyen de 98,6 % (voir le tableau 2.6).

**Tableau 2.6 : Taux de réponses des évaluations auprès des enfants**

	<i>Périodes d'évaluations</i>						
	<i>oct. 2007</i>	<i>fév. 2008</i>	<i>juin 2008</i>	<i>oct. 2008</i>	<i>fév. 2009</i>	<i>juin 2009</i>	<i>oct. 2009</i>
Enfants évalués	245	247	243	240	241	237	237
Absents	4	2	6	4	2	3	3
Retraits	5	5	5	10	11	14	14
Taux de réponse	98,4 %	99,2 %	97,6 %	98,4 %	99,2 %	98,8 %	98,8 %

## 2.5.2 Mesures auprès des enfants

Au cours des différentes vagues de collectes de données, divers outils ont été utilisés pour suivre le développement des enfants sur plusieurs dimensions importantes liées à la préparation à la scolarisation. Le calendrier et les outils de chaque vague d'évaluations se retrouvent à l'annexe A.

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

L'évaluation de la petite enfance – Appréciation directe (ÉPE–AD) mesure quatre domaines de la préparation à la scolarisation et un cinquième domaine conçu spécifiquement aux fins du projet Capacité d'apprentissage:

(Domaine A) Conscience de soi et de l'environnement;

(Domaine B) Habiletés cognitives;

(Domaine C) Langage et communication;

(Domaine D) Développement physique et moteur; et

(Domaine E) Conscience et engagement envers la culture francophone.

Chaque domaine comporte un certain nombre de questions présentées en ordre croissant de difficulté. Le domaine A, Conscience de soi et de l'environnement, est composé de 16 questions permettant de savoir jusqu'à quel point l'enfant peut reconnaître et identifier des éléments présents dans son environnement. Par exemple, on demande à l'enfant de nommer des couleurs, des parties du corps, certains professionnels et leur rôle, en plus de sa date de naissance.

Le domaine B, Habiletés cognitives, est composé de 17 questions mesurant différents aspects logico-mathématiques. Par exemple, on demande à l'enfant de compter, de former des groupes à partir de différents objets, de comparer différentes formes et de distinguer la forme la plus grande de la plus petite. De plus, on demande à l'enfant de nommer quelques lettres de l'alphabet, d'identifier leur son, et de lire 8 mots.

Le domaine C, Langage et communication, comprend 14 items mesurant les capacités de communication et de compréhension de l'enfant. L'évaluatrice demande à l'enfant de lui montrer du doigt des images représentant un mot qu'elle lui dit, de répéter des phrases de sept syllabes, de répondre à des questions avec des phrases complètes, d'utiliser des images pour raconter une histoire et de démontrer le sens de quatre mots d'action. C'est le seul domaine administré à tous

les enfants en français puisque ce domaine fait partie de l'arbre décisionnel utilisé pour décider de la langue d'administration de l'évaluation.<sup>18</sup>

Le domaine D, Développement physique et moteur, est composé de 16 items mesurant la capacité de l'enfant à accomplir certaines activités de motricité fine et de motricité globale. La motricité fine englobe plusieurs éléments de la pré-écriture comme, par exemple, pouvoir tracer des lettres, dessiner une ligne droite ou colorier entre les lignes en tenant son crayon correctement. La motricité globale concerne les habiletés de l'enfant à bouger son corps, par exemple, en sautant au-dessus d'un objet, en se tenant en équilibre sur un pied ou en sautillant sur un pied trois fois de suite. Il est à noter qu'une étude canadienne basée sur les données de l'ELNEJ indique qu'après l'âge de trois ans, on observe peu de différence entre les enfants ayant un développement normal à ce domaine. Ces résultats ont été confirmés lors l'évaluation de février 2009. Pour cette raison, ce domaine n'a pas été mesuré lors les évaluations effectuées en juin et octobre 2009.

Les items sont notés sur une échelle de 5 points pour les domaines A, B et C et de 4 points pour le domaine D. Les trois premiers domaines requièrent l'utilisation de planches d'évaluation ou d'objets que l'enfant doit montrer du doigt ou manipuler. Durant l'administration du test, il est possible de voir un enfant atteindre un « plateau » dans un domaine donné lorsque les questions deviennent trop difficiles. En effet, l'évaluation d'un domaine est arrêtée lorsque l'enfant obtient un score de « 0 » ou « 1 » à trois items consécutifs. L'évaluatrice passe alors au domaine suivant. L'administration du test dure environ de 45 minutes à une heure.

Jusqu'en février 2009, l'enfant pouvait être évalué en anglais ou en français selon le score obtenu au domaine E et celui obtenu au domaine C. Le domaine E, conçu spécifiquement pour le projet Capacité d'apprentissage, mesurant la conscience et l'engagement envers la culture francophone de l'enfant, servait davantage à établir un rapport amical avec l'enfant et déterminait la langue du test.<sup>19</sup> Son utilisation fut discontinuée en février 2009 puisqu'il s'est avéré non fiable.<sup>20</sup>

L'ÉPE-AD présente de bonnes propriétés psychométriques. Le résultat des analyses factorielles de l'ÉPE-AD a confirmé l'unidimensionnalité de chacun des domaines de la version française du test, tel que postulée théoriquement.<sup>21</sup> La cohérence interne de chacune des dimensions s'est avérée acceptable à très bonne avec des alphas de Cronbach variant de 0,60 à 0,94 pour la version française du test et de 0,79 à 0,95 pour la version anglaise (voir le tableau 2.7).

---

<sup>18</sup> La procédure d'administration de l'ÉPE-AD ainsi que l'arbre décisionnel pour déterminer la langue du test sont détaillés à l'annexe B.

<sup>19</sup> Le domaine E contient des questions sur les préférences des enfants en matière de livres, d'émissions de télévision et de chansons, sur la langue de ces ressources (anglais ou français), ainsi que sur les langues parlées avec leurs parents et ami(e)s.

<sup>20</sup> Le lecteur est référé au rapport *Capacité d'apprentissage dans les communautés francophones en situation minoritaire : Rapport de référence* (Legault et al., 2014) pour connaître les raisons sous-jacentes à cette décision.

<sup>21</sup> Il a été impossible d'effectuer une analyse factorielle sur la version anglaise du test en raison du trop petit nombre d'individus ayant répondu dans cette langue (63 cas au domaine C et 44 dans les autres domaines). Un minimum de 160 - 170 cas aurait été souhaitable par domaine afin d'assurer la validité statistique des calculs (Tabachnick & Fidell, 2006).

Tableau 2.7 : Alpha de Cronbach des domaines de l'ÉPE-AD par vague d'évaluation

<i>Alpha de Cronbach</i> [ $\alpha$ ( $n$ )]							
	<i>oct. 2007</i>	<i>fév. 2008</i>	<i>juin 2008</i>	<i>oct. 2008</i>	<i>fév. 2009</i>	<i>juin 2009</i>	<i>oct. 2009</i>
<b>Domaines administrés en français</b>							
<b>A</b>	,91 (213)	,91 (215)	,91 (215)	,90 (220)	,90 (232)	,89 (237)	,90 (236)
<b>B</b>	,85 (213)	,87 (215)	,86 (215)	,83 (220)	,86 (233)	,75 (237)	,78 (237)
<b>C</b>	,91 (245)	,93 (247)	,93 (243)	,93 (240)	,91 (240)	,63 (237)	,60 (237)
<b>D</b>	,89 (212)	,92 (215)	,94 (215)	,93 (220)	,93 (232)	s.o.	s.o.
<b>Domaines administrés en anglais</b>							
<b>A</b>	,87 (30)	,91 (32)	,90 (28)	,91 (20)	,90 (6)	s.o.	s.o.
<b>B</b>	,79 (30)	,87 (32)	,87 (28)	,90 (20)	,88 (7)	s.o.	s.o.
<b>C</b>	,89 (43)	,87 (40)	,90 (35)	,88 (27)	,95 (8)	s.o.	s.o.
<b>D</b>	,86 (30)	,94 (31)	,91 (28)	,95 (20)	,96 (7)	s.o.	s.o.

### **Épreuve de dénomination de Gardner (Ska, 1995)**

L'Épreuve de dénomination de Gardner (1979) mesure le vocabulaire expressif de l'enfant. L'outil est la traduction validée et normalisée en français pour les enfants de la maternelle à la deuxième année de l'outil *Expressive One-Word Picture Vocabulary Test*, version française (EOWPVT-F). Ainsi, les consignes et l'ordre de difficulté des items ont été ajustés pour une population francophone. L'exercice consiste à présenter une série d'images à l'enfant et de lui demander de nommer le mot associé à l'image. Après six erreurs consécutives, le test est arrêté, résultant en un temps d'administration variant entre 10 et 15 minutes. L'avantage de ce test est qu'il est très sensible aux différences dans le niveau de maîtrise de la langue française. Le score total représente la somme des mots correctement identifiés en français par l'enfant. On tient également compte de certaines réponses incorrectes, soit des mots anglais ou des expressions régionales, offertes par l'enfant. Ces réponses incorrectes sont récoltées à titre d'informations seulement, et ne servent pas au calcul d'un score.

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

L'ÉVIP-R mesure le vocabulaire réceptif de l'enfant. Son administration ressemble à un jeu : l'enfant entend un mot prononcé à voix haute et doit désigner, parmi quatre images, celle qui représente le mot entendu. Le test s'arrête après six erreurs parmi huit essais. Selon les règles du test, le point de départ où l'enfant commence le test est déterminé selon son âge et sa performance. Pour les besoins du projet Capacité d'apprentissage, un point de départ commun pour tous les enfants (soit la 15<sup>e</sup> question qui correspond au point de départ des enfants âgés de 3

ans et demi) a été établi peu importe leur âge ou leur niveau de performance. Le but du présent projet n'étant pas de comparer la performance des enfants à une « norme »<sup>22</sup>, l'utilisation d'un point de départ commun suffisamment facile, nous assure de capter le niveau de vocabulaire réceptif auprès des enfants issus de famille exogame.

Le score de l'ÉVIP-R représente la somme des bonnes réponses données par l'enfant à compter de la 15<sup>e</sup> question. Cette façon de calculer le score de l'ÉVIP-R diffère légèrement de celle préconisée dans le manuel du test (Dunn et al., 1993, p. 13-15). Nous observons une corrélation supérieure à .99 entre la méthode adoptée par la SRSA et celle mise de l'avant dans le manuel. C'est donc dire qu'il n'y a aucune perte d'information. La méthode utilisée par la SRSA pour calculer le score de l'enfant à l'ÉVIP-R comporte plusieurs avantages notamment, elle: a) maximise la variance; et b) permet d'inclure les enfants qui n'ont pas pu faire le test parce qu'ils ont échoué aux essais d'entraînement (ils ont obtenu un score de zéro).

### ***Sous-échelles de vocabulaire (ÉPE-AD)***

Au cours de l'hiver 2009, la SRSA a remanié la composition des échelles de l'ÉPE-AD pour créer deux nouvelles sous-échelles de vocabulaire et ce, dans le but d'améliorer la capacité de mesurer de fines différences dans les compétences langagières des enfants. La première sous-échelle mesure le vocabulaire expressif, soit la capacité de l'enfant de dire le mot associé à l'image qui lui est présenté. Cette échelle contient six items du domaine A et deux items du domaine C. La cohérence interne est très bonne pour les items du vocabulaire expressif avec des alphas de Cronbach variant entre 0,81 et 0,88 pour les versions anglaise et française.

La deuxième sous-échelle mesure le vocabulaire réceptif et se compose de deux items du domaine A et trois items du domaine C. On présente à l'enfant une série d'images et celui-ci doit identifier l'image associée au mot prononcé par une évaluatrice. La cohérence interne des items mesurant le vocabulaire réceptif dans la version française est acceptable avec des alphas de Cronbach variant de 0,62 à 0,71. Par contre, la cohérence interne de la version anglaise de la sous-échelle de vocabulaire réceptif est moins acceptable puisqu'elle comporte plusieurs alphas de moins que 0,50 (voir le tableau 2.8).

---

<sup>22</sup> Précisons que les normes actuelles pour la version française du test ont été établies en 1990 et n'ont pas fait l'objet d'une mise à jour. En conséquence, leur validité est inconnue.

**Tableau 2.8 : Alpha de Cronbach des sous-échelles de vocabulaire de l'ÉPE-AD par vague d'évaluation**

	<i>Alpha de Cronbach</i> [ $\alpha$ ( <i>n</i> )]						
	<i>oct. 2007</i>	<i>fév. 2008</i>	<i>juin 2008</i>	<i>oct. 2008</i>	<i>fév. 2009</i>	<i>juin 2009</i>	<i>oct. 2009</i>
<b>Domaines administrés en français</b>							
Vocabulaire expressif (8)	,85 (213)	,88 (215)	,88 (215)	,88 (220)	,86 (233)	,85 (237)	,84 (237)
Vocabulaire réceptif (5)	,71 (213)	,69 (215)	,64 (215)	,52 (220)	,62 (233)	s.o.	s.o.
<b>Domaines administrés en anglais</b>							
Vocabulaire expressif (8)	,83 (30)	,86 (32)	,87 (28)	,81 (20)	,87 (6)	s.o.	s.o.
Vocabulaire réceptif (5)	,37 (30)	,73 (32)	,40 (28)	,53 (20)	,48 (6)	s.o.	s.o.

**Évaluation de la petite enfance — Appréciation directe, version modifiée (SRSA, 2009).**

Au cours de l'hiver 2009, la SRSA a constaté un potentiel élevé d'effet plafond suite aux premières analyses des résultats de l'ÉPE-AD et à des projections de scores pour la 5<sup>e</sup> évaluation prévue en juin 2009. Un effet plafond des scores entraverait la capacité de continuer à suivre l'évolution du développement des enfants et, par extension, il deviendrait impossible de tester pour des différences entre les enfants des trois groupes expérimentaux. RHDCC a donc retenu les services du concepteur de l'outil pour créer de nouvelles questions plus difficiles permettant ainsi de poursuivre la trajectoire développementale des enfants aux domaines A, B, et C jusqu'en octobre 2009. Le concepteur a également éliminé les questions les plus faciles du test afin de ne pas allonger le temps d'administration. La « version allongée de l'ÉPE-AD » (Willms, 2009) fut prête en mai 2009.

En parallèle, la SRSA a procédé à une restructuration des questions composant la version allongée de l'ÉPE-AD. Cette restructuration avait comme objectif de mesurer plus finement des habiletés langagières des enfants tout en conservant la capacité de suivre la trajectoire développementale des enfants selon trois des quatre domaines de l'ÉPE-AD (c.-à-d. conscience de soi et de l'environnement (domaine A); habiletés cognitives (domaine B); et langage et communication (domaine C)).<sup>23</sup> Le nouvel outil, nommé « l'ÉPE-AD modifiée », a aussi donné lieu à des sous-échelles mesurant le vocabulaire expressif et réceptif présentées plus tôt. L'ÉPE-AD modifiée comprend:

- 15 questions au domaine A concernant surtout le vocabulaire expressif et les connaissances générales;
- 12 questions au domaine B mesurant la conscience phonologique (soit l'habileté des enfants à jouer avec les sons qui forment les mots) et les concepts de la numératie (soit l'habileté des enfants à jouer avec les chiffres). Deux autres items du domaine B

<sup>23</sup> Le lecteur intéressé à connaître davantage sur le remaniement de l'ÉPE-AD et des connaissances théoriques l'informant est prié de consulter le *Rapport présentant une revue des outils de mesures directes des dimensions développementales des enfants* remis à RHDCC en juillet 2009.

ont été administrés de façon légèrement différente de l'ÉPE-AD allongée. Ainsi, la connaissance du son de toutes les lettres de l'alphabet a été mesurée au moyen d'un livret où les lettres majuscules sont présentées en ordre croissant de difficulté pour les petits francophones. La connaissance du nom de toutes les lettres de l'alphabet a été mesurée au moyen d'un livret où les lettres minuscules sont présentées en ordre croissant de difficulté pour les petits francophones.

- Enfin, deux questions du domaine C ont été conservés pour poursuivre l'échelle de vocabulaire expressif et deux nouvelles questions plus difficiles ont été ajoutées pour mesurer la conscience phonologique des enfants.

L'ÉPE-AD initiale a été l'outil utilisé jusqu'en février 2009 tandis que «l'ÉPE-AD modifiée», la version de l'outil retravaillée par la SRSA, fut utilisée lors des évaluations de juin et d'octobre 2009. Une comparaison du contenu des trois versions de l'ÉPE-AD se retrouve à l'annexe C.

Dans le cadre du projet Capacité d'apprentissage, il est important de suivre la trajectoire développementale des enfants d'octobre 2007 à octobre 2009. L'établissement d'une continuité dans les dimensions mesurées par deux outils passe par l'examen du patron de corrélations entre les sous-échelles mesurant la même dimension bien que compris dans chacun des outils. Une forte corrélation (soit de 0,75 et plus) observée entre les sous-échelles signifie qu'ils mesurent respectivement le même construit. L'examen du patron de corrélations au tableau 2.9 confirme la présence de fortes corrélations entre les domaines A et B de l'ÉPE-AD initiale (0,83 et 0,75 respectivement, évaluation de février 2009) et ceux de l'ÉPE-AD modifiée (évaluation de juin 2009). L'exception est le domaine de la communication (domaine C) où l'on observe une corrélation de 0,61. Ce résultat plus faible qu'espéré est toutefois pallié par de bonnes corrélations retrouvées entre le domaine C de l'ÉPE-AD initiale et la sous-échelle du vocabulaire expressif créé à partir de l'ÉPE-AD (0,84), ainsi qu'entre le domaine C de l'ÉPE-AD initiale et l'échelle «de vocabulaire expressif» de Gardner (0,75). On note également une forte corrélation entre le domaine C de l'ÉPE-AD initiale et l'échelle du domaine A de l'ÉPE-AD modifiée. L'ensemble de ces résultats permet de conclure que l'utilisation de la version modifiée entraîne une perte d'information négligeable. Par conséquent, nous pouvons suivre la trajectoire développementale des enfants de septembre 2007 à octobre 2009.



**Tableau 2.9 : Coefficients de corrélation Pearson des échelles de l'ÉPE-AD et des échelles de vocabulaire expressif et réceptif**

		<i>ÉPE-AD initiale : février 2009</i>				
Échelles		A	B	C	Sous-échelle vocabulaire expressif (ÉPE-AD)	Sous-échelle vocabulaire réceptif (ÉPE-AD)
<b>ÉPE-AD modifiée : juin 2009</b>	<b>A</b>	,83 (227)	,66 (232)	,84 (237)	,81 (221)	,63 (222)
	<b>B</b>	,55 (227)	,75 (232)	,52 (237)	,53 (221)	,47 (222)
	<b>C</b>	,60 (227)	,63 (232)	,61 (237)	,61 (221)	,47 (222)
	<b>Sous-échelle vocabulaire expressif (ÉPE-AD)</b>	,79 (227)	,52 (232)	,82 (237)	,81 (221)	,58 (222)
	<b>Vocabulaire expressif (Gardner)</b>	,75 (227)	,53 (232)	,75 (237)	,76 (221)	,50 (222)
	<b>Vocabulaire réceptif (ÉVIP-R)</b>	,66 (227)	,56 (232)	,69 (237)	,68 (221)	,43 (222)

*Note* : L'ÉPE-AD modifiée ne comprend pas une sous-échelle de vocabulaire réceptif. Ces dernières ont été éliminées par le concepteur de l'outil dans sa version allongée de l'ÉPE-AD puisque faciles. L'analyse des données du projet Capacité d'apprentissage confirme que ces données étaient trop faciles et par conséquent, ne permettaient plus de discriminer entre les enfants.

### 2.5.3 Mesures auprès des parents

Les sondages de suivi auprès des parents permettent de déterminer à quel point l'environnement de l'enfant s'est modifié au cours des quatre mois précédant le sondage. Dans un premier temps, ils permettent de mesurer la fréquence des activités de littératie, la langue de ces activités et d'autres facteurs qui affecteront la préparation à la scolarisation des enfants et le développement de leur identité francophone. Dans un deuxième temps, ils permettent de vérifier de façon systématique les changements quant au mode de garde de l'enfant et la langue utilisée dans ce nouveau mode de garde pour que les analyses puissent en tenir compte. Dans un troisième temps, ils visent à prendre en note les changements dans la composition de la famille afin de déterminer si l'environnement langagier habituel de l'enfant à la maison en est affecté.

Les échelles utilisées dans les sondages des parents participant au projet Capacité d'apprentissage proviennent d'études effectuées auprès de populations francophones au Canada, telles que l'ELNEJ, l'Étude longitudinale du développement des enfants du Québec (ELDEQ; Institut de la statistique du Québec, 2003) et l'Enquête sur la vitalité des minorités de langue officielle (EVMLO, 2006). Les questions qui ont été sélectionnées pour les sondages du projet Capacité d'apprentissage sont celles pertinentes au groupe d'âge des enfants.

Le lecteur est prié de noter que seules les échelles incluses dans les analyses d'impact sont détaillées dans cette section. Il n'est donc pas étonnant que cette liste d'échelles diffère quelque peu de celles rapportées dans le *Rapport de référence* (Legault et al., 2014). Plusieurs autres construits ont été mesurés sans toutefois être retenus dans les analyses d'impact. La décision de

les inclure ou non est fondée sur les analyses préliminaires dont les résultats sont rapportés au chapitre 4.

### ***Caractéristiques sociodémographiques***

Les caractéristiques sociodémographiques des enfants et de leurs parents proviennent de questions posées dans le formulaire de consentement des parents (par exemple, le sexe de l'enfant et sa date de naissance) et du sondage de base. Les sondages de suivi ont permis d'apporter des modifications, au besoin, à ce portrait initial. Cette section identifie les variables sociodémographiques utilisées pour les analyses d'impact.

Des informations telles que le niveau d'éducation de la mère et celui du père, l'âge de la mère à la naissance de l'enfant et le revenu familial ont été recueillies auprès des parents lors du sondage de base. Le revenu de la famille a été demandé, par la suite, à deux reprises. Une question ouverte portant sur le revenu fut d'abord posée en février 2009. Devant le refus de plusieurs parents (24,6 %) de partager cette information, la même question a été à nouveau posée en octobre 2009, et ce, spécifiquement aux parents ayant refusé de répondre, mais cette fois avec des catégories de revenu incluses dans les choix de réponses.

Plusieurs questions visaient à établir le profil linguistique des familles participantes. Entre autres, le répondant devait identifier ses langues maternelles, sa connaissance des langues officielles ainsi que les langues parlées avec l'enfant et à la maison. Les mêmes questions étaient demandées pour le conjoint ou la conjointe. Il existe deux définitions courantes de la population francophone canadienne. Selon Statistique Canada,<sup>24</sup> on peut tout d'abord calculer la population qui déclare avoir le français comme langue maternelle, c'est-à-dire que le français est la première langue apprise et encore comprise. La deuxième façon est de calculer la « première langue officielle parlée » (PLOP), un score prenant en compte la connaissance des deux langues officielles, la langue maternelle et la langue la plus souvent parlée à la maison (Forgues et Landry, 2006).

Chaque sondage de suivi commence par une série de cinq questions adressées seulement aux parents ayant subi un changement dans la composition de leur famille depuis le dernier sondage complété. Les changements dans la composition de la famille par rapport aux réponses du sondage de base (p.ex. la taille du ménage, les familles monoparentales, les enfants plus vieux, les enfants jumeaux, les enfants plus jeunes) sont pris en considération dans les analyses d'impact.

Le statut d'immigrant des parents a été établi au moyen d'une question dans le sondage d'octobre 2009. Ce sondage a également permis de faire une mise à jour des données sociodémographiques du conjoint ou de la conjointe du répondant. Notons qu'au sondage de base, l'information obtenue définissait les caractéristiques du conjoint ou de la conjointe qui était le parent biologique de l'enfant, même si ce parent n'avait pas de contact avec l'enfant. Afin de clarifier les résultats, ces questions ont été reposerées en octobre 2009 pour obtenir de l'information sur le deuxième adulte qui habite dans la maisonnée avec l'enfant et qui est impliqué de façon active dans les soins de l'enfant.

---

<sup>24</sup> Définition de la population francophone disponible sur le site Web de Statistique Canada à [http://www.statcan.ca/francais/concepts/definitions/language\\_f.htm](http://www.statcan.ca/francais/concepts/definitions/language_f.htm).

### **Habitudes langagières**

Plusieurs questions au sujet des habitudes langagières sont incluses dans le sondage de base et les sondages de suivi. Les répondants choisissent parmi plusieurs options la catégorie la plus représentative de leur vécu (par ex. s'ils parlent anglais seulement, anglais et français, français plus que l'anglais, etc.). Ces catégories ont été combinées pour créer une échelle en cinq points représentant un continuum linguistique où un score de 1 signifie que le répondant parle « anglais seulement », un score de 3 signifie qu'il/elle parle le « français et l'anglais et/ou une autre langue », et un score de 5 signifie qu'il/elle parle le « français seulement ». Les analyses d'impact ont fait usage de ce continuum linguistique basé sur une échelle de cinq points.

Un premier continuum linguistique a été généré représentant la langue habituellement parlée par la mère à l'enfant. Un deuxième continuum linguistique a été calculé en prenant la moyenne des scores de quatre questions portant sur les langues parlées par l'enfant avec sa mère, avec son père, avec ses amis et sa fratrie à la maison, ainsi qu'à l'extérieur de la maison (alpha de Cronbach de 0,95).

Deux indices représentant la langue habituellement parlée au foyer ont été dérivés. Un premier indice fut dérivé en mariant la PLOP de la mère avec celle du père. Un deuxième indice fut créé en combinant la langue habituellement parlée par la mère à l'enfant avec la langue habituellement parlée par le père à son enfant. Ces deux indices ont servi à déterminer l'environnement langagier dans lequel baigne l'enfant à la maison. En outre, la PLOP a servi à déterminer si l'enfant habite un foyer endogame francophone, endogame anglophone ou exogame.

### **Processus familiaux**

Les informations collectées sur les processus familiaux proviennent du sondage de base. Certains de ces processus sont mesurés dans les sondages de suivi. En tout, cinq processus familiaux ont été mesurés au moyen de quatre échelles, notamment les pratiques parentales, le fonctionnement familial, la dépression et les activités d'alphabétisation.

L'échelle des *Pratiques parentales positives* comprend cinq items mesurant la fréquence de contacts positifs entre le parent et l'enfant (par ex., combien de fois le parent félicite l'enfant, rit avec lui ou lui donnant de l'affection). Les items sont notés sur une échelle en cinq points dont 1 indique « jamais » et 5 indique « plusieurs fois durant la journée ». La cohérence interne de l'échelle est acceptable à 0,66.

L'échelle des *Pratiques parentales autoritaires* comprend quatre items mesurant le degré de supervision et de discipline. Elle nous permet de découvrir si, entre autres, l'enfant respecte ou non les punitions qu'on lui impose ou si ces punitions varient selon l'humeur du parent. Les items sont notés sur une échelle en cinq points où 1 indique « jamais » et 5 indique « presque toujours ». La cohérence interne s'avère acceptable avec un alpha de Cronbach de 0,61.

L'échelle de *Fonctionnement familial*, composée de huit items, évalue la qualité des liens dans la famille. Le contenu des items porte sur l'ouverture émotionnelle, l'expression des sentiments et la convivialité à l'intérieur de la famille. Les choix de réponses pour les items de l'échelle passent de 1, soit « absolument en désaccord » à 4, soit « entièrement en accord ». L'échelle du fonctionnement familial démontre une excellente cohérence interne avec un alpha de Cronbach de 0,82.

L'échelle de *Dépression* mesure, par exemple, le nombre de fois que le répondant s'est senti déprimé ou malheureux, qu'il a pleuré ou s'est senti seul, ou qu'il ne jouissait pas de la vie. Les huit items de l'échelle s'étalent sur un score de quatre points, où 1 indique que le répondant s'est senti ainsi « rarement ou jamais » et 4 signifiant qu'il s'est senti ainsi « la plupart du temps ou tout le temps ». L'échelle de la dépression a une cohérence interne très élevée avec un alpha de Cronbach de 0,80.

L'échelle des *Activités de littératie* comprend cinq items mesurant la fréquence des activités de lecture, d'écriture et de numératie. Plus spécifiquement, les items mesurent la fréquence à laquelle les parents racontent des histoires sans livre à leur enfant, chantent des chansons avec lui, lui enseignent à écrire ou à tracer des lettres ou des chiffres, lui enseignent à lire des mots, et l'encouragent à utiliser des nombres dans ses activités quotidiennes. Ces items sont notés sur une échelle de 5 points, où 1 correspond à une activité qui n'a jamais été faite et 5 correspond à une activité faite plusieurs fois par jour.

Chacune des questions portant sur les activités d'alphabétisation est suivie d'une question vérifiant la langue utilisée lors de cette activité. Les répondants choisissent parmi plusieurs options la catégorie la plus représentative de leur vécu (par ex. s'ils parlent anglais seulement, anglais et français, français plus que l'anglais, etc.). Ces catégories ont été combinées pour créer un score en cinq points représentant un continuum linguistique *pour cette activité* où un score de 1 signifie que l'activité se fait en « anglais seulement », un score de 3 signifie que l'activité se fait en « français et anglais et/ou une autre langue », et un score de 5 signifie que l'activité se fait en « français seulement ». Une échelle des *Langues des activités de littératie* est obtenue en faisant la moyenne des scores obtenus à chaque question. La cohérence interne de cette échelle est très élevée avec un alpha de Cronbach variant entre 0,91 et 0,95.

Les questions permettant de construire l'échelle des Activités de littératie ainsi que l'échelle des Langues des activités de littératie ont été posées à chaque sondage, incluant le sondage de base. La cohérence interne pour ces échelles s'avère assez acceptable à chacune des périodes de sondage pour permettre d'avoir confiance dans l'homogénéité de ces échelles (voir le tableau 2.10).

**Tableau 2.10 : Alpha de Cronbach de l'échelle des activités de littératie et de l'échelle des langues utilisées lors des activités de littératie à chaque période de sondage**

	<i>Alpha de Cronbach [<math>\alpha</math> (n)]</i>						
	oct. 2007	fév. 2008	juin 2008	oct. 2008	fév. 2009	juin 2009	oct. 2009
<b>Activités de littératie</b>	,59	,54	,54	,60	,64	,67	,59
<b>Langues utilisées lors des activités</b>	,95	,94	,94	,94	,91	,91	,93

### ***Dimension identitaire***

Le sondage de base comprenait plusieurs questions portant sur la *Vitalité subjective du français dans la communauté francophone* (ci-après nommée Vitalité), *l'Engagement envers la culture francophone* et le *Sentiment d'appartenance aux communautés linguistiques*. Parmi toutes ces dimensions identitaires, seul le sentiment d'appartenance à la culture francophone s'est avéré intéressant dans les analyses d'impact. Il s'agit d'une question visant à déterminer le

groupe culturel auquel les parents s'identifient, soit au groupe francophone, anglophone, aux deux ou à un autre groupe.

#### **2.5.4 Connaissances, attitudes et croyances des parents**

Les sondages pré et post Ateliers familles contiennent une série de questions visant à faire ressortir la nature et la portée des changements au niveau des perceptions et des comportements des parents du groupe programme. Pour les besoins de l'évaluation du programme Ateliers familles, plusieurs mesures ont été créées à partir des questions utilisées dans les études portant sur l'alphabétisation familiale menées par le Centre interdisciplinaire de recherche sur la citoyenneté et les minorités (CIRCEM) (LeTouzé, 2006) et par la Coalition francophone pour l'alphabétisation et la formation de base en Ontario dans le cadre de leur étude sur l'impact de l'alphabétisation familiale sur les familles vivant en milieu minoritaire (Benoît, s.d.). Des mesures ont aussi été créées à partir du questionnaire post-intervention du programme de transition à l'école provenant du Conseil du district scolaire 3 à Edmundston (Gauthier St-Onge, s.d.) et à partir de l'ELNEJ (Statistique Canada, 2006). Enfin, l'approche rétrospective utilisée pour créer les questions a été inspirée d'un sondage rétrospectif créé par Lamb et Tschillard (2005).

Le choix des construits pour étudier les impacts du programme des Ateliers familles s'inspire en partie du modèle de l'action raisonnée de Fishbein et Ajzen (pour une introduction au modèle voir Brigham, 1991). Ce modèle met en relation les croyances et les attitudes avec la prédiction des comportements. Selon le modèle, les croyances des gens relatives à un comportement influencent à la fois leurs attitudes envers ce comportement et leurs perceptions de contrôle (il s'agit d'un concept similaire à celui des sentiments d'efficacité). Les attitudes et les perceptions de contrôle influencent, à leur tour, l'intention d'effectuer le comportement. Enfin, l'intention détermine la probabilité que le comportement sera fait. Dans le projet Capacité d'apprentissage, nous avons mesuré les croyances et les attitudes des participants face aux dimensions ciblées par les Ateliers familles ainsi que leurs sentiments d'efficacité. Nous avons également mesuré les connaissances des participants envers les principaux thèmes abordés dans les Ateliers familles. Bien qu'il ne soit pas inclus dans le modèle de l'action raisonnée, le niveau de connaissance affecte également les attitudes des gens face à une action. Enfin, une échelle de modelage a été ajoutée pour capter l'importance des activités d'alphabétisation effectuées par les parents sur les attitudes et comportements des enfants. En effet, plusieurs études ont démontré l'effet positif d'un environnement riche en littératie familiale, tel que la présence de matériel écrit ou pour écrire, ainsi que des comportements parentaux valorisant la lecture ou l'écriture, comme de lire ou d'écrire devant les enfants (Dionne, 2007; Sénéchal & Lefèvre, 2002; Jordan, Snow, & Porche, 2000). Les quatre construits propres aux sondages pré- et post-ateliers sont détaillés dans les paragraphes suivants.

Quatre questions mesurent certaines *croyances et attitudes* des parents qui sont abordées au cours des Ateliers familles. Ces croyances correspondent à des thèmes majeurs du programme, notamment: le parent comme premier éducateur de l'enfant, la coopération entre l'éducatrice et les parents, l'environnement culturel francophone à la maison, et l'âge propice pour lire à l'enfant. Le répondant indique jusqu'à quel point il est en accord avec chacun des énoncés en utilisant une échelle en 4 points où 1 signifie « absolument en désaccord » et 4 signifie « absolument en accord ». Ces questions ont été posées dans les sondages pré- et post-ateliers.

Notons toutefois que ces questions ne forment pas, à proprement parler, une échelle et que, par conséquent, les informations collectées sont de nature descriptive.

L'échelle de *Perception d'auto-efficacité* évalue la perception qu'ont les parents de leur capacité à effectuer certaines actions discutées lors des Ateliers familles. Ainsi, on demande aux parents à quel point ils sont confiants quant à leur capacité de : gérer les émotions et les comportements de leur enfant, parler en français avec lui, l'aider dans ses apprentissages, et l'aider dans le développement d'une identité culturelle francophone. Chacun de ces aspects fait l'objet de deux questions. La première se veut une rétrospective demandant aux parents à quel point ils se sentaient confiants *avant* les Ateliers familles. La deuxième mesure plutôt l'état de confiance des parents *après* la fin des Ateliers familles (soit au moment du sondage). Ces items sont notés sur une échelle en 5 points, où 1 indique « peu ou pas de confiance » et 5 indique « une confiance complète ». La cohérence interne s'avère bonne avec un alpha de Cronbach de 0,72 pour la sous-échelle mesurant la perception rétrospective du sentiment d'auto-efficacité et de 0,62 pour la sous-échelle mesurant la perception du sentiment d'auto-efficacité suivant les ateliers.

L'échelle des *Connaissances* évalue le niveau de connaissance des parents par rapport à cinq des aspects abordés lors des Ateliers familles, dont le développement des enfants, les façons de faire pour aider à la maison le développement et les apprentissages de l'enfant, les moyens d'aider l'enfant à être prêt pour l'école, le travail d'une éducatrice en garderie, et les services ou ressources offerts en français dans la communauté. Comme pour l'échelle de perception d'auto-efficacité, chacun de ces aspects a fait l'objet de deux questions : l'une rétrospective (Comment décrivez-vous votre niveau de connaissance des sujets *avant* les Ateliers familles?) et l'autre suivant les Ateliers familles (Comment décrivez-vous votre niveau de connaissance des sujets *maintenant, après* les Ateliers familles?).<sup>25</sup> Ces items sont notés sur une échelle en 5 points, où 1 indique « une absence de connaissance » et 5 indique « un excellent niveau de connaissance ». La cohérence interne s'avère acceptable avec un alpha de Cronbach de 0,79 pour la sous-échelle mesurant la perception du niveau de connaissance avant les ateliers et de 0,66 pour la sous-échelle mesurant la perception du niveau de connaissance suivant les ateliers.

L'échelle de *Modelage des activités de littératie* comprend cinq questions basées sur le modèle des activités de littératie. Elle vise à mesurer l'exemple que donne le parent à son enfant quant à l'importance de la lecture et de l'écriture. Les items portent, entre autres, sur la fréquence à laquelle les parents lisent et écrivent devant leur enfant, ainsi que la fréquence à laquelle les parents laissent à la disposition de leur enfant du papier et des crayons. Les réponses aux questions sont notées sur une échelle en 4 points où 1 indique « jamais » et 4 indique « plusieurs fois par jour ». Cette échelle fut remplie par les parents du groupe programme lors du sondage pré-intervention et du sondage post-intervention administré 3 semaines suivant la fin des ateliers (soit en mars 2008 pour la première cohorte et en mars 2009 pour la deuxième cohorte) et un dernier sondage administré 7 mois après la fin des ateliers (soit en octobre 2008 pour la première cohorte et en octobre 2009 pour la deuxième cohorte). La cohérence interne s'avère acceptable avec un alpha de Cronbach de 0,73 et de 0,68 observé respectivement lors de chacune de ces administrations.

---

<sup>25</sup> Pour une discussion éclairée sur les avantages d'utiliser la rétrospective dans le cadre de formations, voir Lamb et Tschillard (2005).

### 2.5.5 Dosage

Le nombre d'heures de garde par semaine est une variable importante à prendre en compte pour mesurer le « dosage » du programme que reçoivent les enfants en garderie. Pour recueillir cette information, la SRSA a utilisé le registre des présences et absences que les éducatrices remplissaient déjà dans le cadre de leurs tâches, incluant l'heure d'arrivée et de départ des enfants. Le nombre d'heures que l'enfant a passé en garderie est ensuite compilé de façon hebdomadaire et envoyé de façon mensuelle au bureau d'Ottawa de la SRSA. Pour les fins des analyses d'impact d'un an, les heures passées en garderie ont été analysées en prenant la moyenne d'heures passées en garderie par semaine durant les quatre mois précédant immédiatement l'évaluation des enfants (voir la Section 5.1.3). Pour les analyses d'impact de la deuxième année, la moyenne d'heures passées par semaine durant les huit premiers mois post-intervention figure dans les analyses (voir la Section 6.2).

En ce qui a trait aux Ateliers familles, l'assiduité des parents, des enfants et des autres adultes présents a été compilée lors de chaque atelier. Le dosage aux ateliers est calculé en prenant la somme d'ateliers suivis par au moins un adulte.

### 2.5.6 Indices de la mise en œuvre du volet en garderie

L'importance de tenir compte du degré de mise en œuvre d'une intervention dans l'interprétation des effets du programme est mise en évidence dans un nombre grandissant d'études (Charlebois et al., 2004; Conduct Problem Prevention Research Group, 1999; Dane et Schneider, 1998; Durlak et Dupré, 2008). Sur ce point, Durlak et Dupré (2008) ont conclu à partir des données de cinq méta-analyses que la taille de l'effet d'une intervention est de deux à trois fois plus grande lorsque le degré de mise en œuvre est pris en considération lors des analyses d'impacts. De plus, la prise en considération de l'intégrité et de la qualité de la mise en œuvre du programme en garderie permet de mieux comprendre et nuancer l'interprétation des résultats des analyses d'impact. Il s'en suit que six indices a été incorporé dans les analyses d'impact dont la fidélité structurelle, la fidélité du contenu, la qualité structurelle, la qualité de la lecture, la qualité éducative, et la sensibilité des éducatrices. Tous ces indices ont été créés à partir des notes d'observations dans les garderies programmes et les garderies témoins. Pour les fins des analyses d'impact, ces notes d'observations ont été transformées en sous-échelles de l'ÉÉEP-R.<sup>26</sup>

Un des premiers éléments utilisés dans les analyses d'impact était la fidélité de la mise en œuvre du programme. Il s'agit de s'assurer que l'ensemble des éléments du programme ont été mis en place dans les garderies. Une grille d'observation a été développée afin de vérifier la présence de certains éléments spécifiques au programme testé (comme la présence d'une grille de planification hebdomadaire visible pour les parents ou encore la présence d'une routine imagée décrivant aux enfants le déroulement de la journée). Les observatrices devaient également prendre en note le déroulement des activités et la réaction des enfants lors des différentes activités. Ces notes d'observations ont permis d'élaborer deux indices. Un premier, nommée *fidélité structurelle*, porte sur la présence de certains éléments dans l'environnement,

---

<sup>26</sup> Il a été décidé au début de projet Capacité d'apprentissage de ne pas remplir directement la grille de l'ÉÉEP-R, puisque ce type d'observation pouvait être perçu comme trop intrusif et nuire à l'obtention de la pleine collaboration des garderies programme et témoins. Il a donc été décidé que les observatrices prendraient des notes sur les différents éléments de l'ÉÉEP-R pour les sous-catégories ciblées.

comme les affiches image-mot ou encore la routine imagée. Le deuxième indice, la **fidélité de contenu**, porte sur l'intégration des éléments du programme dans la programmation du milieu de garde, comme de faire la lecture aux enfants lors du cercle d'amis ou de faire avec eux des activités de raisonnement.

Les notes d'observations ont également permis d'obtenir des renseignements sur la qualité des différents milieux de garde des enfants. Spécifiquement, ces notes d'observations ont permis de remplir huit sous-échelles de l'ÉÉEP-R. L'ÉÉEP-R est un outil validé pour évaluer la qualité des services de garde telle que définie par les spécialistes en éducation à la petite enfance de la *National Association for the Education of Young Children* (NAEYC). D'ailleurs, cet outil a été utilisé à plusieurs reprises dans des projets francophones pour évaluer la qualité des services de garde (par ex., Japel, 2005).

Aux fins des analyses d'impacts, quatre sous-échelles de l'ÉÉEP-R ont été utilisées pour former un premier indice de la qualité des milieux de garde. Le premier indice, la **qualité structurelle**, concerne la qualité globale de l'environnement des milieux de garde. Il comprend les sous-échelles de l'ÉÉEP-R (remplis à partir des notes d'observations) « Espaces physiques », « Santé et sécurité » et « Accueil et Départ ». Le deuxième indice, **qualité de lecture**, comprend uniquement les items de la sous-échelle « Livres et images », étant donnée l'importance toute particulière de la lecture pour l'acquisition de vocabulaires chez les enfants. Le troisième indice, **qualité éducative**, porte une attention particulière aux activités permettant d'encourager la communication chez les enfants et d'enrichir leur vocabulaire. Cet indice regroupe les sous-échelles de l'ÉÉEP-R « Encourage l'enfant à communiquer en français », « Soutien à la communication », et « Utilisation du langage pour développer les habiletés de raisonnement ». Enfin, le quatrième indice tient compte de l'influence particulièrement importante de la **sensibilité de l'éducatrice** pour le développement de l'enfant. À cette fin, les notes d'observations ont servi à remplir la sous-échelle « Interaction entre le personnel et les enfants ».

L'interprétation des indices se fait comme suit : les deux indices de fidélité indiquent la proportion d'éléments du programme mis en place tandis que les quatre indices de qualité reflètent à quel point les éléments centraux du programme sont mis en place selon une échelle de 7-points où 1 correspond à des soins se situant bien en deçà des exigences élémentaires d'un service de garde et 7 correspond à des soins personnalisés et de qualité supérieure (Harms, Clifford & Cryier, 1998). Ces six indices ont servi à examiner, lors des analyses d'impact, le lien entre, d'une part, la fidélité et la qualité du programme en milieu de garde et, d'autre part, le développement et la préparation à la scolarisation des enfants au cours de la première année de la prestation du programme. En ce qui concerne les analyses portant sur la deuxième année de la prestation du programme, les impacts du programme sur les enfants ont été vérifiés au moyen de deux indices globaux de fidélité et de qualité. Ces indices globaux correspondent à la moyenne des indices de fidélité et la moyenne des indices de qualité du programme.

Ces six indices de fidélité et de qualité ainsi que les deux indices globaux de fidélité et de qualité ont servi à valider les résultats des principales analyses d'impacts du programme en garderie (voir Chapitres 5 et 6). Le lecteur est prié de noter que seuls les indices ayant servis dans les analyses d'impacts sont abordés dans le présent rapport. Une description complète des modalités et des outils utilisés pour étudier l'intégrité de la mise en œuvre du programme en garderie se retrouve dans le *Rapport de mise en œuvre du projet* (Bérubé et al., 2014).



## 2.5.7 Étude de la mise en œuvre des Ateliers familles

Les observations des Ateliers familles ont permis de recueillir des données pertinentes sur la prestation du programme. La pertinence des données est déterminée en fonction des objectifs du projet Capacité d'apprentissage et de ceux du programme Ateliers familles. Une grille semi-structurée a été créée pour guider les observations effectuées sur le terrain (quant aux éléments ou aux caractéristiques à observer), bien que les observations elles-mêmes prennent la forme de notes informelles. Les renseignements recueillis permettent d'examiner la fidélité et la qualité avec lesquelles le programme était livré à travers les sites afin de voir si les différences dans certains aspects du déroulement peuvent expliquer la présence ou l'absence de changements chez les parents.

Plusieurs documents ont servi de base à la création des différentes sections de la grille d'observation. Ainsi, les questions portant sur les aspects plus généraux de la livraison des ateliers sont basées sur le cadre d'observation proposé par Merriam (1988). Les questions portant sur le contenu du programme, son organisation et le style d'animation de l'intervenante se fondent sur les recommandations proposées par la Fédération canadienne pour l'alphabétisation en français (FCAF, 2007). Les questions concernant la participation des parents et le style d'animation de l'intervenante sont basées sur celles proposées dans le « *Guide for Using the Classroom Observation Tool* » (Baylor College of Medicine, 2001). Enfin, les questions concernant la présentation claire du but de la leçon/atelier s'inspirent du guide « *Inside the Classroom : Observation and Analytic Protocol* » (Horizon Research, 2000).

La fidélité du programme a été saisie par l'entremise des notes d'observations prises lors de chacun des ateliers. L'indice reflète le nombre d'éléments obligatoires couverts lors de chaque atelier, et ce, pour les 11 séries d'ateliers. En moyenne, on observe une excellente fidélité quant à la **couverture du contenu** (96% des éléments obligatoires ont été couverts) pour l'ensemble des séries de dix ateliers offertes. Par ailleurs, toutes les communautés ont couvert au moins 90 % du matériel, ce qui représente aussi une excellente couverture. Ensuite, la qualité du programme a été évaluée à partir de divers indices dont la **qualité de l'animation** de l'intervenante et le niveau de satisfaction des participants de la **durée des séances**. Les données utilisées pour élaborer ces indices ont été recueillies par l'entremise d'un sondage anonyme administré lors du dernier atelier. La qualité de programme a également été évaluée au moyen d'observations des **réactions positives et négatives** des participants lors de chaque atelier. Enfin, **un indice de qualité global** a été calculé à partir de la moyenne des scores obtenus aux trois indices de couverture du contenu, qualité de l'animation et satisfaction avec la durée des séances. Les résultats des analyses effectuées afin d'investiguer l'impact de ces cinq indices sur les résultantes des parents sont présentés à la section 5.2.3.

## 2.6 HYPOTHÈSES À L'ÉTUDE

### 2.6.1 L'impact du volet en garderie sur les dimensions linguistiques des enfants

Le programme en garderie vise particulièrement la préparation à la scolarisation, l'enrichissement des habiletés langagières des enfants, et le développement d'une identité culturelle francophone. C'est donc sur ces dimensions que nous anticipons observer des effets plus prononcés du programme. Les hypothèses émises sont :

1. Les enfants du groupe programme auront des scores plus élevés dans trois des quatre domaines mesurant la préparation à la scolarisation, dont en langue et communication, en habiletés cognitives, et en conscience de soi et de l'environnement, comparativement aux enfants des groupes témoins. Aucune différence entre le groupe programme et les groupes témoins n'est anticipée au niveau du développement physique et moteur puisque le programme testé ne cible pas cette dimension.

Cette hypothèse vise à tester directement la question de recherche principale. La validité interne de l'étude est assurée par l'inclusion du groupe témoin en garderie puisque l'influence d'une garderie en installation sur le développement des enfants constitue en soi un traitement. Elle est également assurée par l'ajout du deuxième groupe témoin qui vise à tenir compte de l'influence d'un milieu de garde informel sur le développement des habiletés langagières en français de l'enfant.

Enfin, cette hypothèse tient compte des autres conditions existantes dans la communauté pouvant affecter le développement des enfants. Par exemple, l'environnement communautaire joue un rôle important dans le développement de l'enfant dans la mesure où il rend disponible des ressources (p. ex., parcs, pataugeuses, pistes cyclables) et des services (p. ex., librairies, bibliothèques, cours de natation, etc.).

2. Les enfants du groupe programme auront une meilleure performance sur les mesures d'habiletés langagières comparativement aux enfants des groupes témoins.
3. Le programme testé aura un impact plus prononcé sur les habiletés langagières des enfants du groupe programme issus de foyers exogames ou de foyers endogames anglophones.
4. L'ampleur de l'impact du programme en garderie sur les dimensions de la préparation à la scolarisation des enfants sera influencée par le niveau d'exposition de ces derniers au programme. Ainsi, une plus grande exposition au programme testé sera associée à une meilleure performance des enfants du groupe programme dans trois des quatre domaines mesurant la préparation à la scolarisation, dont en langue et communication, en habiletés cognitives, et en conscience de soi et de l'environnement, comparativement aux enfants des groupes témoins. Aucune différence entre le groupe programme et les groupes témoins n'est anticipée au niveau du développement physique et moteur puisque le programme testé ne cible pas cette dimension.
5. Les garderies programmes seront caractérisées par des indices de fidélité et de qualité du programme supérieurs à ceux des garderies témoins.
6. L'ampleur de l'impact du programme en garderie sur les dimensions de la préparation à la scolarisation et sur les habiletés langagières variera en fonction du degré de fidélité et de qualité du programme testé.

## **2.6.2 L'impact du volet en garderie sur les dimensions identitaire et culturelle des enfants**

Le programme testé vise à influencer les divers contextes nécessaires au développement des dimensions identitaire et culturelle des jeunes enfants. L'importance de mesurer l'identité culturelle se fonde sur des recherches soulignant que l'exposition des enfants au français dans plusieurs sphères de leur vie fortifie leur sentiment d'identité et d'appartenance à la communauté francophone (Landry & Allard, 2000). Cependant, le concept d'identité culturelle est mesuré à l'adolescence dans ces études. La mesure de ce concept dans ces études correspond bien à la théorie sur le soi d'Erik Erikson (1994) selon laquelle la formation de l'identité débute vers l'âge de 12 ans pour se terminer autour de 25 ans. Ainsi, une mesure directe de l'identité culturelle auprès d'enfants en très bas âge s'avère impossible.

Par contre, une mesure du degré *d'exposition* des enfants au français est possible via les parents. En effet, les parents sont les mieux placés pour nous informer de l'environnement langagier des milieux que l'enfant fréquente et qui sont d'intérêt pour le projet Capacité d'apprentissage, soit la famille, le milieu de garde de jour et l'accès aux services et ressources de langue française. Les informations collectées serviront de variables contrôles dans les analyses d'impact afin de mieux dégager les effets du programme testé sur le développement des habiletés langagières des jeunes enfants.

## **2.6.3 L'impact du volet Ateliers familles sur les attitudes et comportements des parents**

Le contenu du programme des Ateliers familles met particulièrement l'accent sur quatre aspects :

- le parent comme premier éducateur de son enfant;
- les activités de pré-lecture, pré-écriture et pré-numératie susceptibles de favoriser la préparation à la scolarisation de l'enfant;
- l'importance de maximiser l'exposition de l'enfant au français; et
- la collaboration parent-éducatrice dans les apprentissages de l'enfant.

C'est donc sur ces aspects que nous anticipons observer des effets plus prononcés du programme. Les hypothèses émises sont les suivantes. Suite aux Ateliers familles et comparativement aux parents des groupes témoins, les parents du groupe programme rapporteront:

- 1) effectuer significativement plus fréquemment différentes activités de littératie avec leur enfant;
- 2) parler significativement plus souvent en français avec leur enfant durant les activités de littératie.

De plus, nous anticipons que les parents du groupe programme rapporteront, suite aux ateliers :

- 1) un niveau de connaissance significativement plus élevé des dimensions développementales des enfants, des moyens pour les aider à être prêt pour l'école, et des services ou ressources communautaires de langue française;

- 2) des sentiments d'efficacité significativement supérieurs, notamment au niveau des moyens pour aider leur enfant à être prêt pour l'école;
- 3) une fréquence significativement accrue d'occasions de modelage des activités de littératie;
- 4) un changement significatif dans leurs croyances, particulièrement pour les éléments visés par le volet Ateliers familles, notamment : le parent comme premier éducateur de son enfant, la coopération entre éducatrice et parent dans les apprentissages de son enfant, l'importance de l'environnement culturel francophone à la maison, et l'âge propice pour lire à son enfant.

Dans le prochain chapitre, nous détaillons la stratégie analytique adoptée dans les analyses d'impact pour la première et la deuxième année du programme.

### 3. Stratégie d'analyse

Le présent chapitre traite de la logique sous-jacente aux analyses décrites aux chapitres 5 et 6. Ici, l'exposition au programme testé est conçue comme un continuum allant de l'intensité la plus faible à l'intensité la plus forte. L'intensité d'exposition la plus forte possible serait reçue par un enfant qui fréquente à temps plein une garderie qui met en œuvre le programme précisément tel qu'il a été conçu et avec le plus haut degré de qualité. Dans le texte qui suit, nous expliquons comment l'intensité du traitement a été définie et la façon dont ce concept a été utilisé pour estimer les impacts du programme testé. Les groupes expérimentaux ainsi que des indicateurs plus précis de l'intensité du traitement ont été utilisés à cette fin.

#### 3.1 CONCEPTUALISATION DU TRAITEMENT ET DU DOSAGE

Le modèle le plus simple possible pour capter un effet du traitement contiendrait deux groupes de participants: le groupe « traité » et le groupe « non traité ». Ce type de modèle est le plus valide lorsque la distinction entre un groupe traité et un groupe non traité est absolue, c'est-à-dire que le groupe non traité ne reçoit absolument aucun traitement, alors que tous les membres du groupe traité reçoivent un traitement équivalent (p. ex., un dosage ou une « exposition » équivalent(e)). Dans le cas d'une étude effectuée sur le terrain, des distinctions aussi nettes sont plutôt rares, et le projet Capacité d'apprentissage ne fait pas exception à cette règle. L'intensité du traitement reçu par les participants variait en fonction d'au moins deux dimensions: le degré d'exposition (dosage) au traitement et la qualité du traitement. Nous décrirons maintenant de façon plus détaillée les deux sources de variation dans l'intensité du traitement.

##### 3.1.1 Exposition hétérogène au traitement

Lors de leur inscription au projet, les participants ont été assignés à l'un de trois groupes: le groupe programme en garderie, le groupe témoin en garderie ou le groupe témoin hors garderie. Il n'est peut-être pas étonnant de réaliser que ces assignations aient parfois varié au fil du temps en fonction de l'évolution des choix des parents quant au mode de garde de leur enfant. Par exemple, un pourcentage relativement faible de parents (voir le chapitre 2) a décidé de changer de garderie entre les périodes de test. Par conséquent, un enfant qui faisait partie du groupe programme en garderie pendant les deux premières périodes de test pouvait se trouver dans le groupe témoin en garderie ou dans le groupe témoin hors garderie pour les troisième et quatrième périodes.

Même si les participants n'ont pas migré d'un groupe à un autre, leur temps passé en garderie pouvait varier. Les deux groupes en garderie ne différaient pas quant à leur exposition moyenne à l'environnement d'une garderie durant chacune des sept périodes de quatre mois précédant les évaluations de l'ÉPE-AD (voir les résultats figurant à la section 4.4). Une conséquence pratique de cette équivalence est que cette variable ne constitue pas une menace évidente à la validité de nos estimations sur les effets programme. Néanmoins, il est intéressant de se demander si le degré d'exposition à une condition de traitement a de l'importance et si les effets programme, le cas échéant, ont une interaction synergétique avec le degré d'exposition. En d'autres mots, est-ce

qu'une heure passée dans une garderie programme produit de meilleurs résultats qu'une période de temps équivalente passée dans une garderie témoin? Nous avons traité cette question en insérant les variables liées au degré d'exposition au traitement dans les analyses d'impacts.

### **3.1.2 Évaluation de l'effet du dosage ou de l'exposition**

Deux stratégies ont été adoptées pour gérer la migration des participants au fil du temps. La première est brute, mais sa simplicité lui confère l'avantage de faciliter la comparaison des groupes (c.-à-d. traités c. non traités). La seconde stratégie définit le traitement avec plus de précision en indiquant les heures d'exposition à un milieu de garde donné (c.-à-d. la moyenne d'heures passées en garderie par semaine). Cette dernière stratégie est un peu plus complexe, mais elle possède l'avantage de représenter le degré d'exposition aux deux conditions de traitement en garderie. Ces deux techniques servant à représenter l'hétérogénéité de l'exposition au traitement font l'objet de deux séries d'analyses indépendantes, dont nous traitons plus en détails ci-après.

La première série d'analyses repose sur la définition brute de l'exposition au traitement, où des codes fictifs ont servi à représenter l'appartenance à un groupe expérimental. Pour toute période de temps donnée, un participant codé comme appartenant à un certain groupe expérimental était considéré comme ayant reçu entièrement le traitement associé à ce groupe. L'appartenance à un groupe pouvait changer au fil du temps, mais nous avons maximisé l'intégrité des groupes expérimentaux en exigeant que les participants soient exposés à leur nouveau mode de garde durant une période de temps minimale avant de reconnaître ce changement dans les analyses. Plus précisément, les deux conditions suivantes ont été posées. Premièrement, on considérait qu'un enfant avait changé de groupe uniquement si le changement était survenu plus d'un mois avant l'évaluation. Deuxièmement, les changements de modalités de garde effectués au cours des mois d'été n'ont été considérés valides que si ces nouveaux arrangements avaient persisté pendant quelques mois après la rentrée scolaire. Un participant qui rapportait son départ d'une garderie programme et son inscription subséquente dans une garderie témoin pendant l'été, par exemple, était considéré comme membre du groupe programme en garderie pour l'évaluation automnale. Pour les fins d'analyses, le changement de groupe n'entrait en vigueur que pour l'évaluation hivernale puisque les enfants avaient alors été exposés au programme pendant quelques mois durant lesquels le programme de garderie habituel était en place. Les deux mesures de vocabulaire normalisées étaient exceptionnelles en ce sens qu'elles n'ont été prises qu'une seule fois chacune. Peu importe les résultantes mesurées, nous avons évalué dans tous les cas le biais pouvant résulter d'un changement de groupe.

Dans la seconde série d'analyses visant les données de la première année, le traitement a été défini en utilisant les heures d'exposition à un milieu de garde, conjointement avec les caractéristiques de ce milieu (c.-à-d. le groupe programme en garderie ou le groupe témoin en garderie). Nous avons défini les heures d'exposition comme étant la moyenne d'heures qu'un enfant passe en garderie chaque semaine durant la période de quatre mois précédant une évaluation particulière. L'exception était la période de base pour laquelle on utilisait seulement deux mois (septembre et octobre 2007). L'inclusion de cette variable dans les analyses nous a permis de définir l'exposition à notre programme en garderie de manière plus précise. En croisant la variable des heures d'exposition avec notre variable de groupement (c.-à-d. en spécifiant un terme d'interaction), nous avons pu: a) estimer l'effet moyen du traitement associé à un nombre donné d'heures d'exposition par semaine; et b) tester si l'effet du degré

d'exposition à la garderie varie en fonction du type de programme. Ce dernier test est simplement un prolongement de notre hypothèse de recherche fondamentale selon laquelle en présence d'un degré d'exposition équivalent, les participants du groupe programme seront avantagés quant aux trois domaines de préparation à la scolarisation. Cette conceptualisation plus précise du dosage accroît la puissance statistique de nos analyses de détecter les effets du programme. Elle nous offre aussi la possibilité de découvrir le degré d'exposition au programme requis pour provoquer l'effet souhaité.

Lors, de l'analyse de suivi des résultantes de la deuxième année, nous avons utilisé les données sur les heures d'exposition de façon quelque peu différente. À la fin de la première année, la plupart des enfants de l'étude (c.-à-d., les enfants vivant dans les trois communautés ontariennes) étaient inscrits à la prématernelle à temps partiel ou à temps plein.. Pour cette raison, les données sur la fréquentation des garderies qui ont été recueillies pendant la deuxième année du projet sont difficiles à interpréter en tant que prédicteurs des résultantes parce qu'elles ne reflètent pas adéquatement le temps passé dans un environnement de garde d'enfants de qualité. Les enfants inscrits à temps plein ou à temps partiel dans une école sont exposés à un environnement d'excellente qualité. Par conséquent, l'enfant qui passe quelques heures par semaine dans une garderie programme peut en réalité passer beaucoup d'heures à l'extérieur de la maison et faire l'expérience d'un traitement d'une intensité comparable.

Il est indubitable que les « heures d'exposition en garderie » ne constituent pas une variable importante dans la deuxième année du projet. Il serait techniquement possible d'effectuer une analyse de ce type à l'aide seulement de la communauté d'Edmundston (c.-à-d. qu'aucun enfant de cette communauté n'était inscrit à l'école), mais le nombre d'enfants inclus dans une telle analyse serait trop petit pour fournir de l'information utile. Nous avons plutôt traité les évaluations de la deuxième année en tant que tests de suivi quant à l'effet des heures d'exposition en garderie la première année. L'investissement nécessaire pour fournir des services de garde au cours de la deuxième année du projet était néanmoins important afin d'assurer le maintien des effets du programme pour les enfants qui n'étaient pas inscrits à l'école. Le rendement des enfants en question contribue aux effets du programme rapportés dans les chapitres 5 et 6. Autrement dit, les effets persistants des heures d'exposition au traitement ont été testés dans cette analyse.

Pour capter cette idée, nous avons calculé une variable représentant le nombre moyen d'heures par semaine qu'un enfant passe soit dans une garderie programme ou dans une garderie témoin pendant les huit premiers mois de la mise en œuvre du programme.<sup>27</sup> Cette variable a été croisée avec le groupe expérimental dans un modèle de régression prédisant les résultantes des enfants au bout de la deuxième année. Par conséquent, le groupe expérimental de la première année a été croisé avec le dosage en garderie de la première année pour tenter de prédire les résultantes des enfants pour la deuxième année. Si une année passée dans une garderie programme au cours de la première année se traduit par de meilleures résultantes, en moyenne, qu'une année passée dans une garderie témoin, alors on peut s'attendre à observer un effet bien plus significatif du dosage pour le groupe fréquentant une garderie programme que pour le

---

<sup>27</sup> Nous avons exclu les données relatives aux quatre mois précédant la quatrième évaluation puisqu'elles portaient sur des variations attribuables aux perturbations de la période estivale. La mesure la plus pure de l'effet moyen d'exposition au traitement repose sur l'information collectée au cours de l'année scolaire.

groupe fréquentant une garderie témoin au cours de la deuxième année, si cet effet est soit persistant, soit retardé.

### 3.1.3 Évaluation de l'effet de la fidélité/qualité du programme en garderie

Le mécanisme par lequel l'intervention est censée influencer les résultantes en matière de développement des enfants par l'entremise de la qualité du programme et de la fidélité de sa mise en œuvre. Dans ce sens, l'appartenance à un groupe expérimental est une approximation de la qualité et de la fidélité. La validité interne de l'étude repose fondamentalement sur la véracité de cet énoncé. Il s'ensuit qu'une condition requise pour observer un effet programme sur les résultantes ciblées est que les enfants inscrits dans les garderies où l'intervention a été mise en œuvre (groupe programme en garderie) vivent des expériences qui se comparent favorablement à celles des enfants faisant partie du groupe témoin (groupe témoin en garderie). De même, à l'intérieur de chaque groupe, la nature du programme offert d'une garderie à l'autre doit être la plus semblable que possible, c'est-à-dire qu'elle doit être cohérente au sein de toutes les garderies du même groupe. En d'autres mots, les garderies doivent être groupées de façon cohérente. Ces deux conditions ont été confirmées au moyen d'analyses qualitatives présentées dans le rapport *Capacité d'apprentissage dans les communautés francophones en situation minoritaire : Rapport de mise en œuvre du projet* (Bérubé et al., 2014) et dans une série d'analyses quantitatives préliminaires rapportées à la section 4.4.

L'utilisation des groupes expérimentaux pour estimer l'ampleur des effets du traitement constitue une simplification utile où les dimensions potentiellement continues (p.ex., la fidélité et la qualité) sont réduites à des catégories (c.-à-d. groupe programme en garderie et groupe témoin en garderie), dans le but de faire des comparaisons. Cette simplification est la méthode la plus appropriée lorsque les membres de chaque groupe sont très semblables les uns aux autres quant aux dimensions continues. Bien entendu, la nature du programme offert dans une garderie varie toujours un peu d'une garderie à l'autre, ce qui signifie que le recours aux groupes expérimentaux aboutit à une perte d'information (c.-à-d., la variabilité intra-groupe).

Nous avons vérifié si cette perte d'information, engendrée par l'emploi des groupes expérimentaux, était importante. Pour ce faire, nous avons procédé à une série d'analyses où la présence d'un intermédiaire, soit le groupe expérimental, a été éliminée et remplacée par des indicateurs continus de fidélité et de qualité en tant que prédicteurs des résultantes des enfants. Cette méthode optimise l'utilisation de l'information disponible et elle offre de meilleures possibilités de déceler les effets véritables du programme. Bref, cette méthode constitue une méthodologie de recherche plus efficace.

Nous voulions aussi déterminer si certains effets observés du groupe expérimental étaient attribuables au programme lui-même ou à d'autres caractéristiques des garderies programme. Nous avons testé cette idée en estimant les effets du groupe expérimental après avoir contrôlé statistiquement la fidélité et la qualité. Si c'est spécifiquement le programme testé qui est responsable des effets observés, alors les effets ajustés du groupe expérimental résultant de cette analyse ne devraient pas s'avérer statistiquement significatifs. La logique de cette analyse repose sur un test de médiation (Cohen et al., 2003, p. 457) où l'effet de l'appartenance à un groupe expérimental sur les résultantes est présumé s'exercer indirectement par la qualité ou la fidélité



du service.<sup>28</sup> Tout effet résiduel lié au groupe expérimental serait nécessairement attribuable à un autre facteur (p. ex., des différences non contrôlées entre les groupes au début du projet ou autre source de biais). L'inclusion des analyses en question dans le présent rapport en augmente substantiellement la longueur, mais elle améliore en revanche notre compréhension des résultats rapportés et notre confiance à leur égard.

Dans les analyses rapportées à la section 5.1.3, chaque type d'indicateur a été opérationnalisé à un niveau général et à un niveau plus détaillé. Deux estimations détaillées de fidélité ont été calculées selon l'adhérence aux éléments spécifiques à la structure du programme et au contenu du programme, respectivement. L'estimation générale de fidélité a été calculée simplement en faisant la moyenne de ces deux indicateurs. En ce qui concerne la qualité proprement dite, nous avons calculé les indices représentant les dimensions structurelles, pédagogique et de sensibilité des éducateurs. Encore une fois, un indice de qualité global a été calculé en faisant la moyenne de ces trois indicateurs de qualité. Un quatrième indicateur a saisi la nature des activités d'alphabétisation en salle de classe. Cet indicateur a été traité séparément des autres, en raison de son importance théorique et empirique.

Les indices de fidélité et de qualité ont été insérés dans les analyses de la façon suivante. Nous avons d'abord mené une série d'analyses où les indices de fidélité et de qualité, plutôt que l'appartenance au groupe expérimental, ont servi d'indicateurs de l'exposition au traitement. Le but cette analyse était de vérifier si: a) les résultats obtenus en employant la simple définition du traitement seraient reproduits; et b) l'analyse plus sensible révèle des effets plus marqués, plus persistants, ou les deux. Dans une seconde série d'analyses, nous avons réintroduit l'appartenance au groupe expérimental comme indicateur de l'exposition au traitement, tout en effectuant un contrôle statistique des indices de fidélité et de qualité. Le but de ce deuxième test était de vérifier si la façon plus simple de définir le traitement est redondante avec la fidélité/qualité pour expliquer la fluctuation des résultantes pour les enfants. Si les indices de fidélité et de qualité capturent adéquatement le cheminement par lequel l'appartenance au groupe expérimental exerce ses effets sur les résultantes pour les enfants, nous nous attendrions alors à ce que l'ampleur des estimations des effets du traitement diminue considérablement lorsque la qualité et la fidélité des garderies sont contrôlées lors des analyses.<sup>29</sup> Si l'on obtenait un tel résultat, cela fournirait une justification de plus pour attester que les effets du traitement estimés rapportés ici ne sont pas qu'un artefact de la méthodologie, mais plutôt qu'ils reflètent fidèlement l'impact du programme en garderie testé.

## 3.2 APPROCHE STATISTIQUE

Dans cette section, nous présentons la stratégie d'analyse qui a servi à modéliser l'impact de l'intervention. Nous commençons par introduire la notion d'échantillonnage multiniveaux et d'observations groupées (voir la section 3.2.1). Nous décrivons par la suite les stratégies de spécification des modèles utilisées pour estimer les effets du traitement dans le cadre d'analyses

---

<sup>28</sup> Un test de médiation approprié comporte normalement des séries de tests plus complexes (Baron & Kenny, 1986 ; Frazier, Tix & Barron, 2004). L'une des conditions essentielles à la démonstration de la médiation est que l'effet direct de la variable X (dans ce cas, le traitement) doit être éliminé lors du contrôle du médiateur M (dans ce cas, la qualité/fidélité).

<sup>29</sup> Une exception évidente est le cas où les deux variables en question participent à un effet de suppression (Cohen, Cohen, West & Aiken, 2003, pp.77-78, 457-458). Les effets de suppression dans le contexte de ces analyses sont abordés lorsqu'ils ont une incidence sur l'interprétation des résultats.

diverses (voir les sections 3.2.2, 3.2.3 et 3.2.4). Un compte rendu détaillé des spécifications du modèle de régression est présenté aux chapitres 5 et 6 avec les résultats correspondants.

### **3.2.1 Modélisation linéaire hiérarchique (MLH)**

Les données ont été analysées à l'aide de modèles de régression linéaires. Un postulat fondamental des modèles de régression standards veut que chaque point d'observation ou point de données présent dans l'analyse ait été observé de façon indépendante. Ce postulat n'est pas respecté lorsque les unités d'échantillonnage (p. ex., les garderies ou les enfants) contribuent à de multiples observations d'un ensemble de données. Dans ce cas, on dit des observations faites par une même unité d'échantillonnage qu'elles sont imbriquées ou groupées. Dans un devis longitudinal, les observations sont groupées par participant (c.-à-d. que chaque participant contribue à plusieurs observations) et parfois en fonction d'un type supplémentaire d'unités d'analyse. Le projet Capacité d'apprentissage en particulier présente des données qui sont imbriquées par garderie et par participant.

Le problème posé par le groupement de données dans les analyses de régression standards est que les observations d'un même groupe sont plus similaires entre elles qu'elles ne le sont avec d'autres observations. Par exemple, la connaissance du score d'un enfant pour un test de vocabulaire durant la période de test de base est beaucoup plus utile pour prédire le score de cet enfant à un test futur que le score d'un autre enfant. Si on l'ignore, la présence du groupement et des résidus corrélés qu'elle implique peut nous amener à sous-estimer l'ampleur des erreurs-types et, en conséquence, à surestimer la signification statistique (Hox, 2002; Moulton, 1990).

Une stratégie pour produire les erreurs-types appropriées consiste à estimer un modèle de régression spécifiant des effets aléatoires pour les unités sur lesquelles repose le groupement. Cette stratégie est appropriée (et des plus efficaces) lorsque certains postulats sont respectés (c.-à-d. que les effets aléatoires ne sont pas corrélés avec les variables indépendantes du modèle). Dans la littérature en évaluation, on fait face à cette problématique du groupement à l'aide de ce qu'on appelle un estimateur d'erreurs-types robuste au groupement (Williams, 2000). Les erreurs-types robustes au groupement ont l'avantage d'être relativement stables sous plusieurs configurations de taille d'échantillon dans le temps et l'espace (c.-à-d. période d'évaluation et garderie/enfant, Kézdi (2004); mais voyez Donald et Lang (2007) pour leur piètre rendement avec des groupes contenant très peu de sujets, p. ex. groupes de quatre ou moins). Cette façon d'estimer les erreurs-types est une généralisation pour les échantillonnages complexes de l'estimateur robuste à l'hétérogénéité de Huber-White (White, 1980). On recommande généralement l'estimateur de Huber-White parce qu'il est « robuste » à l'hétérogénéité, c'est-à-dire qu'il produit des estimations d'erreurs-types qui permettent d'interpréter les tests de signification statistique avec confiance lorsque les postulats liés à l'indépendance et à la distribution des scores ne sont pas respectés. Cette façon d'estimer les erreurs-types peut surcorriger celles-ci en leur donnant trop d'importance dans certaines situations, mais nous avons décidé d'être conservateurs en préférant la « cohérence » à l'« efficacité ». (Hayes & Cai, 2007). La façon dont l'estimateur d'erreur-type robuste a été utilisé dans les analyses d'impact présentées dans ce rapport est traitée plus en détails à la section 4.3.

### 3.2.2 Estimateur par différence de différences

Une stratégie courante pour estimer les effets programme dans la littérature en économétrie consiste à employer un estimateur par différence de différences (DD) (Abadie, 2005; Bertrand, Duflos & Mullainathan, 2004). L'estimateur DD est une technique de données recueillies au moyen d'un panel qui est appropriée aux devis de recherche non expérimental à mesures répétées employant une mesure pré-test « de base » et un groupe témoin. Les lecteurs qui connaissent moins bien la littérature en économétrie seront peut-être plus familiers avec l'idée d'une interaction ou d'un effet modérateur pour laquelle l'estimateur DD est un cas spécial. Le terme estimateur DD fait référence à l'interaction impliquant deux variables fictives, l'une représentant les périodes de test pré- et post-traitement (pré-test vs post-test) et l'autre représentant les deux groupes comparés (groupe d'intervention vs groupe témoin).

Cet estimateur peut être généralisé à des scénarios plus complexes en spécifiant plusieurs termes d'interaction de ce type dans le même modèle de régression (p. ex., pour comparer le pré-test à de multiples mesures post-test dans le même modèle de régression) ou pour comparer des groupes multiples. Les analyses d'impact rapportées dans le présent document ont fait appel à des estimateurs DD multiples pour représenter les comparaisons des trois groupes à l'étude en fonction de plusieurs évaluations post-intervention. Les trois groupes ont été inclus dans la même analyse afin de maximiser la stabilité des tests statistiques, ce qui augmente les chances de déceler les effets réels du programme.

Comme son nom le suggère, l'estimateur DD possède deux composantes de base. La première est une estimation du changement ( $\Delta$ ) de l'évaluation pré-test à une évaluation post-test pour chaque groupe ( $\Delta_{\text{groupe d'intervention}} = \text{post-test} - \text{pré-test}$ ;  $\Delta_{\text{groupe témoin}} = \text{post-test} - \text{pré-test}$ ). Ici, la mesure pré-test est la norme à laquelle toutes les évaluations subséquentes sont comparées. En d'autres termes, les mesures post-test sont toujours comparées à la mesure pré-test lors de l'estimation des effets du programme. Ces scores de changement (c.-à-d. différences) ne sont toutefois pas suffisants pour isoler l'effet traitement, car le changement faisant l'objet d'une estimation pourrait être survenu suite aux processus développementaux naturels (par exemple, la maturation).

La deuxième composante de l'estimateur DD élimine cette ambiguïté en enregistrant la différence entre les scores de changement pour le groupe d'intervention et le groupe témoin ( $DD = \Delta_{\text{intervention}} - \Delta_{\text{témoin}}$ ). Ici, l'estimation du changement fournie par le groupe témoin sert à ajuster les estimations de changement pour le groupe d'intervention. La validité de l'estimateur DD part de la prémisse selon laquelle une fois que vous éliminez l'estimation contrefactuelle du changement fournie par le groupe témoin, tout ce qui reste est le changement provoqué par l'effet traitement. Comme avec la plupart des statistiques, les postulats soutenant cette prémisse sont susceptibles d'être violés dans la pratique. Cela est particulièrement vrai dans le contexte d'une étude non expérimentale effectuée sur le terrain, dont le projet Capacité d'apprentissage est un exemple.

La validité de l'estimateur DD dépend d'au moins deux postulats. En ce qui concerne les scores de changement, nous supposons que la composition des groupes demeure constante pour toutes les périodes d'évaluation soumises à la comparaison, ce qui n'est peut-être pas le cas si les participants migrent d'un groupe à l'autre au cours de l'étude. Les changements de groupe posent un problème pour le projet Capacité d'apprentissage et ils doivent être contrôlés. Pour ce qui est de la comparaison des scores de changement, nous supposons qu'en l'absence d'un

traitement, les groupes que l'on compare entre eux ont suivi une évolution identique. Il n'y a évidemment aucune façon de vérifier ce postulat directement, mais les différences initiales entre les groupes quant aux caractéristiques associées à la mesure dépendante soulèvent la possibilité de pentes non parallèles. Les différences de groupes identifiées dans le *Rapport de référence* (Legault et al., 2014) sont un exemple de sources possibles de ce type de biais.

Nous avons minimisé ces menaces à la validité de l'estimateur DD en incluant dans nos analyses des covariables qui ont permis d'effectuer des ajustements qui tenaient compte des différences entre les groupes à leur état initial et au cours de leur trajectoire développementale, en l'absence de traitement. Nous avons intégré sélectivement des variables dans nos modèles de régression lorsqu'une telle intégration produisait une différence significative au niveau du patron de résultats rapporté. Nous traitons de cette stratégie plus en détails à la section 3.3.3.

### **3.2.3 Contrôle statistique dans le cadre d'une étude longitudinale**

L'estimateur DD neutralise efficacement l'effet statique des différences observées et non observées au sein des groupes expérimentaux à la période de base.<sup>30</sup> Par construction, ces différences sont éliminées en distinguant initialement le post-test du pré-test. Néanmoins, comme nous en avons discuté à la section précédente, l'estimateur DD peut être biaisé lorsque la composition des groupes change au fil du temps et quand on soupçonne des différences de pré-test dans les trajectoires développementales des groupes comparés. Nous abordons à tour de rôle les stratégies servant à gérer les deux types de biais.

#### ***Changements dans la composition des groupes***

Comme nous l'avons déjà dit, les changements survenant dans la composition des groupes peuvent biaiser l'estimation DD. Un tel biais peut se produire lorsqu'une caractéristique liée à une résultante varie avec le temps pour un groupe en moyenne. Par exemple, on sait que le sexe est associé à plusieurs résultantes dans la documentation sur le développement. Si la proportion de filles augmente soudainement dans un groupe avant une mesure post-test, les calculs des différences requis pour le calcul de l'estimateur DD seront biaisés.

Cette menace à la validité statistique a été contrôlée de deux manières. D'abord, nous avons surveillé les changements de composition des groupes au fil du temps. Par exemple, nous avons vérifié si le fait de changer de groupe était associé à nos covariables de base ou à la condition de traitement. Nous n'avons pas observé de telle association (résultats disponibles sur demande). Nous avons de plus vérifié s'il existait une association statistique significative entre le groupe expérimental et l'attrition (voir la section 4.2.2). Encore une fois, aucune association du genre n'a été observée. Tout compte fait, nos analyses préliminaires n'ont pas réussi à dégager des problèmes importants concernant les changements de composition des groupes. Nous avons quand même eu recours à une deuxième stratégie consistant à inclure comme covariables toutes les variables de base non redondantes associées à une résultante de façon significative. Cette inclusion des covariables visait à maximiser la validité de l'estimateur DD<sup>31</sup>. Nous présumons

---

<sup>30</sup> L'effet statique d'une caractéristique de base fait référence à une association avec des résultantes qui est stable dans le temps. Un exemple serait le fait que l'effet moyen du sexe est d'une ampleur comparable dans toutes les périodes d'évaluation durant la première année. Si l'effet du sexe était dynamique, sa force d'association avec une résultante varierait considérablement en fonction de la période d'évaluation prise en considération.

<sup>31</sup> Dans ce cas, l'estimateur DD représente l'effet traitement moyen estimé de notre intervention, sous réserve des covariables qui sont incluses dans le modèle. Nous présumons que la condition de traitement ne tempère pas l'effet des covariables.

que l'ensemble de covariables inclus dans notre modèle a contrebalancé tout biais lié à la sélection de participants pour les groupes expérimentaux.

### ***Pentes développementales parallèles***

Même si la composition des groupes demeure constante à travers le temps, des différences initiales entre les groupes expérimentaux peuvent être associées à des tendances développementales non parallèles. Comme nous l'avons vu plus haut, les tendances développementales non parallèles (en l'absence de traitement) compromettent la validité de l'estimateur DD. Si la cause de ce non-parallélisme est observée (c.-à-d. que nous avons collecté les données d'une covariable appropriée), il est possible d'apporter les ajustements statistiques qui corrigeront un tel biais (Abadie, 2005). Un exemple de cet ajustement serait d'inclure, en tant que covariable, l'interaction entre le temps et la variable pertinente. Les variables qu'on pourrait songer à employer pour apporter cet ajustement comprennent les covariables exogènes de base qui sont: a) associées à la mesure de résultantes; et b) réparties différemment dans les groupes comparés. Les limites pratiques imposées par la taille de notre échantillon nous obligeaient à être sélectifs dans nos choix des covariables à utiliser pour apporter un tel ajustement, afin d'éviter le surajustement des données. Ainsi, nous avons seulement inclus des termes interactifs de temps par covariable s'ils étaient des prédicteurs statistiquement significatifs d'une résultante (n'importe quelle résultante) aux fins de notre échantillon (voir la section 4.4).

### ***Validité du dosage estimé et des effets de la qualité***

Dans la discussion précédente, l'exposition au traitement ou «dosage» ainsi que la qualité ont été présentées comme alternatives à la seule appartenance au groupe expérimental pour classer les participants de non traités à traités le long du continuum. Les estimations des effets reposant sur ces définitions plus précises peuvent être plus efficaces, mais elles sont assujetties aux mêmes limites que les estimations reposant seulement sur l'appartenance au groupe expérimental. Nous invoquons ici les postulats qui étayaient la validité de nos estimations de l'effet traitement: que les conditions du traitement, en l'absence de tout nouvel investissement de la part des autres paliers de gouvernement, auraient été comparables.

L'analyse préliminaire visait principalement à établir la comparabilité des groupes à l'étude. Si nous devons comparer des enfants aux degrés de dosage variables à l'intérieur des groupes, et d'un groupe à l'autre, (p. ex., pour estimer l'effet des heures passées en garderie), alors nous devons présumer que les enfants en question sont par ailleurs comparables. Par exemple, nous présumons que les enfants qui passent en moyenne 40 heures par semaine en garderie sont équivalents, en ce qui a trait à toute autre caractéristique pertinente, aux enfants qui passent seulement en moyenne 10 heures par semaine en garderie. Ce postulat a peu de chances d'être respecté, étant donné qu'une diversité de variables sociodémographiques sont potentiellement associées aussi bien au temps passé en garderie qu'aux résultantes. Nous avons aussi présumé que les garderies programme et les garderies témoin auraient été équivalentes pour ce qui est de la fidélité et de la qualité en l'absence d'intervention. Ce postulat n'a pas été vérifié empiriquement (c.-à-d. qu'aucune mesure pré-test n'a été prise des dimensions) et il est peu probable qu'il soit (complètement) respecté étant donné le petit nombre de garderies et l'impossibilité de la mise en correspondance des variables en question.

Nous avons utilisé deux stratégies pour pallier à ces deux sources potentielles de biais dans l'estimation de l'effet du dosage et de la qualité. En premier lieu, le devis longitudinal de la

présente étude permet de recourir à l'estimateur DD, qui neutralise l'effet statique des caractéristiques de base (tant observées que non observées). En second lieu, la spécification finale qui a servi à estimer les effets du traitement comprenait toutes les covariables non redondantes associées à au moins une des mesures de résultantes (voir la section 4.4). En présumant que toutes les variables associées à différents degrés d'exposition à la garderie (ou à la fidélité/qualité) ont été adéquatement prises en compte dans le cadre de cette stratégie (c.-à-d. que le modèle a été correctement spécifié), nous pouvons conclure que les estimations conditionnelles résultant des effets du dosage ne sont pas biaisées. Les mêmes stratégies et postulats ont été retenus pour estimer l'effet de la fidélité au programme et de la qualité du programme.<sup>32</sup>

### **3.2.4 Ateliers d'alphabétisation familles : Un cas spécial**

L'analyse des données sur les ateliers d'alphabétisation familiale doit être traitée séparément, car il s'agit d'un cas spécial dans le contexte plus large des analyses qui figurent dans ce rapport. Seulement deux échelles parentales ont servi dans chaque sondage de suivi: la fréquence des activités d'alphabétisation; et la langue des activités d'alphabétisation. Ces échelles ont fait l'objet d'une analyse comparable à celle de l'ÉPE-AD (c.-à-d. par l'entremise de l'estimateur DD). Pour les autres résultantes (c.-à-d. échelles des connaissances, d'auto-efficacité et de modélisation), les données n'étaient disponibles que pour les parents dont les enfants étaient inscrits dans une garderie programme. Nous appellerons dorénavant ce groupe d'échelles les échelles des ateliers pour parents. Pour l'analyse de ces échelles, nous avons adopté une stratégie quelque peu différente que nous décrivons dans les sections ci-dessous.

#### ***Échelles des ateliers pour parents : Choix de l'échantillon***

L'un des premiers traits distinctifs de l'analyse de ces sous-échelles parentales est qu'elle repose sur un échantillon total de participants (les six communautés) inscrits dans une garderie programme au moment où les ateliers ont été offerts (N=105). Contrairement à l'analyse de résultantes portant sur les enfants, les communautés d'Edmonton et de St-Jean n'ont pas été exclues de cette analyse. Cette décision s'appuyait sur trois raisons. D'abord, contrairement à l'intervention en garderie, les ateliers familles ont été correctement mis en œuvre dans toutes les communautés (voir le *Rapport de mise en œuvre du projet*, 2014). Ensuite, étant donné que l'analyse vise seulement les parents dont les enfants étaient inscrits à l'une des garderies programme, la disponibilité d'un groupe témoin en garderie dans toutes les communautés ne pose pas de problème. Enfin, l'exclusion des communautés de St-Jean et Edmonton aurait réduit l'échantillon à un niveau ne permettant pas d'évaluer correctement les effets des ateliers. Les caractéristiques de l'échantillon total des familles du groupe programme en garderie ne sont pas rapportées ici (voir le tableau 2.5 pour les taux de réponses), mais elles font l'objet d'une description détaillée dans le *Rapport de référence* (Legault et al., 2014).

#### ***Échelles des ateliers pour parents : Stratégie d'analyse***

Le deuxième trait qui distingue l'analyse des échelles des ateliers pour parents est l'absence d'un groupe témoin planifié. Les mesures n'ont été administrées qu'aux parents dont les enfants

---

<sup>32</sup> Toutes les covariables sont au niveau du « participant » plutôt que de la garderie, sauf pour la variable de la communauté. Le nombre de garderies était insuffisant pour permettre d'inclure une large palette de caractéristiques à ce niveau d'analyse.

étaient inscrits à l'une des garderies participant à l'intervention et il n'existe pas de données pour certaines mesures dans le cas des familles ayant participé à moins de trois ateliers (n = 15). Un devis équilibré serait doté d'un ensemble complet de mesures pré-test et post-test pour toutes les variables et pour tous les participants, mais ce n'est pas le cas ici. Par souci d'efficacité, un devis non équilibré a été employé, qui a nécessité une méthode sélective de collecte de données.

Spécifiquement, les familles classées comme « non participantes » n'ont eu qu'à fournir des estimations post-test de leurs attitudes, ce qui signifie que les estimations rétrospectives (Lamb & Tschillard, 2005; Rockwell & Kohn, 1989) ne sont pas disponibles pour ce groupe. Un ensemble complet de données post-test et de données rétrospectives de pré-test était disponible pour les échelles des connaissances et d'auto-efficacité des familles classées comme « participantes ». Les estimations réelles pré-test et post-test ont été recueillies pour l'échelle de modélisation de toutes les familles peu importe la participation. L'approche analytique utilisée dans les analyses associées a été adaptée aux complexités du devis.

Étant donné la structure de l'ensemble de données, nous disposons d'au moins deux façons d'estimer l'impact des ateliers. La première méthode estime l'impact des ateliers en fonction de la différence entre les mesures pré-test et post-test. La deuxième façon d'estimer l'impact des ateliers consiste à comparer les parents qui étaient classés comme participants (n=90) à ceux qui ne l'étaient pas (n=15). Chaque type de comparaison en lui-même manque de validité, mais le patron global produit par une batterie de tests peut être instructif.

La validité des scores de changement (différence entre les mesures pré-test et post-test pour les participants uniquement) est douteuse, non parce que ces derniers sont rétrospectifs (pour une discussion sur les avantages des estimations rétrospectives par rapport aux pré-tests traditionnels, voir Lamb et Tschillard, 2003; 2005), mais plutôt en raison de l'absence d'un groupe témoin servant à contrôler pour la maturation et les effets historiques. Ce problème de validité est exacerbé par le fait que les participants, conscients d'être traités, pourraient produire des estimations de leur état pré-test et post-test conformes à leurs attentes de l'effet positif des ateliers.<sup>33</sup> Dans le même ordre d'idées, la comparaison de participants et de non-participants est invalidée par la possibilité évidente d'une sélection biaisée qui ne peut être contrôlée ni statistiquement ni au moyen de l'appariement en raison du nombre trop faible de participants au sein du groupe de non-participants.

Individuellement, les deux estimations de l'effet traitement sont biaisées, mais ensemble elles peuvent produire des résultats instructifs sans écarter complètement toutes les sources de biais. Par exemple, nous soutenons que le patron de résultats suivant correspond à un véritable effet traitement: a) l'effet traitement basé sur le score de changement pour les participants est statistiquement significatif et positif; b) l'estimation pré-test rétrospective pour les participants est équivalente à l'estimation « post-test » fournie par les non-participants; et c) les participants rapportent des estimations post-atelier supérieures à celle des non-participants en moyenne. Nous présumons que parce que les non-participants n'ont pas participé au programme, leur estimation « post-test » est aussi valide qu'une estimation de leur état pré-test. En d'autres mots, nous

---

<sup>33</sup> Ce type de biais peut prendre plusieurs formes et n'est pas propre aux pré-tests rétrospectifs. Les parents peuvent exagérer l'effet programme en sous-estimant leur état pré-test et en surestimant leur état post-test, ou en produisant des estimations contaminées des deux manières. L'expérience a montré qu'une sous-estimation de l'état pré-test est caractéristique des mesures rétrospectives (Taylor, Russ-Eft & Taylor, 2009). Les mesures pré-test prospectives traditionnelles sont biaisées d'autres façons (p. ex. biais de réponses améliorées; Howard, 1980), ce qui entraîne parfois la surestimation de l'état pré-test (Moore & Tananis, 2009).

postulons que leur état véritable, en moyenne, n'a pas changé durant la période où les ateliers étaient offerts. Si le patron de résultats escompté est obtenu, cela permet de supposer que les ateliers ont eu un effet positif. Si seulement le résultat (b) est obtenu, alors on peut supposer que l'effet du traitement est nul (ou négatif). Si les résultats (a) et (c) sont obtenus sans le résultat (b), alors on peut fortement supposer que les estimations sont biaisées et que l'interprétation des effets positifs du programme doit être nuancée en conséquence. Autrement dit, il y a un patron spécifique de résultats qui appuie fortement l'impact positif réel des ateliers d'alphabétisation familiale, et bon nombre d'autres patrons qui ne procurent pas cet appui. Si notre prédiction « risquée » se confirme, elle prête de la crédibilité à la validité de notre interprétation quant aux effets du traitement.

Les limites imposées par les données disponibles signifient qu'un estimateur comme l'estimateur DD peut être écarté parce qu'il n'y a pas de mesure pré-test pour les deux groupes. Les trois hypothèses formulées ci-dessus peuvent plutôt être testées à l'aide d'une série de tests indépendants, un pour chaque hypothèse. Les covariables n'ont pas pu servir à tester les différences entre les groupes, car la trop petite taille d'échantillon du groupe de non-participants rendait la chose impossible (n=15). De plus, les covariables n'ont pas servi à estimer les scores de changement des participants, parce que: a) l'effet statique des caractéristiques des participants est neutralisé (« c'est-à-dire éliminé absent des différences ») au cours de cette estimation; et b) les changements dans la composition des groupes ne posent pas un problème.

Les estimations de fidélité, de la qualité et du dosage du programme (proportion d'ateliers fréquentés) étaient disponibles pour l'échantillon de parents qui ont assisté aux ateliers (n=90). Nous avons évalué l'impact de ces facteurs en utilisant un estimateur DD, comme pour d'autres analyses rapportées ici. Les estimations produites représentent les impacts des ateliers quant à l'intensité du traitement reçu. Encore une fois, comme les changements dans la composition des groupes ne constituaient pas un problème pour cette analyse, nous avons compté sur le fait que l'estimateur DD n'est pas biaisé par les caractéristiques de base observées et non observées.



## 4. Analyses préliminaires

Avant de procéder aux analyses d'impact, une variété de vérifications et analyses de la qualité préliminaires ont été effectuées. Ce chapitre brosse le portrait des étapes préliminaires franchies afin de préparer les données à l'analyse. Nous décrivons les processus de contrôle de la qualité mis en place pour recueillir les données à la section 4.1 et les résultats de l'analyse des valeurs manquantes ainsi que l'imputation des données à la section 4.2. Puis, les analyses ayant une pertinence directe pour les analyses d'impact sont décrites. Les processus sur lesquels reposent les décisions liées à la spécification des termes d'erreur et la sélection des covariables sont décrits aux sections 4.3 et 4.4, respectivement. Enfin, la question de validité externe est traitée à la section 4.5, où l'on compare les caractéristiques du présent échantillon à celles d'un échantillon conçu pour être représentatif de la population cible (Enquête sur la vitalité des minorités de langue officielle, appelée dorénavant EVMLO).

### 4.1 PROCESSUS DE CONTRÔLE DE LA QUALITÉ

Assurer la qualité des données recueillies comprend plusieurs étapes allant de la sélection ou conception des instruments à la saisie des données, en passant par la collecte de données. La sélection ou la conception des *outils de mesure* (p. ex. sondage, grille d'entrevue, grille d'observation sur le terrain, etc.) est une première étape dans la production d'un ensemble de données « épurées ». Dans la mesure du possible, l'équipe du projet Capacité d'apprentissage a sélectionné des échelles préexistantes qui ont été testées et validées. Lorsque de telles mesures n'étaient pas disponibles, l'équipe de recherche du projet Capacité d'apprentissage a mis au point de nouveaux instruments (p. ex. échelle, grille d'observation ou grille d'entrevue) en appliquant des principes psychométriques éprouvés. Par exemple, nous avons veillé à ce que l'importance des diverses sources d'erreur de mesure soit minimale. Les questions ont été formulées dans un langage clair, précis et simple. Entre autres considérations, le format des instruments de mesure et les questions ont été conçus de façon à ne pas imposer un exercice de mémoire trop lourd aux participants. Tous les outils de mesure ont fait l'objet de tests pilotes avant leur usage sur le terrain. Dans la batterie de procédures de contrôle des tests pilotes, on demandait à un groupe de membres de l'équipe de valider: a) la clarté du libellé des questions (en français et en anglais); b) la cohérence des questions dans les deux langues quant à leur forme et leur contenu; c) la présentation logique des items; et d) la pertinence des choix de réponses.

En ce qui concerne plus particulièrement la mesure de résultantes pour les enfants (c.-à-d. l'ÉPE-AD), un test pilote a été mené au printemps 2007 auprès d'un échantillon d'enfants afin de répondre à trois grandes préoccupations : 1) déterminer si l'outil de mesure pouvait être employé avec des enfants aussi jeunes que les enfants visés par le projet Capacité d'apprentissage (qui étaient de plusieurs mois plus jeunes que les enfants de l'échantillon original utilisé pour mettre au point l'ÉPE-AD lors de sa conception initiale par le concepteur de l'outil); 2) déterminer si l'outil de mesure pouvait être employé auprès d'enfants bilingues (une sous-population qui n'était pas explicitement représentée dans l'échantillon utilisé par le concepteur de l'outil); et 3) acquérir sur le terrain une expérience de l'administration de l'outil de mesure, dans diverses conditions et dans différents milieux (p. ex. foyer, services de garde, etc.).

En résumé, la SRSA a pris des moyens concrets pour réduire les différentes sources d'erreur de mesure qui peuvent apparaître dans ces circonstances. Ces démarches étaient essentielles, étant donné la nouveauté relative de l'instrument de mesure et les circonstances particulières de son utilisation dans le cadre du projet Capacité d'apprentissage.

Le personnel de recherche a veillé à adopter des normes élevées pour assurer la qualité de la collecte de données en élaborant ***un matériel technique et des manuels d'instruction propres à l'instrument employé***. Ces documents ont été distribués à ***tous les employés visés, qui ont aussi reçu une formation pratique concernant l'administration des instruments de mesure***. Les employés visés comprenaient les intervieweurs, les évaluateurs et les observateurs sur le terrain. La formation de tous les employés a été mise à jour avant chaque période de test. Par exemple, les évaluateurs ont reçu une formation continue, qui était prévue avant chaque évaluation. Ces séances de formation ont permis aux évaluateurs de se perfectionner sur l'administration de l'ÉPE-AD et d'apprendre les nouvelles mesures de résultantes pour les enfants qui étaient ajoutées à la batterie d'instruments. De leur côté, les intervieweurs qui administraient les sondages auprès des parents ont été informés de la nature et de l'objectif des questions du sondage avant chaque vague de collecte de données. La formation continue du personnel de recherche a assuré la cohérence dans l'emploi des mesures à travers le temps, ainsi que la cohérence au sein du personnel de collecte de données lui-même. On a répondu à toutes les questions et préoccupations du personnel de collecte des données durant ces séances de formation et d'information. L'équipe de recherche du projet Capacité d'apprentissage a suivi de près les progrès de l'équipe de collecte de données, lui fournissant les techniques, les stratégies et l'aide nécessaires. Dans certains cas, des ajustements ont été apportés au plan de collecte de données en réponse au rendement des mesures prises sur le terrain.

Un ***premier contrôle de la qualité des données*** a été mis en place au début du projet Capacité d'apprentissage. Nous avons recueilli les données brutes en deux étapes. Les coordonnatrices communautaires ont servi de nœud de raccordement, recevant les données recueillies dans leurs communautés respectives. Une fois les renseignements recueillis, ils étaient transmis au bureau d'Ottawa de la SRSA. Par conséquent, les coordonnatrices communautaires étaient chargées de faire une première vérification de la qualité des données, ce qui a permis d'offrir une rétroaction opportune au personnel de collecte de données sur le terrain. Autrement, il aurait fallu attendre que la SRSA reçoive les données, ce qui aurait repoussé la rétroaction de plusieurs jours à cause du temps nécessaire pour transmettre l'information. Les coordonnatrices communautaires ont veillé à ce que les données manquantes soient minimales dans les sondages et les évaluations des parents en vérifiant qu'ils les aient remplis adéquatement. Si on trouvait des erreurs ou des omissions, on demandait aux intervieweurs ou aux évaluateurs de recueillir les renseignements manquants en contactant les parents ou en complétant l'évaluation de l'enfant. Grâce aux coordonnatrices communautaires, il a été possible d'assurer les taux élevés de réponse et de rétention observés dans le projet Capacité d'apprentissage.

Un ***deuxième contrôle de la qualité des données*** a été effectué au bureau d'Ottawa pour vérifier la qualité des données saisies ainsi que les propriétés psychométriques des échelles de mesure. Les données ont été saisies dans une base de données électronique, puis soumises à un système de vérification rigoureux afin d'en assurer l'exactitude. Dans un premier temps, une vérification aléatoire de 10 % des données a été effectuée pour assurer l'exactitude de la saisie de données. Par la suite, des analyses descriptives ont été réalisées pour vérifier si les fréquences des items tombaient dans le domaine de valeurs escomptées. Une batterie de contre-vérifications

ont été réalisées à partir des bases de données électronique pour s'assurer que les réponses étaient cohérentes pour chacun des répondants, dans les sondages individuels aussi bien qu'au fil du temps. Les valeurs incompatibles ou improbables étaient vérifiées grâce à la copie papier ou auprès de la personne responsable de la collecte des données. Nous avons ensuite employé des méthodes statistiques pour confirmer la qualité des échelles de mesure. La cohérence interne des échelles et la validité des dimensions mesurées ont fait l'objet d'une vérification avec l'alpha de Cronbach et l'analyse factorielle, respectivement. Nous avons analysé la validité de construit des variables mesurées en vérifiant si la direction des inter-corrélations observées parmi les variables correspondait aux attentes. En dernier lieu, nous avons régulièrement contrôlé les données pour dégager les observations aberrantes unidimensionnelles et multidimensionnelles avant l'analyse. Toutes les données de sondage collectées ont été soumises à l'imputation des valeurs manquantes, conformément aux procédures reconnues (Cohen, Cohen, West & Aiken, 2003; pp. 431-451; Tabachnick & Fidell, 2007, pp. 62-71).

Un *troisième contrôle de la qualité des données* consistait à employer une approche de méthodes mixtes. Le projet Capacité d'apprentissage a eu recours à une diversité d'outils, tant quantitatifs que qualitatifs, et à de nombreuses sources de données choisies en fonction des objectifs de recherche. Chacune de ces sources de données a servi à trianguler les résultats des recherches, ce qui a renforcé notre confiance dans les conclusions auxquelles nous sommes parvenus. De plus, la complémentarité de l'ensemble des résultats recueillis permet de mieux comprendre le phénomène à l'étude, avec toutes les subtilités qu'il recèle.

En dernier lieu, la *cohérence de la composition des échelles* à travers le temps était essentielle pour préserver la cohérence interne des analyses d'impact. À ce titre, une vérification a été effectuée pour confirmer que les échelles répétées avec le temps contenaient les mêmes items et des choix de réponses identiques. Si la composition des éléments d'échelle variait avec le temps (p. ex. dans le cas des résultantes pour les enfants), on procédait à la normalisation des scores dans chaque version de l'ÉPE-AD pour rendre les mesures comparables au fil du temps et permettre ainsi qu'elles fassent l'objet des analyses qu'il était prévu de mener concernant les impacts du programme (voir la section 5.1.2).<sup>34</sup>

## 4.2 ANALYSE DES VALEURS MANQUANTES

Les valeurs manquantes dans une base de données peuvent compromettre la validité d'une analyse. Cette menace provient de deux sources principales. D'abord, les cas avec des valeurs manquantes sur une variable sont habituellement exclus d'une analyse. L'exclusion des cas avec des valeurs manquantes peut être problématique lorsque ceux-ci diffèrent considérablement des cas qui demeurent dans l'analyse. Lorsque cela se produit, la composition de l'échantillon change, ce qui a des répercussions sur la validité externe des résultats. Par exemple, si les valeurs manquantes surviennent de façon disproportionnée pour les filles, les résultats obtenus pourraient ne pas être généralisables à cette population. Ensuite, la validité interne des estimations de l'effet programme dépend de la stabilité de la composition des groupes au fil du temps. Si les valeurs manquantes perturbent la composition des groupes, cela peut biaiser les estimations de l'effet

---

<sup>34</sup> La standardisation fait ici référence à la transformation statistique des scores bruts ( $x$ ) d'une variable en écarts réduits, où  $z = (x - \text{score moyen}) / \text{écart-type}$ . Toutes les variables normalisées ont une moyenne de 0 et un écart-type de 1.

programme. Nous présentons maintenant une analyse des valeurs manquantes qui évalue le potentiel pour les deux types de biais.

#### **4.2.1 Répartition générale des valeurs manquantes**

Les valeurs manquantes dans une base de données peuvent être divisées en deux types majeurs: planifiées et non planifiées. Les valeurs manquantes planifiées affectent seulement les échelles de l'ÉPE-AD, tant la version française qu'anglaise, pour les deuxième, troisième et cinquième périodes d'évaluation. Ces valeurs manquantes résultaient de l'application des règles de départ pour l'administration de certaines sous-échelles. Aucun enfant ne répondait aux critères pour l'application de la règle de départ lors de la première évaluation; ils étaient tous âgés de moins de quatre ans.

Les valeurs manquantes non planifiées survenaient à la fois pour les mesures administrées aux enfants et les mesures administrées aux adultes. Les valeurs manquantes non planifiées étaient de deux types: a) le résultat de l'incapacité d'amasser quelque donnée que ce soit auprès d'un participant durant une période de test donnée (manquant complètement); et b) le résultat d'un processus où un participant n'a pas répondu à une question particulière ou n'a pas été soumis à une échelle particulière (manquant partiellement).<sup>35</sup> Nous décrivons maintenant à tour de rôle la fréquence de l'incidence de ces deux types de valeurs manquantes.

##### ***Valeurs manquantes planifiées***

Toutes les valeurs manquantes planifiées ont produit des valeurs manquant partiellement. En aucun cas n'avons-nous intentionnellement sauté une évaluation ou une échelle entière. Les éléments faciles au début des échelles de l'ÉPE-AD ont été omis intentionnellement pour certains enfants en raison de l'application de la règle de départ prescrite par le concepteur de l'outil. La règle de départ sert à augmenter l'efficacité du processus d'évaluation et de réduire au minimum les erreurs de mesure attribuables à l'ennui ou à la fatigue des enfants. L'hypothèse qui soutient le recours à la règle de départ suppose que l'enfant aurait obtenu le score maximal pour les items omis s'ils avaient été administrés. Si cette hypothèse se vérifie, alors les valeurs manquantes estimées sont remplacées par le score maximal possible sans aucune perte d'information. Toutefois, elle ne sera accompagnée d'une mesure du rendement que si les items d'une échelle sont présentés en ordre de difficulté et qu'un bon prédicteur du rendement de l'enfant est utilisé pour déterminer le moment où appliquer la règle de départ.

Pour les deuxième et troisième administrations de l'ÉPE-AD (février 2008 et juin 2008, respectivement), des candidats auxquels pouvaient s'appliquer les règles de départ ont été identifiés selon leur âge, conformément aux lignes directrices de l'ÉPE-AD. Les enfants de plus de quatre ans devaient sauter les six premiers items des domaines Langue et communication et Conscience de soi et de l'environnement, ainsi que les huit premiers éléments des domaines Habilités cognitives et Physique/moteur ». Une règle faisait en sorte que les enfants de plus de quatre ans qui n'obtenaient pas un score de 3 ou 4 aux deux premières questions devaient revenir au début de l'échelle. En pratique, cette règle dite « de demi-tour » n'était pas toujours respectée. Même dans les cas où l'évaluateur a décidé à juste titre de ne pas appliquer la règle du demi-tour,

---

<sup>35</sup> Les données manquant partiellement peuvent être attribuées à diverses causes, par exemple une erreur commise par l'administrateur du test, ou dans le cas de l'ÉPE-AD, l'utilisation d'une autre langue pour certaines sous-échelles qui ne peuvent pas par la suite être combinées aux fins des analyses.

l'enfant n'aurait pas nécessairement reçu le score maximal aux items omis (c.-à-d. un score de 4), s'ils avaient été administrés. En somme, l'application de la règle de départ a créé des valeurs manquantes pour les items au début de certaines échelles. Pour la deuxième et la troisième administration de l'échelle de l'ÉPE-AD, le nombre d'enfants ayant des valeurs manquantes à cause de la règle de départ était de 6 et de 42 respectivement.

Par conséquent, il était nécessaire d'estimer la valeur qu'auraient prise les valeurs manquantes si les items omis avaient été administrés, mais la méthode utilisée pour accomplir cette tâche restait à déterminer. Comme on l'a mentionné ci-dessus, le concepteur de l'outil a recommandé que les valeurs manquantes en raison de la règle de départ soient remplacées par le score maximal possible que l'enfant aurait pu obtenir. Nous avons délaissé ce plan en nous fondant sur deux arguments. Premièrement, les éléments des sous-échelles de l'ÉPE-AD n'étaient manifestement pas présentés en ordre ascendant de difficulté. Les analyses préliminaires de la SRSA suggéraient que certains éléments plus faciles étaient présentés après des éléments plus difficiles. Cette observation est corroborée par le fait que les versions plus récentes de l'ÉPE-AD montrent un ordre d'éléments réaménagé. Deuxièmement, le point de coupure basé sur l'âge n'était pas un prédicteur assez fiable des mesures de résultantes pour prédire avec certitude que l'enfant aurait obtenu la valeur maximale sur les éléments de la règle de départ. Ceci, combiné aux erreurs dans l'application de la « règle de demi-tour », signifiait que le score maximal ne représentait pas une approximation adéquate du score véritable d'un enfant. Ceci étant le cas, l'imputation directe des valeurs manquantes planifiées à l'ÉPE-AD avec le score d'élément maximal a été abandonnée à la faveur d'une stratégie qui saurait représenter plus adéquatement la variation de la performance au niveau individuel. Cette stratégie fait l'objet d'une discussion plus détaillée à la section 4.2.3, Stratégie d'imputation des données.

Pour la cinquième période d'évaluation (et l'administration finale de la version initiale de l'ÉPE-AD, en février 2009), nous avons encore appliqué une règle de départ, cette fois selon la performance des enfants au cours de l'évaluation précédente. On croyait qu'en capitalisant sur les données des évaluations antérieures, cela rendrait la règle de départ plus précise et minimiserait toute perte d'information. La règle de départ s'appliquait seulement si l'enfant recevait un score de 10 ou plus aux quatre premiers éléments des échelles Langue et communication pour l'évaluation précédente et un score de 21 ou plus sur les sept éléments de l'échelle Conscience de soi et de l'environnement. Lorsque la règle de départ était appliquée, les enfants pouvaient sauter les trois premiers éléments de l'échelle Langue et communication, ainsi que les six premiers éléments de l'échelle Conscience de soi et de l'environnement lors de l'évaluation subséquente devant se dérouler au cours de la cinquième période d'évaluation. Une autre restriction était appliquée: les quatrième et cinquième évaluations devaient être menées dans la même langue.<sup>36</sup> Cette condition s'appuyait sur le fait que rien ne prouve hors de tout doute que les versions française et anglaise de l'ÉPE-AD sont interchangeables.<sup>37</sup> En l'absence

---

<sup>36</sup> Dans 12 cas, cette condition n'a pas été respectée. Les scores qu'auraient obtenus les participants pour ces items ont été estimés au moyen de l'imputation, comme on l'avait fait lors des deuxième et troisième évaluations. Voir la section 4.2.3, Stratégie d'imputation des données.

<sup>37</sup> Les analyses préliminaires non rapportées ont démontré que les enfants ayant changé de langue d'une évaluation à l'autre produisaient des scores de changement très différents de ceux des autres enfants de l'échantillon. De plus, l'analyse du contenu des versions française et anglaise de l'ÉPE-AD montrait que la traduction française n'était pas bien adaptée. Sans preuve positive du contraire, nous avons conclu que les deux versions de l'ÉPE-AD ne devraient pas être combinées.

d'une telle preuve, nous avons jugé plus sage de ne pas combiner les données des deux langues d'évaluation à quelque fin que ce soit. La liste des enfants susceptibles de sauter les premiers éléments des échelles de l'ÉPE-AD a été préparée en vue de la collecte des données de la cinquième évaluation, menée auprès de 188 des 240 enfants qui participaient encore, à ce moment-là, au projet et pour lesquels la langue de l'évaluation était demeurée constante.

La principale source de valeurs manquantes dans l'ÉPE-AD au niveau des éléments individuels était la mise en œuvre des règles de départ. Toutefois, une poignée ( $n < 5$ ) de valeurs manquantes non planifiées a aussi été observée pour tous les items et périodes de temps de l'ÉPE-AD. Étant donné le petit nombre négligeable de valeurs manquantes et le fait qu'elles manquaient au niveau des items, nous ne les avons pas soumises à une analyse plus détaillée et les avons traitées comme manquantes à cause de la règle de départ aux fins de l'imputation des données.

### ***Valeurs manquantes non planifiées***

Les valeurs manquantes non planifiées ont affecté tant les sondages auprès des parents que l'évaluation des enfants. Ces valeurs manquantes pouvaient provenir d'un sondage ou d'une évaluation manquant(e). De plus, les données provenant des sondages étaient parfois incomplètes, ce qui signifie que des données valides étaient obtenues pour certaines questions, et non pour d'autres, soit en raison d'une erreur commise par l'interviewer, soit parce que le participant avait refusé de répondre.<sup>38</sup> Des deux types de valeurs manquantes, le second est possiblement plus problématique parce que le mécanisme qui l'a généré peut être fonction de la question elle-même.<sup>39</sup> Nous décrivons maintenant ces sources de valeurs manquantes ainsi que leurs conséquences sur la stratégie retenue aux fins de l'imputation ou du remplacement des données.

À chaque période de test, un faible pourcentage de participants ont été classés comme ayant des données manquant complètement. Celles-ci sont causées soit par le défaut d'administrer les mesures, soit par l'attrition des participants (voir la méthodologie au chapitre 5). Les données manquantes en raison de l'attrition des participants sont problématiques, en ceci qu'elles limitent les options d'imputation des données dans un devis longitudinal. Par exemple, dans le cadre d'un devis longitudinal, il est possible de poser la même question à plusieurs reprises afin d'accroître la possibilité d'obtenir l'information recherchée auprès de tous les participants. Pour illustrer ceci dans le contexte du présent projet, nous avons posé des questions sur le revenu familial à trois reprises durant les deux premières années de l'étude.<sup>40</sup> Pour les participants qui sont demeurés avec le projet jusqu'au huitième sondage, il est possible d'exploiter entièrement une telle redondance en imputant de façon croisée les valeurs manquantes pour ces questions lorsqu'elles surviennent.<sup>41</sup> Étant donné ces particularités, les cas isolés de données manquant

---

<sup>38</sup> Ce type de valeur manquante partielle n'a pas été observé pour la version française des échelles de l'ÉPE-AD. Toutes les analyses rapportées dans ce document étant fondées sur la version française de l'ÉPE-AD, la question des valeurs manquant partiellement n'entre plus en considération en ce qui a trait à ces mesures de résultantes.

<sup>39</sup> L'exemple classique servant habituellement à illustrer ce point est le cas où l'on pose des questions sur le revenu familial.

<sup>40</sup> Cette question a été posée dans le sondage de base et lors des septième et huitième sondages de suivi auprès des parents (juin et octobre 2009 respectivement).

<sup>41</sup> Nous devons aussi présumer qu'il est peu probable que la variable mesurée change systématiquement avec le temps ou qu'elle soit affectée par le traitement. Nous considérons que le revenu familial est un excellent exemple de variable pouvant évoluer au fil du temps, mais sans vraisemblablement montrer de relation systématique avec le temps ou avec l'assignation des participants aux groupes à l'étude.

« complètement » ont été traités différemment des cas d'attrition dans l'analyse et la gestion des données manquantes. On trouvera de plus amples renseignements à ce sujet à la section 4.2.3, Stratégie d'imputation des données.

Une autre source de valeurs manquantes dans la base de données était les sondages contenant des données incomplètes. Comme nous l'avons noté ci-dessus, ce type de données manquantes peut causer des problèmes pour l'analyse des données. Le principal problème est de savoir si une quelconque propriété de la question posée augmente systématiquement la possibilité qu'un participant ne réponde pas à la question. Par exemple, il est bien connu que les répondants situés aux deux extrêmes de la fourchette de revenu sont plus susceptibles de ne pas vouloir répondre aux questions liées au revenu. On voit clairement dans cet exemple que le processus produisant les valeurs manquantes est systématique. Si le mécanisme produisant les valeurs manquantes est systématique (non aléatoire), l'exclusion de cas comportant des valeurs manquantes modifiera les caractéristiques de l'échantillon de façon significative.

Un exemple d'un tel mécanisme réside dans la règle utilisée pour décider de la langue d'administration de l'ÉPE-AD. Ce mécanisme cause des valeurs manquantes pour les versions françaises de toutes les échelles, sauf pour Langue et communication chez les enfants ayant les compétences les plus faibles en langue française. En conséquence, l'échantillon réduit est plus fort en français que l'original, ce qui a pour effet de réduire l'étendue de cette dimension, atténuant potentiellement les impacts du programme sur les variables liées à la maîtrise du français. D'autres mécanismes responsables de la création de valeurs manquantes peuvent générer des types similaires de biais. Les analyses présentées dans la section 4.2.2 se rapportent à la description du patron des valeurs manquantes et aux mesures prises afin d'en limiter les conséquences pour la validité de l'étude.

#### 4.2.2 Patron des données manquantes

Les données manquantes sont comme tout autre résultat, c'est-à-dire qu'il est possible de modéliser le processus qui les a engendrées. Les conclusions tirées d'un tel exercice de modélisation déterminent les mesures à prendre pour préserver la validité interne et la validité externe de l'étude. Si le processus qui a engendré les données manquantes est aléatoire, la validité n'est pas menacée. Par contre, s'il n'est pas aléatoire, il faut alors prendre des mesures pour éviter d'introduire un biais en excluant des cas auto-sélectionnés. Nous décrivons maintenant la stratégie qui a servi à discerner la nature du patron des données manquantes dans la base de données du projet Capacité d'apprentissage.

Selon Little et Rubin (1987), on peut classer les valeurs manquantes non planifiées sous trois types: a) manquant complètement aléatoire (*missing completely at random*; MCAR); b) manquant aléatoire (*missing at random*; MAR); et c) manquant non aléatoire (*missing not at random*; MNAR non-négligeable). Le scénario le plus souhaitable est une situation où un petit nombre de valeurs manquantes sont distribuées aléatoirement dans un ensemble de données (c.-à-d. MCAR). En revanche, le pire scénario serait d'avoir un grand nombre de données manquantes distribuées de façon non-aléatoire. Dans le premier cas, le problème des valeurs manquantes peut être résolu par une méthode « *listwise deletion* » sans biaiser les résultats d'une analyse (c.-à-d. suppression des cas avec valeurs manquantes). Traiter ainsi les valeurs manquantes distribuées de façon non-aléatoire peut introduire un biais dans les estimations des effets du traitement. Ce biais peut être évité d'une de ces manières: a) en excluant la ou les

variables irrégulières de l'analyse; b) en employant une stratégie d'imputation des données; ou c) en acceptant le biais dans l'analyse et en décrivant sa nature avant d'interpréter les résultats dans ce contexte (p. ex. les valeurs manquantes étaient plus courantes dans la population X, donc les résultats fondés sur les cas restants peuvent ne pas être généralisables à cette population). Dans ce rapport, nous utilisons toutes les trois stratégies, le cas échéant.

La première étape dans ce processus de décision est d'évaluer la prévalence des valeurs manquantes dans la base de données (c.-à-d. le pourcentage de valeurs manquantes à travers toutes les variables de la base de données) et le patron de ces valeurs manquantes. La prévalence des valeurs manquantes est évaluée dans le cadre d'analyses descriptives de base qui ne requièrent aucune explication supplémentaire. En revanche, la méthode utilisée afin de déterminer le patron de données manquantes exige plus d'explications.

La question la plus fondamentale est de savoir si les valeurs manquantes sont prévisibles ou non. Comme nous l'avons déjà vu, les valeurs manquantes imprévisibles ou « aléatoires » sont classées comme MCAR. Une condition requise pour démontrer MCAR est d'établir que la relation entre l'occurrence des valeurs manquantes et des valeurs observées des variables dans la base de données est statistiquement nulle. On peut vérifier cette propriété en faisant des tests d'association indépendants pour chaque variable d'une base de données. Dans le cas des variables continues, on peut également recourir au test MCAR de Little. Si le test MCAR ou un autre test d'association fait état de valeurs manquantes systématiquement, on peut alors rejeter l'hypothèse de travail d'une distribution des valeurs manquantes de type MCAR et opter pour une hypothèse moins restrictive: la distribution des valeurs manquantes est en réalité de type MAR.

Pour qu'un patron de données manquantes soit considéré comme MAR, deux conditions doivent être respectées: a) les valeurs manquantes doivent être distribuées de façon non aléatoire et b) il doit être possible de prédire la valeur à attribuer à la valeur manquante. La première condition est remplie lorsque la ou les variables de la base de données permettent de prédire l'incidence de valeurs manquantes; la deuxième condition est remplie lorsque de forts prédicteurs de la variable mesurée au moyen des valeurs manquantes existent dans la base de données. Lorsque la première condition est remplie, mais non la deuxième, la distribution de données manquantes est de type MNAR. Dans ce cas, nous n'avons pas d'autres options que de décrire le profil des données manquantes de façon suffisamment détaillée pour que l'on comprenne bien les limites relatives à la validité de l'étude.

Les problématiques que pose l'évaluation des valeurs manquantes ont maintenant été examinées. Dans la section suivante, nous présentons les résultats de l'analyse des valeurs manquantes effectuée en prévision des analyses d'impact du projet Capacité d'apprentissage. Tout d'abord, nous rapportons les résultats d'une analyse quantitative de l'attrition des participants, après quoi nous présentons le résultat d'une analyse des valeurs manquantes complètement en raison d'une incapacité à évaluer un enfant ou à recevoir un sondage. Nous présentons enfin une analyse des valeurs manquantes causées par la langue d'administration de l'ÉPE-AD. Dans chaque analyse, nous décrivons les valeurs manquantes sous deux aspects, d'abord en fonction de leur répartition à travers les groupes expérimentaux, puis en fonction de leur relation avec d'autres covariables dans la base de données.



### **Valeurs manquantes dues à l'attrition**

Les incidents d'attrition sont décrits à la section 2.4.1 du chapitre 2 (Méthodologie). En tout, 14 familles se sont retirées de l'étude à la fin de la deuxième année. À la fin de la première année, 10 familles ont quitté l'étude, ce qui représente un taux d'attrition de 3,9 % par rapport à l'échantillon original de 254 familles recrutées au sein des quatre communautés retenues pour les analyses d'impact. La deuxième année, le taux d'attrition a grimpé à 5,5 % de l'échantillon des quatre communautés. Ce pourcentage est faible comparativement à d'autres études d'évaluation (p. ex. Rogers, Fernandez, Thurber & Smitley, 2004), ce qui est associé à un faible risque de biais potentiel.

Les basses fréquences d'attrition n'ont pas permis de mener un test formel pour voir si la rétention des participants était étroitement liée à l'appartenance au groupe expérimental. Toutefois, une analyse descriptive des fréquences ne révèle pas l'existence d'une relation systématique. En tout, cinq familles du groupe programme en garderie, deux familles du groupe témoin en garderie ainsi que quatre familles du groupe témoin hors garderie ont quitté l'étude. Les trois autres participants ont quitté l'étude avant l'évaluation de base et ils étaient donc exclus de l'analyse par groupe expérimental. Une analyse semblable n'a pas révélé de relation systématique avec la composition linguistique du foyer. Les 14 cas d'attrition étaient répartis de façon relativement égale à travers les catégories : endogame francophone (5), endogame anglophone (2), exogame (5) et bilingue (1). La répartition des sexes était aussi semblable, soit huit garçons et six filles.

Pour les variables qui ont été mesurées au moins sur une échelle ordinale ou à un niveau de mesure encore plus précis, nous avons calculé les coefficients de corrélation de rang tau ( $\tau$ ) de Kendall afin de dégager les caractéristiques associées à l'attrition (pour la liste de covariables, voir la section 4.4). Cette analyse a révélé des liens statistiquement significatifs mais faibles entre l'attrition et les variables suivantes: score de base pour le domaine Langue et communication (N=251,  $\tau = -,125$ ); langue des activités d'alphabétisation (N=254,  $\tau = -,116$ ); langue parlée par la mère à l'enfant (N=251,  $\tau = -,139$ ); et langue de garde de l'enfant de 0 à 12 mois (N=244,  $\tau = -,140$ ). Nous rapportons les statistiques descriptives associées à ces variables au tableau 4.1 pour les cas d'attrition et au tableau 4.4 pour les cas retenus.

Le patron des résultats dans les deux tableaux indique que les enfants et les familles qui ont obtenu un score inférieur pour ce qui est de l'utilisation de la langue française étaient plus susceptibles que les autres de quitter le projet. Le faible taux d'attrition et la distribution égale entre les groupes expérimentaux signifie que la menace qui pèse sur la validité interne et externe de l'étude est minimale.

**Tableau 4.1 : Statistiques descriptives des cas d'attrition des variables corrélées avec l'attrition**

<b>Construit mesuré (étendu)</b>	<b>Moyenne</b>	<b>ET</b>	<b>N</b>
Score de base Langue et communication (score brut)	10,00	8,69	11
Langue des activités de littératie (1-5)	3,08	1,45	14
Langue parlée à l'enfant (mère) (1-3)	2,14	,77	14
Langue de garde (0-12 mois) (1-3)	2,21	,80	14

### ***Valeurs manquantes dues aux données manquant complètement (évaluation ou sondage)***

Les taux de réponses généraux pour les deux sondages et pour l'ÉPE-AD sont présentés à la section 2.5.1 du chapitre 2 (Méthodologie). Nous discutons ci-dessous du profil des valeurs manquantes pour l'évaluation, avant d'explorer le patron des données manquantes pour les sondages. L'information présentée est fondée sur une analyse des données recueillies durant la première année du projet. L'échantillon exclut les enfants qui ont quitté le projet à la fin de la première année.

### **Évaluations des enfants**

Le taux de réponses global fut très élevé pour toutes les évaluations, ce qui signifie que les cas de données manquant complètement étaient trop peu fréquents pour permettre de vérifier de quelle manière les données manquantes étaient liées à la participation à un groupe expérimental (dimension des cellules < 5). Une analyse descriptive nous révèle que les valeurs manquantes étaient réparties également à travers les groupes à l'étude quelle que soit l'évaluation prise en considération, avec trois évaluations manquantes ou moins par groupe (non rapportées). Le tableau 4.2 illustre la distribution des valeurs manquantes groupées sur les quatre premières périodes d'évaluation.

Nous avons utilisé une diversité de résultantes liées au développement comportemental chez l'enfant, de variables sociodémographiques, linguistiques, sociolinguistiques et parentales dans une tentative de prédire le patron des évaluations manquantes. L'analyse fondée sur les indicateurs de développement comportemental (imputés tel que décrit à la section 4.2.3) a révélé que les enfants étaient plus susceptibles d'avoir sauté au moins une évaluation durant la première année s'ils avaient obtenu un score faible aux échelles Langue et communication (N=244,  $\tau=-,17$ ), Vocabulaire expressif (N=228,  $\tau=-,14$ ) ou Vocabulaire réceptif (N=229,  $\tau=-,14$ ) lors de la première évaluation. De plus, une analyse employant des variables qui provenaient des sondages auprès des parents indiquait que les enfants de familles monoparentales (N=244,  $\tau=-,14$ ), ceux dont le revenu familial rapporté était plus faible (N=244,  $\tau=-,19$ ), dont le niveau d'éducation de la mère était inférieur (N=244,  $\tau=-,12$ ) et ceux qui disaient s'identifier moins à la communauté francophone (N=243,  $\tau=-,13$ ) étaient plus susceptibles de sauter au moins une évaluation. Par conséquent, il existe des données probantes suggérant que le patron qui donne lieu à ces données manquantes (c.-à-d. les évaluations manquantes) est systématique. Ce problème a été facilement résolu puisque les mêmes instruments ont été administrés plusieurs fois aux enfants, établissant d'excellents prédicteurs des valeurs non observées. En résumé, le profil des valeurs manquantes est MAR, ce qui justifie leur remplacement par imputation.

**Tableau 4.2 : Nombre d'évaluations manquantes durant la première année du projet en fonction de l'appartenance au groupe expérimental**

Évaluations manquantes	Groupe programme en garderie	Groupe témoin en garderie	Groupe témoin hors garderie
Aucune	70	89	72
1 manquante	2	6	3
2 manquantes	0	1	1
% manquantes	2,8 %	7,8 %	5,6 %

*Note* : Évaluations manquantes durant la première année de l'étude en fonction de l'appartenance au groupe expérimental. Pour toute période de temps donnée, le pourcentage manquant par groupe est < 1 %.

### **Sondages des parents**

Les données manquant complètement ne sont problématiques que dans le cas des sondages de suivi auprès des parents. Comme on le voit à la section 2.4.1, tous les sondages auprès des parents ont été reçus pour la période de base. Dans le cas des sondages de suivi, le nombre de sondages manquants pour chaque période de temps était relativement petit (à l'instar des évaluations), ce qui a limité encore une fois le type d'analyse pouvant être mené pour décrire la répartition des valeurs manquantes (p. ex., dimensions des cellules < 5 quand une variable de groupement est utilisée).

Comme pour les évaluations, nous nous limitons à une description simple des valeurs manquantes en tant que fonction du groupe expérimental. En prenant l'appartenance aux groupes expérimentaux de base comme point de référence, nous observons huit sondages manquants pour le groupe témoin hors garderie, quatre sondages manquants pour le groupe programme en garderie et trois sondages manquants pour le groupe témoin en garderie pendant la première année du projet. Comme l'objectif principal des sondages était de suivre l'évolution des variables raisonnablement stables dans le temps (p. ex. la composition des familles), il était possible d'inférer, par interpolation, une estimation raisonnable de la valeur du score non observé. Cette façon de procéder ne pouvait s'appliquer à toutes les variables, ce qui signifie que certains cas manquant complètement ont été classés comme ayant des valeurs manquant partiellement à la suite de cette étape d'imputation (voir ci-dessous).

### ***Valeurs manquantes dues à la langue d'administration (évaluation seulement)***

Une autre source de données manquantes pour la version française de l'ÉPE-AD était la langue dans laquelle le test était administré. Toutes les échelles, sauf Langue et communication, qui devait être administrée en français à tous les enfants (puis en anglais pour certains), étaient touchées. Ce type de valeur manquante était évidemment non aléatoire parce que fonction directe d'une règle de décision fondée sur la capacité de l'enfant de communiquer en français. La seule question qui reste à déterminer est de savoir si les données manquantes sont systématiquement associées au groupe expérimental.

Le nombre total d'enfants qui ont complété les autres échelles de l'ÉPE-AD en anglais était de 28, 31, 27, 20 et 44 pour les évaluations 1 à 5. Par la suite, toutes les évaluations ont été menées en français. Les tests statistiques de khi carré indiquaient que le choix de la langue d'administration ne variait pas de façon significative en fonction de l'appartenance au groupe

expérimental pour toute période d'évaluation, en employant un seuil de signification alpha de 0,05 (la série complète de résultats est disponible sur demande).

Précédemment, la SRDC a décidé que les versions anglaise et française de l'ÉPE-AD ne pourraient être combinées aux fins d'analyse (voir la note 45). En raison du petit nombre d'observations en anglais, nous avons choisi d'analyser seulement les scores provenant de la version française du test. La validité externe de l'expérience peut être influencée par cette décision étant donné qu'elle résulte en un échantillon dont l'éventail de compétences en langue française est restreint (p. ex., à cause de la troncation de l'extrémité inférieure de la distribution). Le patron de résultats obtenu pour les échelles touchées ne pourra par conséquent qu'être reproduit ultérieurement avec un échantillon qui, à l'instar de l'échantillon actuel, sous-représente l'extrémité inférieure de la gamme d'habiletés. La menace à la validité interne est minimale étant donné que les valeurs manquantes en question sont distribuées également entre les groupes à l'étude. Néanmoins, les valeurs manquantes ont été imputées chaque fois que possible pour réduire au minimum la possibilité de biais (voir Section 4.2.3).

#### ***Valeurs manquantes dues aux données manquant partiellement (sondage seulement)***

Certaines questions du sondage ont été posées à plusieurs reprises au cours de l'étude. Cette redondance a permis d'estimer directement et exactement plusieurs valeurs manquantes, ainsi que d'entraîner une réduction importante du nombre de valeurs manquantes dans la base de données. La stratégie fait l'objet d'une description plus détaillée à la section 4.2.3. Malgré ces mesures, plusieurs des 21 variables sélectionnées pour les analyses d'impact comportaient des valeurs manquantes (pour la liste complète, voir la section 4.4). Nous décrivons le patron de ces données manquantes plus en détail dans cette section.

Une seule variable montrait un taux de données manquantes supérieur à 5 %, soit un taux de 14,3 % (N=244): la variable Vitalité (c.-à-d. la vitalité de la culture francophone dans la communauté aux yeux du répondant).<sup>42</sup> Aucune autre variable n'avait plus de 2 % de données manquantes. Le score de base Langue et communication accompagnait les variables du sondage dans un test formel du profil des données manquantes. Dans le cadre de cette analyse, nous avons spécifié le groupe à l'étude et le sexe comme variables catégoriques et traité les autres variables comme continues. Le test MCAR de Little indiquait que les valeurs manquantes réparties aléatoirement (c.-à-d. l'hypothèse nulle) devraient être rejetées au niveau alpha 0,05,  $\chi^2(288)=337,31, p=,02$ .

Les tests-t comparant les cas avec valeurs manquantes aux cas sans valeurs manquantes montrent que c'est surtout la Vitalité qui a entraîné ce résultat. Selon des tests-t indépendants, les cas avec des valeurs manquantes pour la variable Vitalité diffèrent du reste de l'échantillon de plusieurs façons: a) revenu plus faible, 7,21 c. 5,92,  $t(42)=2,4$ ; b) niveau d'éducation inférieur de la mère, 3,21 c. 2,77,  $t(42)=2,7$ ; c) français moins parlé par l'enfant, 15,96 c. 13,12,  $t(42)=2,3$ ; et d) scores de base en Langue et communication inférieurs, 19,25 c. 15,01,  $t(26)=2,0$ . À la lumière de ces résultats, nous avons décidé d'exclure la Vitalité de l'analyse, parce qu'elle ne pouvait servir qu'à effectuer des ajustements visant à éliminer l'effet des différences actuelles entre le groupe témoin en garderie et le groupe témoin hors garderie (le groupe programme en garderie n'est pas affecté). Exclure la variable Vitalité représente donc une perte minime d'information,

---

<sup>42</sup> La valeur arbitraire de 5 % est un seuil de coupure conventionnel pour signaler les variables problématiques (Tabachnick & Fidell, 2007, p. 63).

tout en simplifiant grandement les analyses d'impact. Une fois la Vitalité exclue, le test MCAR de Little n'a pas réussi à rejeter l'hypothèse des valeurs manquantes réparties aléatoirement. Cela s'est avéré vrai pour l'échantillon total,  $\chi^2(150)=151,41, p=,45$ ; mais aussi lorsque l'hypothèse a été testée pour le groupe programme en garderie seulement,  $\chi^2(74)=75,99, p=,55$ . Nous concluons que les autres valeurs manquantes parmi les covariables sont réparties aléatoirement (MCAR) et peuvent donc être aménagées dans les analyses grâce à la méthode *listwise deletion* sans trop affecter la validité externe ou interne de l'étude.

### 4.2.3 Stratégie d'imputation des données

La stratégie servant à traiter les valeurs manquantes changeait en fonction de la variable traitée. Dans tous les cas, nous avons misé sur le fait que notre étude s'appuyait sur un devis à mesures répétées. Le fait que nous les mêmes variables (ou des variables très semblables) auprès des mêmes personnes de façon répétée au cours de l'étude nous a permis d'estimer avec beaucoup de précision la valeur qui aurait été obtenue si les valeurs manquantes avaient été observées. Nous décrivons dans des sections séparées la stratégie d'imputation employée pour les mesures d'évaluation (c.-à-d. les résultantes pour les enfants) et pour les variables de sondage (c.-à-d. les covariables et les résultantes pour les parents).

#### ***Imputation des résultantes des enfants***

Nous avons adopté une stratégie d'imputation pour les échelles de l'ÉPE-AD qui capitalisait sur la nature longitudinale de l'étude. L'ÉPE-AD originale a été administrée en français ou en anglais durant les cinq premières évaluations. Par la suite, la version révisée de l'ÉPE-AD a été administrée en français seulement, avec un ensemble d'items différent. Aucune règle de départ n'a été appliquée lors de l'ÉPE-AD révisée, éliminant la nécessité d'imputer les données manquantes pour chacun des items. Le problème des valeurs manquantes ne s'applique qu'aux cinq premières évaluations.

Nous avons divisé la procédure d'imputation en deux étapes. La première consistait à imputer les valeurs manquantes au niveau des items par suite de l'application de la règle de départ. La seconde étape consistait à imputer les valeurs manquantes au niveau des échelles au moyen d'une procédure semblable. La procédure d'imputation était basée uniquement sur les scores recueillis au moyen d'évaluations menées en français. Nous décrivons maintenant les deux étapes de cette procédure.<sup>43</sup>

#### **Imputation au niveau des items**

Comme première étape de la procédure d'imputation, nous avons remplacé toutes les valeurs manquantes au niveau des items par suite de l'application de la règle de départ, qui n'existent

---

<sup>43</sup> À noter que la procédure d'imputation employée ici sous-estime le degré de variabilité chez l'enfant qui existe vraiment dans la population, parce que l'algorithme n'ajoute pas de bruit aux estimations (Cohen et al., 2003). Dans le cas de l'imputation au niveau des items, cette critique n'est pas aussi pertinente puisqu'il s'agit d'une amélioration par rapport à la procédure d'imputation prévue à l'origine par le concepteur de l'outil, qui aurait imputé un score de 4 à toutes les valeurs manquantes à cause de la règle de départ, établissant ainsi une sous-estimation encore plus importante de la variance. Rien ne prévoyait l'imputation au niveau des échelles, et la procédure que nous avons adoptée peut faire l'objet de critiques en ce sens que le remède ici serait plus nocif que la maladie. Il est vrai que la précision des estimations de coefficients de régression risque d'être surestimée. Le potentiel de biais est le plus grand lorsque le nombre de valeurs manquantes est le plus élevé (voir le tableau 4.3). De toute manière, l'échelle Langue et communication n'est pas touchée et le biais potentiel relatif à l'estimation de la précision est atténué par l'exactitude des estimations pour les participants.

que pour les deuxième et troisième évaluations (et la cinquième évaluation, comme il est décrit ci-dessous). Nous avons d'abord imputé les valeurs de la deuxième évaluation, puis imputé les valeurs manquantes de la troisième évaluation. Toutes les valeurs manquantes ont été estimées avec succès au moyen de cette procédure (le nombre et les items affectés sont rapportés antérieurement dans ce chapitre, voir la section 4.2.1), que nous allons maintenant décrire.

L'algorithme ayant servi à imputer les scores des items manquants utilisait l'information du participant tirée uniquement de la version française de l'ÉPE-AD.<sup>44</sup> D'abord, nous avons estimé un score de gain « moyen » pour chaque combinaison enfant-item. Ces scores de gain moyen représentent les changements moyens du score observé chez l'enfant entre deux évaluations consécutives.<sup>45</sup>

Une limite de cette méthode est que la procédure d'imputation ne pouvait seulement être appliquée qu'aux enfants qui avaient complété au moins deux évaluations en français durant les cinq premières évaluations. Les valeurs observées non manquantes ont ensuite servi de points d'ancrage pour la procédure d'estimation. Par exemple, les valeurs manquantes dans la deuxième évaluation ont été imputées en ajoutant le score de gain moyen d'un item donné au score observé pour cet item durant la première évaluation. Si des renseignements manquaient dans la première évaluation, la procédure estimait alors « à rebours » à partir du score observé à la troisième évaluation, en soustrayant le score de gain moyen.

Les valeurs manquantes à cause de la règle de départ lors de la cinquième évaluation ont été imputées dans le cadre d'une procédure plus simple, soit l'imputation du score maximal pour chaque item. Comme nous l'avons déjà précisé, l'application de la règle de départ dans ce cas était fondée sur la performance antérieure de l'enfant et, en conséquence, le postulat selon lequel l'enfant aurait obtenu le score maximal sur les items était beaucoup plus raisonnable.<sup>46</sup>

Suite à l'imputation des valeurs manquantes au niveau des items, les scores des six échelles de l'ÉPE-AD ont été calculés pour chaque période d'évaluation. Les scores au niveau des échelles ont alors été soumis à une procédure additionnelle d'imputation conçue pour remplacer les valeurs non observées dues soit aux évaluations manquantes ou à l'utilisation de la version anglaise de l'instrument à une ou plusieurs reprises.

---

<sup>44</sup> L'emploi de l'information du participant fait en sorte que les attributs uniques du participant sont correctement représentés par la procédure d'imputation. La procédure n'exagère donc pas la congruence entre les scores d'un enfant donné et le reste du groupe.

<sup>45</sup> Les scores de changement moyen ont été calculés en fonction de la différence moyenne prévue entre deux périodes d'évaluation consécutives. Si l'enfant a été testé deux fois au cours de la première et quatrième périodes d'évaluation et a obtenu des scores respectifs de 4 et 14, les scores de changement moyens serait la différence observée entre les évaluations ( $14-4=10$ ), divisée par le nombre d'intervalles ou d'étapes entre les évaluations, soit, dans le cas qui nous intéresse, trois (score de changement moyen de  $10/3 = 3,33$ ). Toutes les différences entre les évaluations ont contribué au calcul, pourvu qu'elles se situent à l'intérieur de trois étapes l'une de l'autre. De plus, aucune imputation n'était effectuée si les deux scores observés pour l'enfant sur une échelle donnée s'appliquaient aux première et cinquième évaluations (c.-à-d.. quatre étapes).

<sup>46</sup> L'exception montrait 11 cas où la règle de départ n'a pas été mise en œuvre correctement sur le terrain pour l'échelle Conscience de soi et de l'environnement. Dans ces cas, l'imputation du score maximal pour chaque item n'était plus justifiée par le fait que la langue d'administration de l'ÉPE-AD avait changé de l'anglais à la quatrième évaluation au français à la cinquième évaluation. Il a été impossible d'exécuter une imputation de ces scores en fonction des réponses antérieures aux items en question, comme on l'avait fait avec la règle de départ manquant des deuxième et troisième évaluations. Les valeurs manquantes en raison de la règle de départ pour ces cas étaient donc imputées selon des estimations de régression dérivées de leur rendement sur l'échelle de conscience de soi lors de la sixième évaluation.

### Imputation au niveau des échelles

Le même algorithme d'imputation ayant servi à imputer les items a été appliqué à l'imputation des scores au niveau des échelles. Comme auparavant, l'imputation n'était possible que si un enfant avait complété l'ÉPE-AD en français au moins deux fois au cours des cinq premières évaluations. Dans un premier temps, nous avons estimé un score de gain moyen pour chaque enfant. Dans un deuxième temps, nous avons combiné ce score aux valeurs observées pour chaque enfant dans l'ÉPE-AD française pour combler, dans la mesure du possible, les lacunes d'observation.<sup>47</sup>

Plusieurs valeurs manquantes restaient pour les échelles de l'ÉPE-AD (autres que l'échelle Langage et communication), parce qu'un sous-échantillon d'enfants ont complété la plupart de leurs évaluations en anglais. Ceci a entraîné la perte de 26 enfants pour l'analyse d'impact connexe, accroissant en fait la capacité de l'échantillon en question. Le nombre de données observées, manquantes imputées et manquantes non imputées, figure au tableau 4.3, par évaluation et par groupe expérimental.

**Tableau 4.3 : Valeurs françaises imputées de l'ÉPE-AD en fonction de l'évaluation et du groupe expérimental**

	<i>Administrations de l'ÉPE-AD</i>				
	<i>1e</i>	<i>2e</i>	<i>3e</i>	<i>4e</i>	<i>5e</i>
<b>Échelle Langage et communication</b>	# imputée / n	# imputée / n	# imputée / n	# imputée / n	# imputée / n
Groupe programme en garderie	2/72	0/72	0/70	0/70	0/65
Groupe témoin en garderie	3/96	1/94	1/95	2/95	2/112
Groupe témoin hors garderie	2/76	0/78	3/79	2/79	1/63
Total	7/244	1/244	4/244	4/244	3/240
Total (% imputé)	2,9 %	,01 %	1,6 %	1,6 %	,02 %
<b>Autres échelles de l'ÉPE-AD</b>	# imputée / n (# non-imputée manquante)	# imputée / n (# non-imputée manquante)	# imputée / n (# non-imputée manquante)	# imputée / n (# non-imputée manquante)	# imputée / n (# non-imputée manquante)
Groupe programme en garderie	10/72 (3)	6/72 (5)	3/70 (4)	1/70 (4)	0/65 (0)
Groupe témoin en garderie	9/96 (6)	6/94 (5)	5/95 (4)	2/95 (4)	1/112 (1)
Groupe témoin hors garderie	1/68 (7)	0/78 (10)	4/79 (11)	2/79 (11)	2/63 (5)
Total	20/244 (16)	12/244 (20)	12/244 (19)	5/244 (19)	3/240 (6)
Total (% imputé / % non-imputé manquant)	8,1 % / 6,6 %	4,9 % / 8,1 %	4,9 % / 7,8 %	2,0 % / 7,8 %	1,2 % / 2,5 %

<sup>47</sup> Nous avons limité la procédure de manière à ce que l'imputation ne soit pas exécutée que si les deux seules observations provenaient des première et cinquième évaluations.

*Note* : Les résultats pour les « Autres échelles de l'ÉPE-AD » sont basés sur l'échelle Conscience de soi et de l'environnement. Les nombres exacts peuvent varier légèrement pour certaines des autres dimensions, mais la différence est négligeable (1-2 dans les comptes de fréquence). Les valeurs manquantes qui restent après l'imputation sont indiquées entre parenthèses. La méthode *listwise deletion*, supprimant des cas avec valeurs manquantes, entraîne l'élimination de 26 enfants des analyses effectuées sur les échelles autres que l'échelle Langage et communication.

### ***Imputation des données de sondage***

Lorsqu'il était logique de le faire, les valeurs manquantes de ces variables étaient imputées directement selon les scores des variables provenant d'autres vagues de collecte des données. Lorsqu'il n'était pas possible d'appliquer cette procédure d'imputation simple, les données manquantes ont été ignorées en raison de leur répartition MCAR. Il faut noter que l'imputation décrite plus loin a été effectuée avant de tester la répartition aléatoire des valeurs manquantes à la section 4.2.2.

Par exemple, nous avons mesuré la variable du revenu trois fois durant les deux premières années de l'étude. La plus utile de ces mesures fut obtenue avant la septième période d'évaluation, qui faisait la distinction entre 10 tranches de revenu (29 cas manquants, correspondant à 11,9 % de l'échantillon). La variable du revenu pour le sondage de base ne distinguait que six tranches de revenu et présentait un problème d'effet de plafonnement. Les deux variables étant fortement corrélées ( $r = 0,78$ ), la variable du revenu telle que mesurée avant la septième évaluation pouvait servir à effectuer une imputation par régression des valeurs manquantes observées avec la variable du revenu telle que mesuré au sondage de base.

Pour la période de base, nous considérons que la covariable la plus utile était le Continuum de français parlé par l'enfant (19 manquants, 7,8 % de l'échantillon). Cette variable a été mesurée dans plusieurs sondages de suivi, mais la mesure prise à la suite de l'exposition du participant au programme a pu être affectée, et donc ce type d'imputation n'était pas approprié. Nous avons plutôt effectué une simple imputation par régression en utilisant une variable hautement corrélée ( $r = ,81$ ) qui n'était liée d'aucune façon aux analyses rapportées ici, mais qui avait été mesurée sur une échelle semblable (c.-à-d. de l'anglais seulement au français seulement) : la langue maternelle de l'enfant.<sup>48</sup>

Les deux principales mesures dépendantes tirées des sondages étaient la fréquence et la langue des activités de littératie, respectivement. Les valeurs manquantes pour les deux variables ont été imputées suivant la procédure décrite à la section sur les résultantes pour l'enfant. Nous avons pris cette décision en reconnaissant le fait que ces variables pouvaient changer avec le temps. Grâce à cette procédure, nous avons réussi à remplacer toutes les valeurs manquantes dues aux sondages manquants (pour les taux de réponses, voir le tableau 2.4).

## **4.3 SPÉCIFICATION DES TERMES D'ERREURS**

Comme nous l'avons déjà vu, la présente étude a été conçue de manière à faire l'échantillonnage d'observations provenant de deux unités ou niveaux d'analyse. Le premier niveau concerne les garderies et le deuxième vise les participants ou, peut-être plus exactement,

---

<sup>48</sup> La corrélation très élevée entre les deux variables signifie que la possibilité d'un biais descendant dans l'estimation des erreurs-types est minime malgré le recours à l'imputation simple (Cohen et al., 2003, p.446). Le fait que la variable utilisée pour exécuter l'imputation ne soit aucunement liée aux analyses rapportées ici signifie qu'il n'y a aucun danger d'introduire la multicollinéarité dans les analyses.



la « famille ». Selon le formalisme développé au chapitre 3, on peut dire des observations qu'elles sont « groupées » en fonction de ces unités d'analyse. Le principal défi que cela pose lors d'une analyse de régression est que les estimations d'erreurs-types peuvent être sous-estimées (Moulton, 1990). Dans ce cas, les tests statistiques inférentiels seront trop sensibles aux différences entre nos groupes à l'étude. Le résultat est une inflation du taux d'erreur de type I, ce qui dans le cas présent, réfère à la probabilité de conclure incorrectement qu'une différence réelle existe entre les groupes et qu'elle risque d'être reproduite dans les études subséquentes.

Le projet dont il est question ici a fait appel à un devis à mesures répétées. Lorsque l'on fait des observations auprès des mêmes participants de manière répétitive, on induit par le fait même au fil du temps une corrélation en série dans les observations qui peut biaiser à la baisse les estimations de l'erreur-type. Les effets du traitement rapportés dans ce document reposent sur l'estimateur DD. Par construction, l'estimateur DD neutralise ce type de biais. En fait, l'estimation des erreurs-types associée à l'estimateur DD équivaut essentiellement au traitement des participants en tant qu'effet fixe.<sup>49</sup> Toutefois, même après que cette source de biais ait été neutralisée, les résidus produits par le modèle de régression peuvent encore être corrélés (Donald & Lang, 2007). La performance de chaque enfant, par exemple, pourrait être influencée par le milieu de garde où il évolue. Comme plusieurs enfants d'une même garderie participent au projet Capacité d'apprentissage, ce milieu peut induire une corrélation dans les résidus des modèles de régression. Si l'on spécifie incorrectement la structure de cette corrélation, les estimations des erreurs-types associées seront biaisées.

Nous avons adopté la stratégie suivante pour modéliser la structure de notre erreur résiduelle. D'abord, nous avons dégagé un nombre limité de spécifications de groupement plausibles. Nous avons envisagé le fait que les corrélations associées tant à la garderie qu'au participant pourraient être présentes dans les résidus. Nous avons donc tenté d'établir des spécifications pour lesquelles les erreurs-types étaient calculées en groupant les résidus en fonction de la garderie et du participant.<sup>50</sup> Des deux possibilités, le groupement de garderie a engendré les plus importantes erreurs-types. En d'autres mots, le groupement de garderie a rapporté des estimations relativement conservatrices sur la précision des estimations des effets de notre programme. Nous avons préféré sous-évaluer plutôt que surévaluer la précision de nos estimations, ce qui explique pourquoi la majorité des analyses présentent des erreurs-types calculées au niveau des garderies. Cette décision a fourni un total de 20 degrés de liberté pour les tests t impliquant les estimations DD.<sup>51</sup>

Pour l'analyse des échelles des ateliers pour parents, le groupement des erreurs-types en fonction de la garderie n'a pas été pratique en raison du petit nombre de garderies participant à cette analyse (n=6). Nous avons comparé les deux alternatives: a) groupement par individu; et b)

---

<sup>49</sup> Cette équivalence tient lorsque les données sont équilibrées, c'est-à-dire que l'on observe les mêmes participants à chaque période de temps et qu'il n'y a pas de données manquantes. Les inférences qui résultent de ces erreurs-types s'appliquent au groupe étudié plutôt qu'à une population d'individus.

<sup>50</sup> Nous avons mené des analyses où les erreurs-types étaient calculées en groupant les résidus selon la période de temps en garderie. Le calcul des erreurs-types à ce niveau visait à neutraliser la corrélation négative dans les résidus qui a été observée au fil du temps chez les individus lors de plusieurs analyses. Dans certains cas, cette stratégie a été plus conservatrice que le groupement des résidus par garderie seulement (c.-à-d. les principales analyses d'impact visant l'appartenance au groupe expérimental seulement). Dans d'autres cas, les erreurs-types engendrées étaient plus libérales (c.-à-d. les analyses d'impact où les heures passées en garderie étaient incorporées dans les spécifications du modèle). Par souci de cohérence, nous avons décidé de calculer les erreurs-types au niveau des garderies pour toutes les analyses fondées sur les résultantes pour les enfants.

<sup>51</sup> Le nombre exact de degrés de liberté peut varier en fonction de l'analyse.

erreurs-types robustes sans groupement. La seconde méthode a produit des tests de signification statistique plus conservateurs avec les plus importantes estimations d'erreurs-types. Nous rapportons les résultats fondés sur les estimations les plus conservatrices.

#### 4.4 Liste des variables de contrôle

La validité interne de toute comparaison entre les conditions de traitement dépend de la démonstration préalable que les trois groupes sont similaires quant à toutes les dimensions, sauf celle qui est manipulée par le programme testé. Dans la recherche non expérimentale (et même la recherche expérimentale), la validité interne est maximisée en effectuant un certain type de contrôle à l'aide de covariables. L'idée ici est de mesurer toutes les variables pertinentes susceptibles d'être liées à une résultante donnée. Quand de telles mesures sont disponibles, on rend les groupes à comparer aussi semblables que possible soit en jumelant les participants, soit en faisant un contrôle statistique (p. ex. Behrman, Cheng & Todd, 2004). Il nous fut impossible de jumeler les participants en raison des limites que posait la taille de notre échantillon. Nous avons plutôt adopté une stratégie qui consistait à entreprendre le contrôle statistique sélectif des covariables quand nous constatons qu'elles étaient liées à la condition de traitement.<sup>52</sup> Il en résulte que les effets du programme rapportés aux chapitres 5 et 6 devraient être interprétés comme l'effet programme selon les covariables incluses dans le modèle.

Cette stratégie suppose un processus en deux temps où les variables de contrôle potentielles sont d'abord dégagées, puis insérées dans les analyses d'impact. Une analyse exhaustive des différences entre les groupes basées sur les variables mesurées durant le sondage de base apparaît dans le *Rapport de référence* (Legault et al., 2014). Cependant, cette analyse était fondée sur l'échantillon total qui a accepté de participer au projet, alors que les analyses rapportées ici sont fondées sur le sous-échantillon de participants des quatre communautés sélectionnées pour l'analyse des effets sur les résultantes se rapportant au développement des enfants (l'analyse des résultantes liées aux parents était fondée sur les six mêmes collectivités comme dans le *Rapport de référence* (Legault et al., 2014). Il est possible que pour le présent échantillon de participants, un ensemble différent de covariables soit nécessaire pour assurer la validité interne de l'étude.

Nous avons exploré cette possibilité en analysant de nouveau les données de sondage de base avec le sous-échantillon de participants qui fait l'objet de la présente analyse. Nous ne nous attendions pas à ce que le résultat soit très différent des résultats rapportés dans le *Rapport de référence* (Legault et al., 2014). Néanmoins, nous avons jugé prudent de ne pas prendre cette cohérence pour acquise. De surcroît, avec l'extension de la base de données au fur et à mesure que de nouvelles collectes de données sont effectuées, la possibilité se présente de tester les différences entre les groupes à chaque vague de collecte de données. Ce faisant, on reconnaît le fait que certaines variables changent avec le temps (p. ex. variables de la composition de la famille comme la taille du ménage), y compris la composition des groupes expérimentaux (p. ex. dû à l'attrition ou aux changements de groupe). Cette variation temporelle soulève la possibilité que des différences apparaissent au sein des groupes avec le temps, différences qui n'étaient pas apparentes lors du sondage de base. La stratégie que nous venons tout juste de décrire suppose

---

<sup>52</sup> En sélectionnant nos covariables selon leurs liens observés avec les résultantes, nous présumons que les estimations observées sont valides (c.-à-d. non biaisées par les erreurs d'échantillonnage ou de mesure). Le contrôle statistique (et même l'appariement des cas) ne peuvent avoir pour effet que d'ajuster les différences qui sont observées dans les variables mesurées. Les différences réelles qui ne sont pas décelées (p. ex., en mesurant imparfaitement la variable prise en défaut) ne peuvent être contrôlées de cette manière.

un grand nombre de tests statistiques, un pour chaque combinaison de variables et période de temps (variables K x période de temps T). Rapporter les résultats détaillés de cette analyse préliminaire ajouterait peu de valeur au rapport, tout en l'allongeant considérablement. Pour rendre la présentation plus efficace et plus limpide, nous présentons la synthèse des résultats dans les sections qui suivent.

### ***Différence entre les groupes : Première et deuxième années***

La stratégie suivante a servi à dégager les variables potentiellement confondues avec le groupe expérimental. Les variables confondues sont des caractéristiques sur lesquelles les groupes expérimentaux diffèrent. Elles constituent un problème puisqu'elles fournissent des explications alternatives aux effets observés du programme, alors que le but de la recherche est de les éliminer.

Nous avons par conséquent testé les associations entre les groupes expérimentaux et les covariables en recherchant les variables confondues. Nous avons répété les tests en question pour chaque période, pour les raisons expliquées dans la section précédente. Suivant le cas, nous avons employé les tests statistiques de khi carré ou l'analyse de la variance (ANOVA) pour tester les différences entre les groupes. Pour les fins des analyses d'impact, toutes les variables étaient considérées comme étant continues, sauf celles de la communauté, du type linguistique du foyer (p. ex. exogame français), du foyer monoparental et du sexe. La notation de toutes les autres variables était codée pour refléter une progression ordonnée de valeurs ascendantes dans une dimension donnée, traitées comme si elles étaient mesurées sur une échelle de variation par intervalle. Nous présentons au tableau 4.4 un résumé de ces analyses pour les variables qui: a) répondaient aux critères d'inclusion en tant que covariables; et b) n'étaient pas redondantes avec les variables déjà incluses dans les analyses.

**Tableau 4.4 : Statistiques descriptives, différences observées entre les groupes expérimentaux quant aux caractéristiques observées et leur association avec les résultantes des enfants**

Type de covariable	Statistiques (Niveau de base)		Variable confondue au groupe expérimental?			Pertinence <sup>b</sup>
	Moyenne	ET	G1 c. G2	G1 c. G3	G2 c. G3	Liée aux VD ?
<b>Variables continues</b>						
<b>Sociodémographique</b>						
Taille du ménage	4,00	,92	NS	G1 < G3	NS	Oui
Nombre d'enfants plus vieux	,68	,80	NS	NS	NS	Oui
Nombre d'enfants plus jeunes	,38	,50	NS	G1 < G3	NS	Oui
Âge de l'enfant (mois)	38,36	3,89	NS	NS	NS	Oui
Revenu (échelle 1-10)	7,02	2,63	NS	NS	NS	Oui
Éducation (mère, 1-4)	3,15	,81	NS	NS	NS	Oui
Âge à la première naissance (mère, échelle 1-5)	2,37	,93	G1 > G2	G1 > G3	NS	Oui
<b>Linguistique et sociolinguistique (Des valeurs plus élevées signifient de meilleures « en francophonie »)</b>						
Continuum linguistique (enfant, 5-20)	15,47	5,79	NS	G1 < G3	G2 < G3	Oui
Langue parlée à l'enfant (mère, échelle 1-5)	4,11	1,37	NS	G1 < G3	G2 < G3	Oui
Langue des activités de littératie (échelle 1-5)	4,03	1,17	NS	G1 < G3	G2 < G3	Oui
Langue de garde de jour (0-12 mois) (échelle 1-3)	2,61	,70	NS	G1 < G3	NS	Oui
Engagement envers la culture francophone (échelle 1-3)	2,47	,73	NS	G1 < G3	NS	Oui
Vitalité (échelle 4-24)	16,41	5,64	NS	NS	G2 < G3	Oui
<b>Échelles parentales</b>						
Fonctionnement familial (échelle 8-32)	29,89	3,02	NS	NS	NS	Oui
Dépression (mère) (échelle 8-32)	10,48	3,39	NS	NS	NS	Oui
Style parental autoritaire (échelle 4-16)	12,98	2,10	NS	NS	NS	Oui
Style parental positif (échelle 5-25)	23,30	1,79	NS	NS	NS	Oui
<b>Variables nominales</b>						
Foyer monoparental <sup>c</sup>	9,4 % de l'échant.		ND	ND	ND	Non
Statut d'immigrant	4,1 % de l'échant.		NS	NS	NS	Oui
Sexe	47,5 % garçons		NS	NS	NS	Oui
Type de foyer (exogame, etc.)	56,6 % endog. F, 2,5 % endog. A <sup>d</sup> , 23 % exog, et 16,8 % bilingue		NS <sup>d</sup>	G3>endog. F <sup>d</sup> G3<bilingue <sup>d</sup>	G3>endog. F <sup>d</sup> G3<bilingue <sup>d</sup>	Oui

**Note :** Les statistiques descriptives rapportées visent l'échantillon de quatre communautés. Les variables de la composition de la famille s'appuient sur l'évaluation de base (et peuvent varier quelque peu avec le temps). L'association entre le groupe expérimental et les variables continues est déterminée par le test F de Welch et les tests post-hoc de Ryan-Einot-Gabriel-Welsch (robuste à N inégal, variance hétérogène). Nous énumérons seulement les variables qui: a) étaient fortement associées au groupe expérimental; ou b) étaient fortement associées à l'une des variables dépendantes. VD=variables dépendantes; NS=qui n'est pas statistiquement significatif en employant un seuil de signification alpha de 0,05; ND=information non disponible due aux cellules ne contenant pas un nombre suffisant de cas (<5); endog. F=endogame francophone; endog. E=endogame anglophone; et exog.=exogame.

<sup>a</sup> G1 désigne le groupe programme en garderie, G2 le groupe témoin en garderie, G3 le groupe témoin hors garderie. Les différences entre les groupes sont rapportées si elles sont statistiquement significatives pour toute période d'évaluation (première à septième) ayant eu lieu durant les deux années de l'étude.

<sup>b</sup> Covariable corrélée significativement à au moins une VD au seuil de signification alpha de 0,05 selon la régression/ANOVA lors d'au moins une évaluation au cours de la première année du programme.

<sup>c</sup> La variable Vitalité a été exclue des analyses d'impact à cause d'un problème avec les valeurs manquantes. La variable dichotomique du foyer monoparental a été exclue en raison d'une asymétrie excessive. Aucune de ces deux variables n'avait une influence marquée sur les résultats.

<sup>d</sup> Les participants exogames anglophones ont été exclus à cause du trop petit nombre de cas contenus dans les cellules associées à cette catégorie.

Dans le tableau 4.4, nous ne rapportons que les associations observées entre une covariable donnée et l'appartenance au groupe expérimental pour l'une ou l'autre des sept périodes d'évaluation qui font l'objet de l'étude dont il est question ici. De plus, nous avons envisagé la possibilité que l'importance des variables en question pour le développement de l'enfant ait changée considérablement au cours de la première année ou de la deuxième année de l'étude (non rapportée dans le tableau). Pour tester cette idée, nous avons estimé une série de modèles de régression pour chaque covariable énumérée dans le tableau 4.4 en fonction de la spécification décrite dans le modèle 1 de la section 5.1.2. Cette spécification initiale a été adaptée de sorte que l'interaction entre Temps et Covariable soit utilisée dans la spécification plutôt que l'interaction entre Temps et Groupe expérimentaux. À l'aide de cette spécification, nous avons ré-estimé le modèle pour chaque covariable et chaque mesure de résultante sur l'ÉPE-AD. À la fin, seulement une des interactions testées a été retenue pour les analyses d'impact en raison de son interaction significative avec le temps, spécifiquement la langue de garde (de 0 à 12 mois), pour la variable Langage et communication: Wald F (3, 18) = 3,07, p = 0,05.<sup>53</sup> Lorsque cette variable était insérée dans les modèles de régression pour les analyses d'impact telles qu'elles sont présentées dans les chapitres 5 et 6, elle était invariablement un effet principal et interagissait avec le Temps. Elle visait à corriger toute différence de pente qu'on aurait observée entre les groupes expérimentaux en raison de cette caractéristique (Langue de garde (de 0 à 12 mois), en l'absence de traitement. Cette question a été abordée dans la section 3.2.3.

### ***Autres facteurs confusionnels possibles***

En plus des associations entre les covariables et le groupe expérimental, nous avons effectué des tests pour déterminer si le groupe programme en garderie (G1) et le groupe témoin en garderie (G2) différaient beaucoup en ce qui concerne les heures passées en garderie par semaine. Les analyses d'impact ont vérifié si l'efficacité prédictive de cette information variait en fonction du groupe expérimental. La validité de ces tests dépend du respect du postulat selon

---

<sup>53</sup> Une autre variable linguistique qui a interagi de manière significative avec le temps était le Continuum de linguistique parlé par l'enfant. Pour réduire au minimum la redondance dans le modèle de régression, cette variable n'a pas été utilisée dans les analyses principales rapportées dans les sections 5.1.2 et 6.1 Cette décision n'influence en rien l'essentiel des conclusions à tirer des analyses rapportées. Ce fait a été confirmé par une série d'analyses non rapportées qui incluaient comme covariable l'interaction entre le Temps et cette variable.

lequel la distribution de cette variable est équivalente pour les deux groupes. Le test F de Welch indiquait que c'était le cas pour la période de quatre mois précédant :

- la première évaluation : G1 : 30,80 (ET=10,76) c. G2 : 29,35 (ET=11,50),  $F(1, 158) = 0,70$ ;
- la seconde évaluation, G1 : 27,32 (ET=8,60) c. G2 : 26,88 (ET=9,36),  $F(1, 158) = 0,10$ ;
- la troisième évaluation, G1 : 30,01 (ET=8,92) c. G2 : 28,61 (ET=10,29),  $F(1, 158) = 0,87$ ;
- et la quatrième évaluation, G1 : 19,98 (ET=9,80) c. 18,31 (ET=9,54),  $F(1, 142)=1,16$ .

Ces résultats nuls ont été confirmés par des tests U de Mann-Whitney non paramétriques. Nous concluons que la distribution des variables de dosage est statistiquement équivalente et que les tests des interactions de dosage avec le groupe expérimental ne sont pas biaisés par ces facteurs confusionnels potentiels.

Les différences individuelles entre les éducatrices constituent un autre facteur qui peut avoir un effet sur les différences entre les garderies programme et les garderies témoins au niveau de base. Un total de 25 femmes ont fait office d'éducatrice principale de la petite enfance au sein des garderies participantes pour la période de référence dans les quatre communautés qui ont fait l'objet des analyses (8 dans le groupe programme et 17 dans le groupe témoin en garderie). Le nombre total d'années d'expérience que ces femmes ont déclaré n'a pas varié de manière significative entre les groupes de garderies: groupe programme en garderie = 6,00 (ET = 2,71) par rapport à la moyenne du groupe témoin en garderie = 7,19 (ET = 8,22) selon le test-t de Welch, [ $t(20,19) = 0,523$ ] ou selon le test de la somme des rangs U de Mann-Whitney = 50,50, [ $Z = -1,02$ ]. La taille minimale de la cellule était insuffisante pour tester si les caractéristiques restantes ont varié en fonction du groupe de garderie ( $n < 5$ ).

Enfin, à la fin de la première année du programme (octobre 2008), environ 52 % des enfants de l'échantillon étaient inscrits à l'école à temps plein ou à temps partiel. Si l'inscription avait été inégale entre les groupes expérimentaux, cette inégalité aurait représenté une difficulté pour la validité interne de l'étude, ce qui aurait alors exigé un contrôle statistique. En octobre 2008 (quatrième évaluation), il n'y avait pas de différences systématiques pour ce qui est de l'inscription à l'école selon l'application d'un test de khi carré [ $\chi^2(2) = 2,11$ ]. En février 2009 (cinquième évaluation), l'inscription à l'école a bel et bien varié de manière significative en fonction du groupe expérimental [ $\chi^2(2) = 7,08, p = 0,03$ ]. L'effet était attribuable au degré plus faible d'inscription à l'école parmi le groupe témoin hors garderie (38 %) que parmi le groupe témoin en garderie (57 %), ou le groupe programme en garderie (60 %). Cette relation ne peut s'expliquer par une représentation disproportionnée des enfants du groupe témoin hors garderie à Edmundston où il n'y avait pas de maternelle pour les enfants de quatre ans. Une explication plausible est que les parents qui n'ont pas à payer de garderie en dehors de la maison sont moins motivés à envoyer leur enfant à la prématernelle. Dans un cas comme dans l'autre, l'inscription à l'école a été entrée comme covariable pour contrôler statistiquement tout biais induit par les différences observées. Le biais potentiel n'aurait eu d'effet que sur les estimations de l'effet du programme pour le groupe témoin hors garderie.

## 4.5 REPRÉSENTATIVITÉ DE L'ÉCHANTILLON : PROJET CAPACITÉ D'APPRENTISSAGE VS. EVMLO (ENQUÊTE SUR LA VITALITÉ DES MINORITÉS DE LANGUE OFFICIELLE)

La question de la validité externe de l'échantillon du projet Capacité d'apprentissage a été abordée précédemment dans le chapitre 5 du rapport *Capacité d'apprentissage dans les communautés francophones en situation minoritaire : Rapport de référence* (Legault et al., 2014, ci-après nommé *Rapport de référence*), où une série de comparaisons détaillées ont été effectuées qui mettaient en cause l'échantillon projet Capacité d'apprentissage et de l'Enquête sur la vitalité des minorités de langue officielle (EVMLO). L'objectif du chapitre en question était tout simplement de répondre à la question suivante : si le programme de garderie et les ateliers familles étaient étendus à toute la minorité francophone dans les communautés participant au projet, est-ce que les effets observés seraient semblables à ceux obtenus dans le projet Capacité d'apprentissage? Cette question est reprise ici avec les quatre communautés qui ont été retenues pour les analyses d'impact sur le développement de l'enfant. Si les analyses comparant le projet Capacité d'apprentissage et l'EVMLO doivent servir la fin pour laquelle elles ont été conçues, il est essentiel que les analyses d'impact soient fondées sur un échantillon comparable de participants.

Les stratégies d'échantillonnage employées dans le cadre du projet Capacité d'apprentissage et de l'EVMLO limitent la comparabilité des deux échantillons de plusieurs points de vue. Comme on l'a expliqué dans le *Rapport de référence*, la procédure d'échantillonnage du projet Capacité d'apprentissage est plus restrictive que celle de l'EVMLO. Des différences démographiques importantes entre les deux échantillons pourraient alors exister. Afin de veiller à ce qu'un échantillon suffisamment grand soit extrait de la base de données EVMLO, les données provenant d'enfants de 3 à 5 ans ont été prises en considération dans les analyses comparatives. Ce fait contraste avec l'âge moyen de l'échantillon projet Capacité d'apprentissage lorsque le sondage de base a été administré, qui était d'environ 3 ans. Enfin, les participants du projet Capacité d'apprentissage ont été sélectionnés d'une manière non probabiliste en fonction de la fréquentation d'une garderie, ce qui signifie que la distribution géographique de l'échantillon est localisée au sein de communautés données. Par contraste, la distribution de l'échantillon EVMLO est plus diffuse du point de vue géographique en raison du fait que l'étude dont il est question ici recourt à l'échantillonnage aléatoire. L'interprétation de toute différence observée entre les échantillons doit être nuancée à la lumière des différences liées à la méthodologie.

Les analyses rapportées dans le *Rapport de référence* (Legault et al., 2014) ont été fondées sur un échantillon projet Capacité d'apprentissage substantiellement plus grand, qui comprenait des enfants provenant des communautés d'Edmonton (AB) et de Saint-Jean (N.-B.). Dans les analyses en question, certaines différences ont été observées entre les échantillons projet Capacité d'apprentissage et EVMLO quant à des caractéristiques sociodémographiques et linguistiques. Il reste à voir si le même patron de résultats s'obtient lorsque les communautés d'Edmonton et de Saint-Jean sont exclues de l'échantillon projet Capacité d'apprentissage, comme elles le sont dans les analyses d'impact décrites aux chapitres 5 et 6. Cette question a été abordée ici à l'aide de la même stratégie analytique qui a été employée dans le *Rapport de référence* (Legault et al., 2014).<sup>54</sup> L'exception réside dans le fait que les analyses à suivre sont

---

<sup>54</sup> Les lecteurs intéressés sont renvoyés à ce document pour les détails techniques.

menées exclusivement au niveau global. Aucune analyse par communauté n'est présentée étant donné que de telles analyses seraient redondantes avec celles qui sont rapportées dans le *Rapport de référence* (Legault et al., 2014). Chaque fois que c'était possible, nous avons utilisé les données disponibles pour réévaluer les statistiques de l'EVMLO en fonction de l'échantillon réduit des quatre communautés de sorte à maximiser la validité de la comparaison avec l'échantillon projet Capacité d'apprentissage employée dans les analyses d'impact.<sup>55</sup> Lorsque cette stratégie était irréalisable pour des raisons pratiques, les comparaisons à l'échantillon EVMLO glanées à partir des six régions géographiques sont rapportées et reconnues dans le corps du texte.

#### 4.5.1. Statut d'immigrant et profil linguistique

Les procédures d'échantillonnage du projet Capacité d'apprentissage et de l'EVMLO ont été soigneusement détaillées et mises en contraste dans le *Rapport de référence* (Legault et al., 2014). La citation qui suit résume les conclusions de cette analyse :

*«Selon Forgues et Landry (2006), une population francophone (comme celle qui est utilisée dans le projet Capacité d'apprentissage) qui est définie selon le critère « ayant droit » aboutirait à un bassin bien plus restrictif, tandis qu'une population francophone (comme celle qui est utilisée dans l'EVMLO) qui est définie à l'aide de plusieurs critères (p. ex., langue maternelle, connaissance des langues officielles et langues parlées à la maison) aboutirait à un nombre supérieur de personnes admissibles.»*

Deux prédictions ont été avancées à la lumière de cette analyse: a) l'échantillon EVMLO devrait comprendre une proportion plus élevée d'immigrants que le projet Capacité d'apprentissage et b) relativement moins d'enfants devraient déclarer le français comme langue maternelle dans l'EVMLO. La première question n'a pu être abordée directement dans le *Rapport de référence* (Legault et al., 2014) à cause du fait que les parents n'ont pas été interrogés sur leur statut d'immigrants. Cette lacune a été comblée dans le huitième sondage de suivi. Les résultats de l'analyse des informations recueillies sont présentés ici. Les comparaisons fondées sur la langue maternelle des enfants et des parents sont aussi rapportées.<sup>56</sup> Le patron des résultats rapportés ici pour l'analyse par langue maternelle est équivalent à celui qui a été communiqué dans le *Rapport de référence* (Legault et al., 2014).

#### **Répondant né au Canada**

Le statut d'immigrant des répondants des échantillons projet Capacité d'apprentissage et EVMLO (quatre communautés) est rapporté dans le tableau 4.5. La première rangée rapporte la fréquence des répondants qui signalent être nés au Canada et la deuxième, celle des personnes nées hors du Canada. Comme on pouvait s'y attendre, l'inspection de la distribution des réponses

---

<sup>55</sup> La SRDC a actuellement accès ux données de fréquence groupées par communauté, ce qui permettait le calcul d'estimations appropriées pour l'échantillon global comprenant les quatre communautés. Toutefois, ce calcul n'était pas possible pour certaines variables où l'analyse par communauté a abouti à des tailles d'échantillon trop petites pour être extraites de Statistique Canada. Les six communautés dans l'échantillon EVMLO représentent néanmoins un groupe comparatif intéressant aux fins d'établir la validité externe de l'échantillon projet Capacité d'apprentissage des quatre communautés.

<sup>56</sup> La PLOP (première langue officielle parlée) n'a pas été utilisée ici pour comparer les échantillons du fait que cette donnée a été recueillie dans l'EVMLO de telle sorte que toutes les comparaisons soient invalidées (voir le Rapport de référence, 2014).



pour les deux enquêtes indique clairement que le projet Capacité d'apprentissage comprend une proportion plus grande de répondants nés au Canada. Environ 92 % des membres de l'échantillon projet Capacité d'apprentissage sont nés au Canada alors que ce fait n'était vrai que pour 76,4 % des répondants de l'échantillon EVMLO. Cette différence apparente a été confirmée par l'application d'un test statistique de khi carré qui s'est révélé significatif [ $X^2(2, N = 1\ 025) = 29,14, p < 0,001$ ].

**Tableau 4.5 : Comparaison entre le projet Capacité d'apprentissage et l'EVMLO**

Statut d'immigrant	projet Capacité d'apprentissage	EVMLO	Différences statistiquement significatives entre les deux groupes
	N (%)	N (%)	Khi carré
Né au Canada	216 (91,9)	598 (75,7)	Oui***
Né hors du Canada	19 (8,1)	192 (23,6)	

*Note* : Échantillon projet Capacité d'apprentissage fondé seulement sur les familles qui ont répondu au huitième sondage. Seuils de signification : \*\*\*  $p < 0,1\%$ ; \*\*  $p < 1\%$ ; \*  $p < 5\%$ .

### **Première langue apprise et toujours comprise – Enfants**

La langue maternelle des enfants du projet Capacité d'apprentissage est recueillie à partir du formulaire de consentement rempli par les parents. Pour l'EVMLO, la langue maternelle de l'enfant est déduite de la question suivante: (Statistique Canada, 2006, p. 35): « Quelle est la langue que [nom de l'enfant] a appris en premier à la maison et qu'il comprend encore? ».

Le tableau 4.6 montre que l'échantillon projet Capacité d'apprentissage est plus francophone en somme que celui de l'EVMLO. En effet, un pourcentage plus important d'enfants du projet Capacité d'apprentissage déclare seulement le français comme langue maternelle (première rangée du tableau). Le pourcentage d'enfants dont la langue maternelle était l'anglais seulement ou l'anglais et une autre langue était plus grand dans l'échantillon EVMLO (troisième rangée du tableau). La représentation des bilingues dans les deux échantillons était pratiquement identique (deuxième rangée du tableau).

L'application d'un test statistique de khi carré a confirmé que la distribution des enfants du projet Capacité d'apprentissage (quatre communautés) entre les différentes catégories de langue maternelle n'est pas représentative de la population minoritaire francophone dans les six régions géographiques selon les données de l'EVMLO [ $X^2(2, N = 1\ 015) = 83,16, p < 0,001$ ]. Ces résultats ne sont pas surprenants, étant donné les différences que nous avons déjà signalées au sujet des stratégies d'échantillonnage pour les deux projets de recherche.

**Tableau 4.6 : Comparaison entre le projet Capacité d'apprentissage et l'EVML0 – Enfants groupés par langue maternelle**

Langue maternelle	projet Capacité d'apprentissage	EVML0	Différences statistiquement significatives entre les deux groupes
	N (%)	N (%)	<i>Khi carré</i>
Français seulement	184 (72,4)	306 (40,2)	Oui***
Anglais et français également OU français et une autre langue	23 (9,1)	89 (11,7)	
Anglais seulement OU anglais et une autre langue OU autre(s) langue(s)	47 (18,5)	366 (48,1)	

*Note* : Seuils de signification : \*\*\*  $p < 0,1\%$ ; \*\*  $p < 1\%$ ; \*  $p < 5\%$ .

### **Première langue apprise et toujours comprise – Mères**

Le tableau 4.7 indique que la plupart des mères des échantillons du projet Capacité d'apprentissage (69,7 %) et de l'EVML0 (58,3 %) déclarent le français comme seule langue maternelle (première rangée du tableau), bien que la proportion à cet égard soit légèrement plus élevée dans l'échantillon projet Capacité d'apprentissage. De plus, une proportion moindre de mères du projet Capacité d'apprentissage déclarent la catégorie « anglais seulement OU anglais et une autre langue OU autre(s) langue(s) » (23,6 % dans la troisième rangée). L'application du test de khi carré suggère que la distribution des mères du projet Capacité d'apprentissage (quatre communautés) entre les différentes catégories de langue maternelle n'est pas représentative de la population minoritaire francophone dans les six régions géographiques fondées sur les données de l'EVML0 [ $X^2(2, N = 774) = 10,49, p < 0,01$ ].

**Tableau 4.7 : Comparaison entre les mères du projet Capacité d'apprentissage et de l'EVML0 – Mères groupées par langue maternelle**

Langue maternelle	projet Capacité d'apprentissage	EVML0	Différences statistiquement significatives entre les deux groupes
	N (%)	N (%)	<i>Khi carré</i>
Français seulement	177 (69,7)	460 (58,3)	Oui**
Anglais et français également OU français et une autre langue	17 (6,7)	75 (9,5)	
Anglais seulement OU anglais et une autre langue OU autre(s) langue(s)	60 (23,6)	254 (32,2)	

*Note* : Seuils de signification : \*\*\*  $p < 0,1\%$ ; \*\*  $p < 1\%$ ; \*  $p < 5\%$ .

### **Première langue apprise et toujours comprise – Pères**

Le tableau 4.8 montre le profil linguistique des pères du projet Capacité d'apprentissage et de l'EVML0 en fonction de leur langue maternelle. Le patron apparaît au premier coup d'œil semblable à celui qui a été observé chez les mères. La principale différence réside dans le

nombre important de pères de l'EVMLLO qui ont déclaré « anglais seulement OU anglais et une autre langue OU autre(s) langue(s) », comme langue maternelle. La proportion de pères de l'EVMLLO aux deux extrémités de la distribution dans le tableau 4.8 est virtuellement identique (46,7 % contre 47,7 %). Par contraste, les pères de l'échantillon projet Capacité d'apprentissage sont plus massivement représentés dans la catégorie « français seulement » (60,8 % dans la première rangée) que dans la catégorie « anglais seulement OU anglais et une autre langue OU autre(s) langue(s)» (32,2 % dans la troisième rangée). Ce dernier patron ressemble à celui qui a été observé dans les deux échantillons pour la langue maternelle de la mère.

L'application du test de khi carré confirme que la distribution des pères du projet Capacité d'apprentissage (quatre communautés) entre les différentes catégories de langue maternelle n'est pas représentative de la population minoritaire francophone dans les six régions géographiques fondées sur les données de l'EVMLLO [ $\chi^2(2, N = 1\ 031) = 18.16, p < 0,001$ ].

**Tableau 4.8 : Comparaison entre les pères du projet Capacité d'apprentissage et de l'EVMLLO – Pères groupés par langue maternelle**

Langue maternelle	projet Capacité d'apprentissage	EVMLLO	Différences statistiquement significatives entre les deux groupes
	N (%)	N (%)	<i>Khi carré</i>
Français seulement	149 (60,8)	367 (46,7)	Oui***
Anglais et français également OU français et une autre langue	17 (6,9)	44 (5,6)	
Anglais seulement OU anglais et une autre langue OU autre(s) langue(s)	79 (32,2)	375 (47,7)	

*Note* : Seuils de signification : \*\*\*  $p < 0,1\%$ ; \*\*  $p < 1\%$ ; \*  $p < 5\%$ .

#### 4.5.2. Caractéristiques sociodémographiques

Dans le *Rapport de référence* (Legault et al., 2014), l'échantillon projet Capacité d'apprentissage a été comparé avec l'échantillon EVMLLO en fonction du sexe de l'enfant, de la composition de la famille (taille et structure de la famille) et de la situation socio-économique (scolarité parentale, revenu familial). Le rapport conclut que les deux échantillons étaient équivalents en ce qui concerne les variables sexe de l'enfant, structure familiale, taille de la famille et revenu familial. Par contre, on a constaté que la distribution des réponses variait entre les échantillons pour la scolarité parentale (père et mère) et le nombre de frères et sœurs. Dans tous les cas, le patron général des résultats s'est reproduit dans l'analyse fondée sur les quatre communautés du projet Capacité d'apprentissage présentées ci-dessous.

##### **Revenu familial total**

Le tableau 4.9 montre que les familles du projet Capacité d'apprentissage et de l'EVMLLO sont distribuées d'une façon similaire entre les catégories de revenu considérées ici. Dans les deux cas, la catégorie modale et médiane pour les deux échantillons est de 60 000 \$ ou plus par année. L'application du test de khi carré confirme que la distribution des parents du projet Capacité d'apprentissage (quatre communautés) au sein des différentes classes de revenu est statistiquement équivalente à celle qui a été observée auprès de la population minoritaire

francophone des six régions géographiques fondées sur les données de l'EVMLO [ $X^2(5, N = 1031) = 5,42, p > 0,05$ ]. Les résultats suggèrent que la plupart des enfants dans les deux échantillons jouissent d'une bonne qualité et d'une bonne quantité de ressources matérielles pour leur développement.

**Tableau 4.9 : Comparaison entre le projet Capacité d'apprentissage et l'EVMLO – Familles par classe de revenu**

Classe de revenu	projet Capacité d'apprentissage	EVMLO	Différences statistiquement significatives entre les deux groupes
	N (%)	N (%)	<i>Khi carré</i>
10 000 \$ ou moins	16 (6,6)	54 (3,5)	Non
Entre 20 000 \$ et 29 999 \$	13 (5,3)	23 (3,5)	
Entre 30 000 \$ et 39 999 \$	15 (6,2)	64 (8,5)	
Entre 40 000 \$ et 49 999 \$	13 (5,4)	57 (4,3)	
Entre 50 000 et 59 999 \$	33 (13,6)	95 (13,7)	
60 000 \$ et plus	152 (62,8)	496 (66,3)	

*Note* : Seuils de signification : \*\*\*  $p < 0,1\%$ ; \*\*  $p < 1\%$ ; \*  $p < 5\%$ .

#### **Degré de scolarité des mères**

Le tableau 4.10 révèle trois points principaux. Premièrement, les mères du projet Capacité d'apprentissage ont en moyenne plus de scolarité que les mères de l'échantillon de l'EVMLO. En fait, près de 80 % des mères du projet Capacité d'apprentissage ont un diplôme d'études collégiales (DEC) ou un diplôme universitaire, tandis qu'environ 70 % seulement des mères de l'EVMLO ont un niveau de scolarité équivalent. Cette différence était principalement attribuable au petit nombre de mères ayant au moins un diplôme ou un certificat d'études collégiales dans l'EVMLO par comparaison à celles du projet Capacité d'apprentissage (deuxième rangée). Deuxièmement, il y a autant de mères avec un diplôme d'études collégiales (39,0 %) qu'il y a de mères avec un diplôme universitaire (39,0 %) dans le projet Capacité d'apprentissage. Troisièmement, il y a plus de mères dans l'EVMLO qui ont fréquenté l'université (42,7 %) que dans le projet Capacité d'apprentissage (39,0 %), mais la différence est négligeable. L'application du test de khi carré a confirmé que la distribution des mères du projet Capacité d'apprentissage entre les différents degrés de scolarité n'est pas représentative de la population minoritaire francophone dans les quatre régions géographiques fondées sur les données de l'EVMLO [ $X^2(2, N = 795) = 16,8, p < 0,01$ ].

**Tableau 4.10 : Comparaison des degrés de scolarité des mères du projet Capacité d'apprentissage et de l'EVMLO**

Degré de scolarité	projet Capacité d'apprentissage	EVMLO	Différences statistiquement significatives entre les deux groupes
	N (%)	N (%)	
Diplôme d'études secondaires ou moins OU quelques cours postsecondaires	56 (22,0)	172 (31,8)	Oui***
Diplôme ou certificat collégial (p. ex. école de métier)	99 (39,0)	138 (25,5)	
Diplôme universitaire (baccalauréat, maîtrise ou doctorat)	99 (39,0)	231 (42,7)	

*Note* : Seuils de signification : \*\*\*  $p < 0,1\%$ ; \*\*  $p < 1\%$ ; \*  $p < 5\%$ .

### **Degré de scolarité des pères**

Le tableau 4.11 indique qu'une comparaison entre la scolarité des pères du projet Capacité d'apprentissage et de l'EVMLO ressemble à la comparaison entre les mères. Comme c'était le cas pour les mères (voir tableau 4.10), il y a plus de pères qui ont fréquenté l'université dans l'EVMLO (35,9 %) que dans le projet Capacité d'apprentissage (31,2 %). Toutefois, les pères du projet Capacité d'apprentissage sont généralement plus scolarisés que les pères de l'EVMLO. En effet, près des deux tiers d'entre eux ont un diplôme d'études collégiales (DEC) ou un diplôme universitaire, tandis que 60 % des pères de l'EVMLO ont un degré de scolarité équivalent. Enfin, l'application du test khi carré suggère que la distribution des pères du projet Capacité d'apprentissage entre les différents degrés de scolarité n'est pas représentative de la population minoritaire francophone dans les quatre régions géographiques fondées sur les données de l'EVMLO [ $X^2(2, N = 788) = 13,5, p < 0,01$ ].

**Tableau 4.11 : Comparaison entre les degrés de scolarité des pères du projet Capacité d'apprentissage et de l'EVMLO**

Degré de scolarité	projet Capacité d'apprentissage	EVMLO	Différences statistiquement significatives entre les deux groupes
	N (%)	N (%)	
Diplôme d'études secondaires ou moins OU quelques cours postsecondaires	79 (32,0)	216 (39,9)	Oui**
Diplôme ou certificat d'études collégiales (p. ex., école de métier)	91 (36,8)	131 (24,2)	
Diplôme universitaire (baccalauréat, maîtrise ou doctorat)	77 (31,2)	194 (35,9)	

*Note* : Seuils de signification : \*\*\*  $p < 0,1\%$ ; \*\*  $p < 1\%$ ; \*  $p < 5\%$ .

### Taille de la famille

Selon le tableau 4.12, la taille modale et médiane de la famille est de quatre personnes pour les deux échantillons (quatre communautés). Dans les deux enquêtes, ces familles représentent environ la moitié de l'échantillon. Les familles restantes sont distribuées assez également entre les familles de trois personnes ou moins et les familles de cinq personnes ou plus. L'application du test de khi carré suggère qu'il n'y a pas de différence significative entre le projet Capacité d'apprentissage et l'EVMLO [ $X^2(2, N = 792) = 0,31, p > 0,05$ ] dans la distribution de la taille des familles.

**Tableau 4.12 : Comparaison de la taille des familles<sup>1</sup> dans le projet Capacité d'apprentissage et l'EVMLO**

Nombre de personnes	projet Capacité d'apprentissage	EVMLO	Différences statistiquement significatives entre les deux groupes
	N (%)	N (%)	Khi carré
Trois personnes ou moins	68 (26,8)	143 (26,6)	Non
Quatre personnes	136 (53,5)	280 (52,0)	
Cinq personnes ou plus	50 (19,7)	115 (21,4)	

*Note* : Seuils de signification : \*\*\* < 0,1%; \*\* < 1%; \* < 5%.

<sup>1</sup> Le nombre de personnes dans une famille comprend seulement le nombre de parents et d'enfants.

### Fratrie

Le tableau 4.13 indique que le nombre modal et médian d'enfants par répondant (famille) est de deux pour le projet Capacité d'apprentissage et l'EVMLO. Toutefois, il y a un peu plus de familles avec exactement deux enfants dans le projet Capacité d'apprentissage (57,9 %) que dans l'EVMLO (50,4 %). De plus, il y a plus de familles avec trois enfants ou plus dans l'EVMLO (30,1 %) que dans le projet Capacité d'apprentissage (20,1 %). En revanche, le nombre de familles avec un enfant seulement, environ 20 %, est à peu près le même dans les deux enquêtes. L'application du test de khi carré montre que la distribution du nombre d'enfants par répondant est significativement différente dans les deux échantillons [ $X^2(2, N = 1\ 040) = 9,52, p < 0,01$ ].

**Tableau 4.13 : Comparaison entre le projet Capacité d'apprentissage et l'EVMLO – Nombre d'enfants par répondant**

Nombre d'enfants	projet Capacité d'apprentissage	EVMLO	Différences statistiquement significative entre les deux groupes
	N (%)	N (%)	Khi carré
Un enfant	56 (22,0)	154 (19,5)	Oui**
Deux enfants	147 (57,9)	396 (50,4)	
Trois enfants ou plus	51 (20,1)	236 (30,1)	

*Note* : Seuils de signification : \*\*\*  $p < 0,1\%$ ; \*\*  $p < 1\%$ ; \*  $p < 5\%$ .

### Structure familiale

Nous avons eu à redéfinir les familles du projet Capacité d'apprentissage de sorte qu'elles soient catégorisées comme monoparentales ou biparentales afin de comparer la structure

familiale des échantillons projet Capacité d'apprentissage et EVMLO (voir tableau 4.14). Cette dernière catégorie comprenait autant les familles intactes que les familles reconstituées où les deux parents (ou un parent et son conjoint) vivent avec l'enfant. La catégorie monoparentale comprend les familles où seul un parent vit dans la maison avec l'enfant.

Il convient de mentionner que le père ou la mère de l'enfant peut être soit le parent biologique, soit le parent adoptif. Nous aimerions aussi mentionner que les couples de même sexe ont été exclus de l'analyse, ainsi d'ailleurs que les enfants élevés par quelqu'un d'autre que la mère ou le père biologique ou adoptif de l'enfant. Comme le montre l'application du test de khi carré, la distribution des enfants dans les foyers monoparentaux ou biparentaux est représentative de la population minoritaire francophone dans les six régions géographiques fondées sur les données de l'EVMLO ( $X^2(1, N = 1\ 043) = 0,79, p > 0,05$ ).

**Tableau 4.14 : Comparaison entre le projet Capacité d'apprentissage et l'EVMLO – Nombre de familles monoparentales et biparentales**

Structure familiale	projet Capacité d'apprentissage	EVMLO	Différences statistiquement significatives entre les deux groupes
	N (%)	N (%)	<i>Khi carré</i>
Monoparentale	23 (9,1)	87 (11,0)	Non
Biparentale	231 (90,9)	702 (89,0)	

*Note* : Seuils de signification : \*\*\*  $p < 0,1\%$ ; \*\*  $p < 1\%$ ; \*  $p < 5\%$ .

## 4.6. SOMMAIRE

Dans ce chapitre, nous avons abordé les diverses questions méthodologiques touchant la validité interne et externe des analyses décrites aux chapitres 5 et 6. Les questions techniques liées aux procédures de contrôle de la qualité, aux valeurs manquantes, à l'estimation des erreurs-types, aux variables confusionnelles et à la validité externe ont été examinées. Nous résumons à tour de rôle les points majeurs de chaque section.

Nous avons dressé la liste exhaustive des procédures de vérification et de contrôle des données à la section 4.1. Elles ont été conçues pour minimiser les problèmes (p. ex. les erreurs de mesure) posés par le processus de collecte de données et pour identifier et corriger les problèmes qui ont atteint les bases de données électroniques. Ces procédures et le respect général de l'approche des méthodes mixtes (ou « opérations convergentes ») en matière de recherche assurent la validité des résultats découlant des analyses d'impact.

L'analyse des valeurs manquantes rapportée à la section 4.2 montre que, dans l'ensemble, les valeurs manquantes constituent un problème d'une ampleur très limitée et elles sont réparties également à travers les groupes expérimentaux. Le taux d'attrition était très bas durant les deux premières années du projet, alors que le taux de réponses était très élevé. Les analyses indiquaient que les familles ayant obtenu les scores moins élevés aux indicateurs francophones étaient plus susceptibles de se retirer du projet. Les tests statistiques révélaient que les autres valeurs manquantes associées aux données de sondage étaient produites aléatoirement. Les données manquantes non aléatoires ont été problématiques avec toutes les échelles de l'ÉPE-AD, sauf pour le domaine Langage et communication. Ces données ont été causées par le fait

que certains enfants ont complété l'ÉPE-AD en anglais plutôt qu'en français. La stratégie d'imputation adoptée ici a permis d'imputer certaines données manquantes comme MAR (c.-à-d. manquante non aléatoire qui peut être imputée de façon fiable selon d'autres variables dans la base de données), mais pas toutes (MNAR, c.-à-d. manquante non aléatoire qui ne peut pas être imputée). En conséquence, les analyses d'impact reposant sur ces échelles ne peuvent être généralisées qu'à l'égard d'une population de francophones minoritaires dont les enfants sont susceptibles de rencontrer les critères nécessaires pour compléter l'ÉPE-AD en français (c.-à-d. les enfants dont le score est relativement élevé sur l'échelle Langage et communication en français, domaine C, et l'échelle Conscience et engagement envers la culture francophone, domaine E) ou, plus exactement, à une population d'enfants qui ont complété l'ÉPE-AD en français deux fois ou plus au cours des cinq premières évaluations.

La stratégie d'analyse adoptée ici (c.-à-d. l'analyse d'un panel) a permis une certaine flexibilité pour déterminer comment les erreurs-types devraient être estimées (section 4.3). La logique appuyant les décisions prises dans le cas présent a fait l'objet d'une description sommaire. La stratégie générale consistait à adopter une approche conservatrice qui nous empêcherait de surestimer l'exactitude de nos résultats. Par conséquent, les tests de l'impact du programme sont relativement inefficaces, mais cette inefficacité est contrebalancée par le caractère robuste qu'ils imposent aux violations du postulat d'hétérogénéité de la variance/dispersion. En pratique, la mise en œuvre de cette stratégie a entraîné le groupement des erreurs-types au niveau des garderies.

Une liste exhaustive des covariables apparaît à la section 4.4, et son but était de: a) faire des ajustements statistiques compensant pour les changements survenant dans la composition des groupes au fil du temps; et b) faire des ajustements statistiques compensant pour les différences dans la trajectoire développementale qui se seraient manifestées entre les groupes en l'absence d'un traitement. L'insertion de ces covariables dans les analyses a permis d'améliorer la validité statistique des résultats.

Enfin, nous avons examiné la question de la validité externe de l'étude. Des travaux antérieurs ont montré que l'échantillon total du projet Capacité d'apprentissage était plus francophone que prévu, si on le comparait à l'échantillon de l'EVMLO. Ce résultat a été répété ici pour l'échantillon des quatre communautés du projet Capacité d'apprentissage. Dans le *Rapport de référence* (Legault et al., 2014), on avait avancé l'hypothèse que ce biais apparent est probablement inhérent à la population d'enfants qui sont en garderies francophones. Dans la mesure où cet argument est fondé, tous les résultats obtenus pour cet échantillon s'appliqueraient de façon générale à la population d'enfants francophones inscrits en garderie. Cette limite est naturelle, puisque cette population d'enfants est celle qui serait touchée par une intervention en garderie.

Si, contrairement à l'argument présenté ci-dessus, les différences de langue maternelle qui ont été observées entre le projet Capacité d'apprentissage et l'EVMLO sont réelles, ce qui revient à dire que l'échantillon projet Capacité d'apprentissage est en fait non représentatif de la population qu'il cible, alors les résultats des analyses d'impact figurant aux chapitres 5 et 6 du présent rapport peuvent en réalité sous-estimer l'importance de l'effet réel du programme pour une population qui est plus diverse du point de vue linguistique. On a examiné la possibilité que les variables linguistiques accentuent ou atténuent l'effet du programme testé en réévaluant les effets du programme pour les enfants qui utilisent davantage la langue française et encore pour



ceux qui l'utilisent moins (voir les sections 5.2.3 et 6.2.3). L'information qui se dégage de ces analyses peut aider à déterminer si l'intervention serait plus efficace si elle ciblait des sous-populations données, comme celle des enfants qui ont la plus grande exposition aux langues autres que le français dans leur environnement domiciliaire.

Le fait que les échantillons projet Capacité d'apprentissage et EVMLO soient généralement similaires lorsqu'on examine les caractéristiques non linguistiques soutient l'argument selon lequel l'échantillon projet Capacité d'apprentissage est représentatif des francophones dans un contexte minoritaire. La seule différence qui mérite d'être mentionnée tient au fait que les parents de l'échantillon projet Capacité d'apprentissage sont légèrement plus susceptibles que ceux de l'autre échantillon d'atteindre un niveau de scolarité supérieur au niveau secondaire.

En somme, nous avons pris de grandes précautions afin de nous assurer que les effets estimés du programme présenté aux chapitres 5 et 6 sont aussi valides que possible, qu'ils soient internes ou externes. D'autres vérifications font l'objet de discussions à mesure qu'elles deviennent pertinentes à l'interprétation des résultats dans les sections suivantes du présent rapport.

## 5. Analyse d'impact : Première année

Les deux composantes centrales de l'intervention étaient le nouveau programme en garderie et les ateliers familles. Le but principal de la première composante était d'influencer directement les résultantes pour les enfants, tandis que celui de la seconde composante était d'influencer directement les attitudes et les comportements des parents. Par conséquent, le succès de cette intervention a été évalué en analysant les résultats provenant des mesures de résultantes pour les enfants et leurs parents. Les résultats de la première année du projet figurent dans ce chapitre et ceux de la deuxième année figurent au chapitre 6. Nous avons divisé les analyses de cette manière puisqu'il n'y avait pas suffisamment de degrés de liberté pour tenir compte de toutes les évaluations au sein du même modèle de régression. De plus, l'entrée de la plupart des enfants en prématernelle au début de la deuxième année du programme constituait un point de rupture conceptuel naturel.<sup>57</sup>

Dans les analyses de la première année, les effets du programme de garderie testé sur les résultantes développementales des enfants ont été évalués: (a) en comparant les trois groupes expérimentaux, (b) en évaluant si les effets du temps passé en garderie (en heures) étaient touchés par le programme testé, (c) en testant l'effet direct de la qualité du programme et de la fidélité au programme (pour les enfants fréquentant une garderie), (d) en testant si la qualité du programme et la fidélité au programme expliquent adéquatement les différences observées entre les groupes expérimentaux et e) en réévaluant les différences entre les groupes expérimentaux, d'une part, pour les enfants dont l'exposition au français est élevée et, d'autre part, pour ceux dont elle est faible. L'objet de ces dernières analyses était de déterminer les sous-populations pour lesquelles le programme de garderie testé est particulièrement efficace.

Les répercussions des ateliers familles sur les attitudes et les comportements des parents ont été testées en comparant les familles du groupe programme en garderie en fonction de leur fréquentation des ateliers ainsi que de la qualité et de la fidélité de la formation reçue. Toutes les familles de l'échantillon ont été prises en considération dans les analyses de la fréquence et de la langue des activités de littératie là où il était possible de comparer les trois groupes expérimentaux.

Ces deux questions de recherche distinctes, portant sur les enfants, d'une part, et sur les parents, d'autre part, sont abordées dans les sections 5.1 et 5.2 respectivement. Un sommaire des résultats est présenté à la section 5.3.

---

<sup>57</sup> Comme on le décrit dans les chapitres 3 et 4, nous avons choisi de grouper les erreurs-types au niveau des garderies, décision qui donne environ 20 degrés de liberté pour les tests de signification. L'inclusion de toutes les mesures répétées dans la même analyse porterait les degrés de liberté sous la barre des dix degrés pour certains tests de signification, ce qui veut dire qu'un test équitable du programme ne serait pas possible.

## 5.1 IMPACTS SUR LES ENFANTS

### 5.1.1 Variables retenues pour les analyses

Parmi les nombreuses variables mesurées au cours de l'étude, les suivantes ont été retenues pour les analyses d'impacts de la première année. Ces variables peuvent être divisées en deux grandes catégories, soit les mesures de résultantes et les prédicteurs. Nous décrivons maintenant de façon succincte les deux types de variables.

#### *Mesures de résultantes*

Ce rapport d'impact concerne la préparation à la scolarisation. Pour cette raison, nous avons sélectionné les résultantes provenant des échelles de l'ÉPE-AD (version française seulement) suivantes : Langage et communication, Conscience de soi et de l'environnement, Habiletés cognitives et Physique/moteur. Les données complètes n'étaient disponibles que pour l'échelle Langage et communication.<sup>58</sup> Nous avons créé des sous-échelles de vocabulaire plus précises en choisissant et en combinant les items pertinents dans les quatre domaines de résultantes principaux. En élaborant ces échelles, nous avons fait la distinction bien documentée entre le Vocabulaire réceptif et le Vocabulaire expressif. En tout, six résultantes de l'ÉPE-AD ont fait l'objet des analyses qui suivent.

#### **Standardisation des résultantes**

Avant de procéder aux analyses, nous avons remis les résultantes à l'échelle afin qu'elles soient standardisées pour chaque période de temps, faisant en sorte que chaque variable de résultante avait une moyenne de 0 et un écart-type de 1 pour l'échantillon total de participants. Une conséquence pratique de cette transformation est que les scores pour chaque période de temps peuvent être interprétés en termes d'échelle commune. La décision de standardiser les scores s'appuyait sur le fait que la nature des scores d'échelle bruts a changé sur le plan qualitatif durant la deuxième année de l'étude. Plus précisément, la composition des échelles a été altérée (c.-à-d. que des items ont été éliminés) avec des modifications correspondantes pour le nombre total d'items. La forte corrélation entre les échelles originales et les échelles révisées appuie notre thèse voulant que très peu d'informations aient été perdues suite à ces modifications. Néanmoins, les différences d'échelle (p. ex. score maximal) de nos mesures au fil du temps posaient un obstacle technique aux analyses statistiques : la signification des valeurs absolues des scores varie avec le temps, ce qui invalide toute comparaison effectuée d'une période de temps à l'autre. Nous avons relevé ce défi technique mineur en recourant à la procédure de standardisation discutée ci-dessus, qui assure que les scores des résultantes ont la même signification au fil du temps.

---

<sup>58</sup> Pour les trois autres échelles, les données n'étaient disponibles que pour les participants qui avaient complété l'ÉPE-AD en français au moins deux fois au cours des cinq premières évaluations. Pour des raisons évidentes, cela signifie que les analyses menées avec les scores des échelles Conscience de soi et de l'environnement, Habiletés cognitives, Physique/moteur, Vocabulaire réceptif et Vocabulaire expressif sont des analyses de sous-échantillon efficaces fondées sur un sous-groupe excluant les enfants dont les compétences langagières en français sont les plus faibles. Cette exclusion n'était pas fortement liée au groupe expérimental, ce qui signifie que la validité interne des estimations sur les effets programme n'est pas compromise. En quelque sorte, cette question affecte la validité externe en limitant l'application générale des résultats aux enfants qui répondaient aux critères établis pour la prise du test en français.

## Interprétation des scores standardisés

Les scores standardisés employés dans les analyses d'impact sont interprétés comme suit. Le score standardisé de chaque participant représente la différence entre le score du participant et le score moyen de l'échantillon. De façon éclairée, cette différence est exprimée en unités d'écart-type. Par exemple, un score de 1,11 signifie que les participants ont obtenu un score qui était à 1,11 écart-type au-dessus du score moyen obtenu par les participants de l'échantillon du projet Capacité d'apprentissage pour l'évaluation correspondante. Les estimations DD des effets programme sont interprétées de manière semblable comme des différences entre les groupes en unités d'écart-type. Cohen (1988) fournit des références conventionnelles pour interpréter l'ampleur des effets dans une échelle standardisée. Une différence standardisée entre les groupes de 0,20 est considérée comme légère, une différence de 0,50 est considérée comme moyenne et une différence de 0,80 est considérée comme importante. Ces références doivent seulement servir de guide général pour juger l'importance d'un effet.

## Évaluation des conséquences de la standardisation

Pour être en mesure de procéder à la standardisation, nous devons sacrifier notre capacité de: a) comparer les scores bruts de notre échantillon à ceux d'une population normative; et b) évaluer la trajectoire développementale de notre échantillon au fil du temps.<sup>59</sup> Nous faisons valoir que les conséquences de la standardisation sont négligeables parce que: a) aucun échantillon normatif n'existe pour la version de l'ÉPE-AD utilisée dans le cadre du projet Capacité d'apprentissage; et que b) les enfants de l'étude ont complété le test plusieurs fois, ce qui a sans doute induit des effets d'entraînement massifs (c.-à-d. de meilleurs scores suite à la répétition d'une tâche). Ces effets d'entraînement sont nécessairement confondus avec toute tentative d'estimer le développement de notre échantillon au fil du temps. Autrement dit, la signification intrinsèque de la valeur absolue des scores bruts observés est limitée, alors que les changements de scores au fil du temps (c.-à-d. la trajectoire développementale de notre échantillon) n'ont pu être différenciés dans un cas ou l'autre des effets d'entraînement. Ce qui importe est que nous puissions comparer de façon significative les *trajectoires de développement* pour nos trois groupes expérimentaux d'une évaluation à l'autre. Cette estimation de la « différence des différences » reflète adéquatement ces comparaisons et s'applique aussi bien aux scores standardisés qu'aux scores bruts.<sup>60</sup>

### **Prédicteurs importants et covariables**

Plusieurs variables ont été incluses dans les équations de régression servant à modéliser les résultantes pour les enfants en matière de préparation à la scolarisation. La plus grande distinction que l'on puisse faire ici est celle qui existe entre les prédicteurs importants des résultantes pour les enfants et les covariables. Les prédicteurs importants font l'objet d'un intérêt particulier en matière de recherche, alors que les covariables ne servent simplement qu'à

---

<sup>59</sup> La standardisation des scores donne une moyenne de 0. La standardisation à l'intérieur de chaque période de temps donne une moyenne de 0 à chaque période de temps. Si le score moyen est de 0 à chaque période de temps, alors la moyenne est constante à travers le temps et l'effet du « Temps » dans un modèle de régression sera nécessairement nul. Lorsque l'ensemble des gains développementaux pour l'échantillon de participants n'est pas d'intérêt direct, cela représente une perte d'information négligeable.

<sup>60</sup> Les analyses préliminaires indiquaient que toutes les mesures de résultantes pour les enfants étaient sensibles au changement au fil du temps. De plus, aucune des mesures n'avait plafonné à la 5e évaluation.

améliorer la validité interne des tests plus importants. Pour des raisons d'économie et de limpidité de la présentation, notre rapport met l'accent sur les prédicteurs importants présents dans nos modèles.

Les prédicteurs importants comprennent le temps (période d'évaluation), le groupe expérimental auquel appartient l'enfant, la moyenne d'heures qu'il passe en garderie chaque semaine, ainsi que la qualité de ces services de garde. Cette liste sera allongée à la section 5.1.3, dans laquelle nous traitons des modérateurs et des médiateurs de l'effet expérimental, afin d'y inclure le profil linguistique des participants (c.-à-d. type de foyer, langue des activités de littératie au niveau de base et les langues habituellement parlées par l'enfant au foyer au niveau de base). Ces variables sont décrites dans la méthodologie (chapitre 2) et la liste détaillée des covariables à inclure dans les analyses apparaît à la section 4.4. En tout, nous avons inséré huit variables dans l'analyse pour contrôler formellement les différences entre les groupes, plus 11 autres variables en raison de leur relation statistiquement significative avec les variables de résultantes.

### **5.1.2 Résultats de la comparaison des groupes**

Cette section présente les résultats des analyses d'impact de la première année. Les estimations de l'effet traitement reposent sur la trajectoire développementale relative des trois groupes à l'étude au cours des quatre premières évaluations. La première évaluation est considérée comme une évaluation pré-test et les trois évaluations subséquentes, comme des évaluations post-test. Tel que mentionné auparavant, l'effet traitement est mesuré par l'estimateur DD, qui capte les différences entre les trajectoires développementales des groupes. Au total, six mesures de résultantes ont fait l'objet d'analyses basées sur les données collectées de la version française de l'ÉPE-AD, à savoir les échelles de Langage et communication, Conscience de soi et de l'environnement, Habiletés cognitives et Physique/moteur, ainsi que les sous-échelles improvisées de Vocabulaire réceptif et de Vocabulaire expressif développées par la SRSA. Nous prévoyons que le programme aura des impacts positifs sur toutes les mesures, sauf sur l'échelle d'habileté physique, qui évalue une habileté qui n'est pas directement ciblée par le programme. Nous décrivons maintenant comment ces mesures de résultantes ont été traitées dans les analyses d'impact.

#### ***Survol des analyses et des détails de spécification***

Le but de cette section est d'étayer la logique derrière les analyses d'impact. Nous expliquons d'abord le raisonnement sous-jacent à la structure adoptée dans la présentation des résultats. Ensuite, nous décrivons en détail les spécifications de notre modèle de régression.

#### **Structure de cette section**

Les résultats de cette série d'analyses sont présentés en deux parties. La première partie rend compte en détail des analyses des scores à l'échelle Langage et communication (réalisée en français) au fil du temps pour l'échantillon total de participants. La deuxième partie présente une analyse davantage synthétisée de la série entière de six échelles de l'ÉPE-AD, incluant les scores de Langage et communication. Les résultats présentés dans cette deuxième partie reposent sur le sous-échantillon de participants qui ont complété l'ÉPE-AD en français au moins deux fois au cours des cinq premières évaluations. Rappelons que les enfants ne poursuivaient leur évaluation

en français que s'ils avaient obtenu un score suffisant à la fois à l'échelle courte servant à évaluer l'environnement familial linguistique de l'enfant ainsi qu'à l'échelle Langage et communication de l'ÉPE-AD (voir l'annexe B pour les détails rattachés à la règle de décision). Par conséquent, il y a des données manquantes sur les échelles Conscience de soi et de l'environnement, Habiletés cognitives et Physique/moteur puisqu'il était impossible de combiner les scores de l'ÉPE-AD en français et en anglais.

On a accordé un statut spécial aux analyses basées sur l'échantillon total de participants, car il représente un très bon test des effets du programme. La taille de l'échantillon pour ces premières analyses est relativement grande et non biaisée. La seconde série d'analyses est basée sur un échantillon qui est à la fois restreint et biaisé en faveur des enfants qui possèdent des compétences langagières relativement fortes en français. Ce biais a été inévitable à cause de la nature des données manquantes qui ont causé l'exclusion de cas.<sup>61</sup> Nous avons décidé d'effectuer un deuxième ensemble d'analyses des scores de Langage et communication avec ce sous-échantillon de sorte qu'un échantillon commun de participants soit utilisé pour générer des estimations pour toutes les échelles. La comparaison de ce deuxième ensemble d'analyses des scores de communication par rapport au premier ensemble d'analyses donne une idée générale de l'ampleur du biais engendré par l'exclusion des enfants ayant complété l'ÉPE-AD en anglais. En fait, une comparaison des deux ensembles de résultats indique que l'effet du biais en question sur les estimations de l'effet du programme est négligeable.

### Détails de spécification du modèle

Un grand nombre de covariables sont incluses dans l'estimation des impacts du programme. Plutôt que d'entrer toutes les covariables en même temps, nous avons opté pour une stratégie servant à intégrer progressivement des groupes de covariables dans le modèle de régression. Un des avantages de cette approche est qu'elle permet de suivre les changements dans l'ampleur des effets à chaque nouvelle spécification. Nous pourrions, par exemple, isoler le groupe des variables annulant l'effet traitement. Un autre avantage de cette stratégie est qu'elle permet de diagnostiquer à chacune des étapes l'effet de certains problèmes techniques liés aux estimations de régression tels la suppression (c.-à-d. de grands changements dans l'ampleur ou la direction d'un effet) et la multicollinéarité (c.-à-d. l'inflation des erreurs-types). Les spécifications employées à chaque étape sont discutées ci-dessous. Chaque modèle de spécification est identifié par une valeur numérique placée entre parenthèses (voir ci-dessous).

Le modèle initial (1) dans l'évaluation des impacts du programme comprenait des variables fictives représentant les périodes de temps, des variables fictives représentant l'appartenance au groupe expérimental et un terme représentant les interactions entre ces indicateurs de variables fictives (c.-à-d. les estimateurs DD). Conformément à la logique de l'estimateur DD, le niveau de base ou le temps de pré-test a été placé en « référence ». Autrement dit, les scores de toutes les périodes de temps subséquentes ont été comparés à ce niveau de base. Pour les variables fictives de l'effet traitement, nous avons choisi de placer le groupe programme en garderie en référence, ce qui a facilité l'interprétation des estimations DD générées par le modèle en permettant de comparer les trajectoires développementales des deux groupes témoins avec celles du groupe

---

<sup>61</sup> Une stratégie d'imputation solide ferait appel à un échantillon d'enfants qui auraient dû effectuer le test en anglais, mais à qui on l'aurait néanmoins administré en français. Dans ce cas, il serait possible d'obtenir des estimations précises des scores que les enfants auraient obtenus dans la version française.

programme. En d'autres mots, nous avons obtenu des estimations des impacts de notre programme relativement au groupe témoin en garderie et au groupe témoin hors garderie par l'entremise d'une seule analyse de régression. *Il est important de noter que la décision de placer le groupe programme en référence signifie que les valeurs négatives des estimations DD pour toutes les comparaisons entre les groupes représentent un effet traitement positif (c.-à-d. un avantage pour le groupe traité).* Pour les fins de ce modèle de base, nous avons exercé des contrôles statistiques de l'effet de la communauté en incluant une série de variables fictives.

À la deuxième étape (2), des covariables de nature linguistique ont été ajoutées au modèle. Ces variables incluent: a) le continuum des langues habituellement parlées par l'enfant au niveau de base; b) la langue des activités de littératie au niveau de base; c) la composition linguistique du foyer (p. ex. endogame francophone, exogame francophone); d) la langue de garde de jour lorsque l'enfant avait de 0 à 12 mois; et d) l'interaction entre la variable de la langue de garde de jour et le temps. Ces variables ont été incluses pour tenir compte des différences linguistiques statistiquement significatives observées principalement entre le groupe programme et le groupe témoin hors garderie. L'interaction entre le temps et la langue de garde de jour a été incluse parce que ce facteur est fortement lié à la plupart des mesures dépendantes. Les autres covariables, elles, ne l'étaient pas. Son inclusion dans le modèle améliore la validité des estimations DD en contrebalançant les différences dans la pente qui pouvaient exister lors de l'évaluation de base. Cette question a été discutée plus à fond à la section 3.3.3.

À la troisième étape (3), des covariables de nature démographique et socioéconomique ont été insérées dans le modèle. Le sexe de l'enfant a été ajouté au modèle à cette étape, ainsi que son âge en mois au moment de l'administration de l'ÉPE-AD. Les variables de contrôle socioéconomiques incluaient le niveau d'éducation atteint par la mère, l'âge de la mère à la naissance de l'enfant et la classe du revenu familial.

À la quatrième étape (4), des covariables reliées à la composition de la famille ont été incluses dans le modèle. Ces covariables sont dynamiques puisque les valeurs prises par ces variables peuvent changer avec le temps. Par contre, les autres covariables ne changent pas avec le temps : on dit alors qu'elles sont « fixes ». Cette liste de covariables comprend les variables suivantes: taille du ménage (nombre de personnes habitant dans la maison); nombre d'enfants plus jeunes demeurant dans la maison; et nombre d'enfants plus âgés habitant dans la maison. La variable fictive représentant le fait que l'enfant vit ou non dans une famille monoparentale a été exclue de la dernière série de résultats rapportés en raison de sa grande asymétrie (dépassant un ratio de 90:10; Tabachnick & Fidell, 2007, p. 73) et de sa corrélation élevée avec les autres variables de composition de la famille.

À la cinquième (5) et dernière étape, nous avons intégré dans le modèle des échelles parentales en tant que covariables. Certaines de ces échelles ont capté des dimensions relatives au style de parentage, tel que l'échelle de Style parental autoritaire et l'échelle de Style parental positif. Une échelle a été ajoutée au modèle pour estimer le Fonctionnement familial.

La stratégie consistant à inclure graduellement une série de covariables a été appliquée à toutes les analyses. Toutefois, les résultats obtenus à chaque inclusion sont seulement rapportés ici pour les analyses de l'échelle Langage et communication effectuées sur l'échantillon total de participants. Les résultats obtenus à l'une de ces étapes progressives ne sont discutés dans le corps du texte que lorsqu'ils affectent l'interprétation de fond des résultats.

## **RÉSULTATS D'IMPACT : Score standardisé de Langage et communication (en français) - Échantillon total**

L'impact du programme testé sur les compétences de communication en français est évalué dans la présente section. Seules les estimations des effets programme en comparaison avec le groupe témoin en garderie et le groupe témoin hors garderie sont présentées sous forme d'estimations DD. Ces estimations devraient être interprétées comme étant des estimations des effets programme en considérant les covariables incluses dans le modèle (les effets des covariables ne sont pas présentés). Un sommaire des résultats de ces analyses se trouve au tableau 5.1.

Plusieurs points sont dignes d'attention quant aux résultats présentés au tableau 5.1.<sup>62</sup> Tout d'abord, nous remarquons que les différences entre le groupe programme en garderie et les groupes témoins pour la période de base ne sont pas statistiquement significatives pour cette mesure de résultantes au seuil de signification alpha de 0,05.<sup>63</sup> Il y a très peu de données probantes montrant qu'il existait des différences réelles entre les groupes avant l'intervention. Nous pouvons donc conclure que la validité interne de l'effet de notre traitement n'est pas menacée par l'existence de différences initiales entre les groupes.

En ce qui a trait à l'effet traitement, la première chose à noter est que le test global de l'interaction Groupe expérimental par Temps n'est pas statistiquement significatif à aucune étape des analyses (voir la rangée intitulée « Test F composite de Wald »).<sup>64</sup> Ce test F évalue l'effet

---

<sup>62</sup> Le nombre de participants visés par les analyses diminue légèrement à mesure que l'on insère des covariables dans le modèle. Cette diminution de la taille de l'échantillon s'explique par la présence de valeurs manquantes réparties à travers de ces variables. En principe, cet écart relativement à la taille de l'échantillon ne pose pas de problème pour ce qui est de l'interprétation des résultats puisque le nombre de valeurs manquantes est peu élevé et que le patron n'est pas systématique.<sup>63</sup> Les résultats de ce test semblent contredire ceux qui ont été rapportés dans le *Rapport de référence* (Legault et al., 2014), où une différence significative a en fait été observée, mais la contradiction est seulement apparente. Nous rappelons au lecteur que l'échantillon des enfants associés à la présente analyse est plus petit après l'exclusion de deux communautés (Edmonton et Saint-Jean) et des dix cas d'attrition de la première année. Cette modification de la composition de l'échantillon est responsable de la différence entre les rapports pour ce qui est de la similarité des groupes expérimentaux pour la période de base.<sup>64</sup> Une technique communément utilisée pour réduire le nombre de tests inférentiels effectués est de ne pas tenir compte des contrastes simples lorsque le test de signification global est nul. Il s'agit de réduire le nombre de tests statistiques et, en conséquence, la probabilité de commettre des erreurs de Type-1. Si nous avions appliqué cette approche ici, nous aurions été contraints de conclure que le programme n'avait aucun impact sur le score de Communication. Toutefois, nous n'avons pas appliqué cette approche parce que a) l'estimation DD associé au premier post-test revêtait un intérêt particulier (c.-à-d. un contraste a priori) et b) une présentation plus détaillée des résultats permet d'extraire un patron cohérent de résultats qui peut ensuite être intégré au cadre plus large de résultats présentés dans ce rapport. En somme, nous avons décidé de laisser parler les données.

<sup>63</sup> Les résultats de ce test semblent contredire ceux qui ont été rapportés dans le *Rapport de référence* (Legault et al., 2014), où une différence significative a en fait été observée, mais la contradiction est seulement apparente. Nous rappelons au lecteur que l'échantillon des enfants associés à la présente analyse est plus petit après l'exclusion de deux communautés (Edmonton et Saint-Jean) et des dix cas d'attrition de la première année. Cette modification de la composition de l'échantillon est responsable de la différence entre les rapports pour ce qui est de la similarité des groupes expérimentaux pour la période de base.<sup>64</sup> Une technique communément utilisée pour réduire le nombre de tests inférentiels effectués est de ne pas tenir compte des contrastes simples lorsque le test de signification global est nul. Il s'agit de réduire le nombre de tests statistiques et, en conséquence, la probabilité de commettre des erreurs de Type-1. Si nous avions appliqué cette approche ici, nous aurions été contraints de conclure que le programme n'avait aucun impact sur le score de Communication. Toutefois, nous n'avons pas appliqué cette approche parce que a) l'estimation DD associé au premier post-test revêtait un intérêt particulier (c.-à-d. un contraste a priori) et b) une présentation plus détaillée des résultats permet d'extraire un patron cohérent de résultats qui peut ensuite être intégré au cadre plus large de résultats présentés dans ce rapport. En somme, nous avons décidé de laisser parler les données.

<sup>64</sup> Une technique communément utilisée pour réduire le nombre de tests inférentiels effectués est de ne pas tenir compte des contrastes simples lorsque le test de signification global est nul. Il s'agit de réduire le nombre de tests statistiques et, en conséquence, la probabilité de commettre des erreurs de Type-1. Si nous avions appliqué cette approche ici, nous aurions été contraints de conclure que le programme n'avait aucun impact sur le score de Communication. Toutefois, nous n'avons pas appliqué cette approche parce que a) l'estimation DD associé au premier post-test revêtait un intérêt particulier (c.-à-d. un



global de toutes les estimations DD du modèle de régression. Le fait que ce test général ne se soit pas révélé significatif suggère, dans l'ensemble, des trajectoires développementales parallèles pour les trois groupes. Ce résultat donne à penser que l'impact du programme est soit absent, faible ou transitoire. Cette dernière possibilité semble la plus probable, étant donné que la prestation du programme était plus constante au cours de la période précédant immédiatement le premier pré-test (voir le *Rapport de mise en œuvre du projet*, 2014). Nous avons donc prédit que la taille de l'effet programme serait la plus importante au cours de la première période de pré-test et avons vérifié cette notion en examinant les estimations DD individuelles.<sup>65</sup>

Nous avons alors constaté que le groupe programme en garderie démontre invariablement de meilleurs gains pour le domaine de Langage et communication au fil du temps comparativement aux deux groupes témoins. Pour ce qui est du groupe témoin hors garderie, cet avantage n'atteint nullement le seuil de signification statistique et ce, pour aucune des périodes post-test. Comme nous l'avions prévu, les estimations DD les plus importantes ont été observées au premier post-test (4 mois après l'évaluation de base). Seule la comparaison des garderies programme et des garderies témoins produit un effet significatif du point de vue statistique.

Le coefficient associé indique que les enfants du groupe programme en garderie ont, en moyenne, amélioré leurs scores de communication d'environ 0,30 écart-type de plus que le groupe témoin en garderie. La taille de cet effet traitement est compatible avec les références conventionnelles pour un léger et un moyen effet, soit 0,20 et 0,50, respectivement (Cohen, 1988). Les estimations post-test subséquentes révèlent une diminution de l'effet de 50 % et l'absence de signification statistique. En résumé, les gains initiaux résultant de l'exposition au programme testé étaient minimes et n'ont pas persisté après la première évaluation post-test.

---

contraste a priori) et b) une présentation plus détaillée des résultats permet d'extraire un patron cohérent de résultats qui peut ensuite être intégré au cadre plus large de résultats présentés dans ce rapport. En somme, nous avons décidé de laisser parler les données.

<sup>65</sup> Chaque test de signification statistique est associé avec une probabilité différente de zéro de conclure incorrectement que l'effet évalué est réel plutôt qu'une anomalie statistique. Ce type d'erreur s'appelle une erreur de type I. Traditionnellement, la probabilité de commettre une erreur de type I est contrôlée en fixant le seuil nominal de signification statistique à 5 % ( $p < 0,05$ ). Pour chaque test statistique effectué, la probabilité d'avoir commis au moins une erreur de type I s'additionne de sorte que lorsque deux tests sont effectués, la probabilité de commettre au moins une erreur de type I est de 10 % et non plus de 5%. Afin de fournir une information complète, nous rapportons les résultats de tous les tests DD. Plutôt que d'utiliser un ajustement trop conservateur du taux d'erreur nominal de type I, nous avons décidé de rapporter les étendues observées de signification statistique (test bilatéral) en appliquant une erreur nominale de type I non ajustée de 0,05 pour évaluer la signification statistique. Lorsque les participants de la deuxième cohorte seront insérés dans les analyses, la puissance statistique des tests permettra une approche plus conservatrice. En interprétant les résultats présentés dans le présent rapport, le lecteur est libre d'appliquer sa propre correction au taux d'erreur de type I (p. ex. Bonferroni). Une alternative utile pour se protéger contre les erreurs de type I est d'accorder une plus grande confiance aux résultats interprétés qui sont conformes aux prédictions et lorsque qu'ils cadrent de manière cohérente avec l'ensemble des résultats.

**Tableau 5.1 : Première année – Différence des différences (DD) : Effets programme pour les scores de communication standardisés de l'ÉPE-AD (version française, échantillon total)**

	<i>Inclusion séquentielle des covariables – Erreurs groupées par garderie</i>									
	(1)		(2)		(3)		(4)		(5)	
Effets programme	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
Garderie programme v. garderie témoin, au niveau de base (G1 <sup>a</sup> v. G2)	,241	,147	-,184 <sup>m</sup>	,112	,195 <sup>m</sup>	,110	,223 <sup>m</sup>	,110	,183 <sup>m</sup>	,117
DD 1 <sup>er</sup> post-test	-,338*	,144	-,309 <sup>m</sup>	,149	-,313*	,147	-,312*	,146	-,314*	,151
DD 2 <sup>e</sup> post-test	-,194	,149	-,192	,167	-,177	,168	-,176	,171	-,189	,175
DD 3 <sup>e</sup> post-test	-,158	,149	-,139	,160	-,115	,156	-,117	,158	-,135	,153
Garderie programme v. groupe hors garderie, au niveau de base (G1 v. G3)	-,045	,105	,205	,150	-,169	,081	-,108	,081	-,164	,088
DD 1 <sup>er</sup> post-test	-,221	,144	-,217	,143	-,200	,133	-,204	,132	-,223	,137
DD 2 <sup>e</sup> post-test	-,083	,174	-,051	,180	-,032	,175	-,033	,177	-,060	,188
DD 3 <sup>e</sup> post-test	-,105	,195	-,081	,188	-,058	,174	-,070	,172	-,115	,186
Test F composite de Wald	F (6, 15) = 1,44		F (6, 15) = 1,26		F (6, 15) = 1,46		F (6, 15) = 1,60		F (6, 15) = 1,59	
Nombre de participants	243		238		237		237		230	

*Note* : La résultante a été standardisée, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type. **Le groupe programme en garderie (G1) est mis en référence, ce qui signifie que les valeurs négatives des estimations DD indiquent un effet programme positif.** Les erreurs-types ont été estimées grâce au programme SPSS pour le calcul de l'estimateur robuste à l'hétérogénéité de White (White, 1980). De plus, les erreurs-types ont été groupées par garderie pour les ajuster à l'auto-corrélation des résidus dans le temps. Le test F composite de Wald évalue l'effet combiné de toutes les estimations DD.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests bilatéraux, 20 degrés de liberté).

<sup>a</sup> G1, G2 et G3 dénotent le groupe programme en garderie, le groupe témoin en garderie et le groupe témoin hors garderie, respectivement. Les détails de spécification du modèle sont présentés à la section 5.1.2. RÉSULTATS D'IMPACT: Échelles standardisées de l'ÉPE-AD (en français) - Sous-échantillon

L'impact du programme préscolaire sur les cinq autres résultantes de l'ÉPE-AD est évalué dans cette section. L'échelle de Communication est évaluée à nouveau pour le même sous-échantillon de participants, qui exclut les 26 enfants n'ayant pas complété l'ÉPE-AD en français au moins deux fois au cours de la première année. Pour des raisons de parcimonie, nous rapportons seulement les estimations de l'impact du programme qui proviennent du modèle de régression comprenant l'ensemble des covariables. Les résultats omis n'ont pas de répercussions importantes sur l'interprétation des résultats, mais ils sont disponibles sur demande.

Le tableau 5.2 présente les estimations générées par ces analyses de régression pour les échelles standardisées de Communication, de Conscience de soi, d'Habilités cognitives et d'Habilité physique, ainsi que les estimations associées aux sous-échelles improvisées de Vocabulaire réceptif et de Vocabulaire expressif développées par la SRSA. Tel que mentionné précédemment, une seconde série d'analyses des scores de Communication a été faite avec ce sous-échantillon de participants afin d'obtenir, pour toutes les échelles de l'ÉPE-AD, un ensemble de résultats basé sur un échantillon commun de participants. Étant donné les objectifs du programme préscolaire testé, nous prévoyons que les effets programme seront plus importants pour les échelles qui évaluent directement l'habileté des enfants à communiquer et à raisonner en français, soit les échelles de Communication, de Conscience de soi, de vocabulaires et, dans une moindre mesure, d'Habilités cognitives. Par contre, nous ne prévoyons pas d'effets programme sur l'échelle d'Habilité physique puisqu'elle est constituée de tâches qui font appel à des habiletés qui ne sont pas directement liées aux objectifs du programme testé. Les résultats liés à chacune des résultantes sont discutés ci-dessous.

### **Échelle de Communication (standardisée)**

Les résultats obtenus des nouvelles analyses des scores de Communication issus du sous-échantillon de participants sont cohérents avec ceux qui ont été décrits précédemment. Il est à noter que le test de base n'a pas révélé de différence significative entre notre groupe programme et chacun des deux groupes témoins. Nous concluons ainsi qu'avant l'intervention, les groupes étaient équivalents sur cette mesure.

Les effets programme étaient les mêmes qu'auparavant, avec une taille d'effet de 0,30 écart-type pour l'estimation DD issue du premier post-test. Encore une fois, la taille de cet effet diminue considérablement au cours des périodes de temps subséquentes, devenant alors statistiquement non significative. Enfin, les comparaisons visant le groupe témoin hors garderie n'ont pas affiché d'effet significatif.

Certaines différences mineures entre les estimations initiales de l'effet programme et les nouvelles estimations méritent d'être soulignées. D'une part, la taille des effets programme diminue de façon plus marquée avec le temps au cours de cette deuxième série d'analyses, de façon à être presque nulle au moment du troisième post-test. D'autre part, le test combiné de tous les effets DD (c.-à-d. le test F composite) dans cette série d'analyses est marginalement significatif du point de vue statistique, à  $p < 0,10$ . Les conclusions de ces analyses n'ont toutefois pas besoin d'être nuancées : des effets programme modérés mais significatifs ont été observés pour les scores de Communication au premier post-test (quatre mois suivant l'évaluation de base), mais ne persistent pas assez longtemps pour être décelés huit mois après l'évaluation de base (deuxième post-test). Même à la suite d'ajustements statistiques, le groupe programme

garderie ne s'est, à aucun moment, écarté de la trajectoire développementale du groupe témoin hors garderie.

### **Échelle de Conscience de soi (standardisée)**

Les analyses des scores sur l'échelle de Conscience de soi ont mené à des résultats identiques à ceux obtenus avec l'échelle de Communication. Tous les groupes étaient équivalents lors de l'évaluation de base et un avantage statistiquement significatif, de l'ordre de 0,30 écart-type, est observé chez le groupe programme en garderie, comparativement au groupe témoin en garderie, pour la première évaluation post-test seulement. Les erreurs-types inférieurs rapportés pour ces analyses indiquent que les estimations associées sont plus précises que celles associées aux analyses de l'échelle de Communication. Ainsi, cela peut expliquer pourquoi le test combiné des effets DD est, pour ces analyses, statistiquement significatif à 0,05.<sup>66</sup> Les analyses préliminaires rapportées au chapitre de la méthodologie (chapitre 2) indiquent que cette échelle est fortement associée aux mesures de vocabulaire. Nous attribuons donc cet effet à l'amélioration de l'habileté des enfants à communiquer en français.

### **Échelle d'Habilités cognitives (standardisée)**

Les résultats qui précèdent se répètent dans les analyses des scores touchant les habiletés cognitives. Les différences de base entre le groupe programme en garderie et les deux groupes témoins ne sont pas statistiquement significatives. Encore une fois, ce résultat se précise par une interaction statistiquement significative entre les groupes expérimentaux et les périodes de temps (test F composite), interaction provenant de l'effet DD statistiquement significatif issu de la comparaison des deux groupes en garderie. De ces analyses, aucune autre estimation DD ne s'est avérée significative.

### **Échelle d'Habilité physique (standardisée)**

Il existe peu de données probantes suggérant que les enfants francophones en milieu minoritaire sont désavantagés par rapport à leurs pairs en ce qui a trait à leur développement physique. Par conséquent, le programme préscolaire testé n'a pas été conçu pour influencer ces habiletés. Nous ne nous attendions donc pas à observer un effet programme pour cette résultante et nos attentes ont été confirmées.

### **Échelle de Vocabulaire réceptif (standardisée)**

Les analyses de la performance des enfants pour la sous-échelle de Vocabulaire réceptif révélaient que dans l'ensemble, les trajectoires de développement des trois groupes expérimentaux n'étaient pas parallèles, tel qu'indiqué par le test F composite statistiquement significatif de l'effet combiné de toutes les variables DD. Toutefois, ce résultat positif initial n'était pas dû à des effets DD statistiquement significatifs issus de la comparaison entre le groupe programme et les groupes témoins en garderie et hors garderie. Contrairement à nos attentes, les analyses de la sous-échelle de Vocabulaire réceptif n'ont pas révélé d'effet

---

<sup>66</sup> Nous rappelons que, lors de l'analyse des effets d'interaction, le test combiné (test F composite) agit communément à titre de filtre. Selon cette approche, nous n'estimerions et ne testerions les estimations DD que si l'effet global était statistiquement significatif. Si nous avons adopté cette approche, l'échelle de Conscience de soi aurait été la première à démontrer un effet du traitement statistiquement significatif.

programme statistiquement significatif, ni pour la première période post-test, ni pour toute autre période subséquente. L'insuffisance de puissance statistique ne pouvait en être la cause, puisque ni la taille ni la direction des estimations DD ne laissaient entrevoir d'effets. Dans certains cas, les indications d'effets pointaient vers un effet programme positif alors que dans d'autres, ils pointaient vers un impact négatif. Les fluctuations de cette nature laissent fortement suggérer des trajectoires développementales équivalentes.

### **Échelle de Vocabulaire expressif (standardisée)**

Les habiletés évaluées par les mesures de vocabulaire réceptif et expressif sont reconnues comme étant distinctes. Des recherches antérieures ont démontré que chaque mesure a typiquement une relation unique avec les diverses mesures de résultantes (p. ex. Wise, Sevcik, Morris, Lovette & Wolf 2007). Il ne serait donc pas surprenant d'observer un effet programme sur le Vocabulaire expressif en l'absence d'un tel effet sur le Vocabulaire réceptif. Ce type de résultat serait possible peu importe que l'objet des analyses soit l'une des échelles de vocabulaire improvisées ou celles bien validées (p. ex. l'ÉVIP-R et l'EOWPVT-F) présentées au sein de l'analyse d'impact de la deuxième année.

Ainsi, les analyses des scores de Vocabulaire expressif ont révélé, pour la première évaluation post-test, un effet programme relativement important en fonction du groupe témoin en garderie. L'effet observé était de l'ordre de 0,50 écart-type, ce qui correspond à la valeur de référence pour un effet dit moyen. Ce résultat est nuancé par la non-équivalence observée entre ces groupes lors de la période de base où les enfants du groupe programme présentent un désavantage de 0,30 écart-type. Rappelons que, selon le postulat des pentes développementales parallèles (voir la section 5.3 pour une discussion plus approfondie sur ce sujet), l'estimateur DD élimine ces différences initiales dans le calcul des impacts du programme. Invariablement, nous observons une équivalence statistique entre les trajectoires de développement du groupe programme en garderie et du groupe témoin hors garderie.

**Tableau 5.2 : Première année – Différence des différences (DD) : Effets programme pour les scores standardisés des sous-échelles de l'ÉPE-AD (participants ayant répondu aux échelles en français seulement)**

	<i>Analyses par sous-échelle – Erreurs groupées par garderie</i>											
	Communication		Conscience de soi		Habiletés cognitives		Habileté physique		Vocabulaire réceptif		Vocabulaire expressif	
Effets programme (Modèle 5)	DD	ET robuste	DD	ET robuste	DD	ET robuste	DD	ET robuste	DD	ET robuste	DD	ET robuste
Garderie programme v. garderie témoin, au niveau de base (G1 <sup>a</sup> v. G2)	,182	,116	,160	,145	,187	,145	,110	,133	-,098	,180	,319*	,151
DD 1 <sup>er</sup> post-test	-,318*	,149	-,337 <sup>†</sup>	,084	-,279 <sup>†</sup>	,126	-,055	,187	,061	,373	-,543 <sup>†</sup>	,144
DD 2 <sup>e</sup> post-test	-,179	,178	-,257	,149	-,069	,158	-,276 <sup>m</sup>	,154	-,330	,298	-,230	,200
DD 3 <sup>e</sup> post-test	-,094	,148	-,084	,163	-,071	,166	-,217	,164	-,140	,182	-,147	,154
Garderie programme v. groupe hors garderie, au niveau de base (G1 v. G3)	-,135	,094	-,081	,140	,101	,149	-,129	,137	-,429 <sup>m</sup>	,210	-,040	,125
DD 1 <sup>er</sup> post-test	-,175	,133	-,062	,112	-,214 <sup>m</sup>	,115	,019	,229	,111	,361	-,154	,168
DD 2 <sup>e</sup> post-test	,048	,178	-,181	,142	-,132	,143	-,059	,190	,220	,227	,049	,220
DD 3 <sup>e</sup> post-test	,015	,180	,025	,110	-,222	,159	,089	,248	-,017	,153	,096	,139
Test F composite de Wald	F (6, 15) = 2,53 <sup>m</sup>		F (6, 15) = 2,89*		F (6, 15) = 2,93*		F (6, 15) = ,758		F (6, 15) = 2,94*		F (6, 15) = 3,67*	
Nombre de participants	217		215		216		215		216		215	

**Note** : Les résultats ont été standardisés, ce qui veut dire que les effets DD peuvent être interprétés en unités d'écart-type. Les analyses sont basées sur un sous-échantillon de participants qui ont complété l'ÉPE-AD en français au moins deux fois pendant la première année de l'étude. Ainsi, ce sous-échantillon est biaisé en faveur des enfants qui possèdent des compétences langagières relativement fortes en français. **Le groupe programme en garderie (G1) est mis en référence, ce qui signifie que les valeurs négatives des estimations DD indiquent un effet programme positif.** Les erreurs-types ont été estimées grâce au programme SPSS pour le calcul de l'estimateur robuste à l'hétérogénéité de White (White, 1980). De plus, les erreurs-types ont été groupées par garderie pour les ajuster à l'auto-corrélation des résidus dans le temps.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests bilatéraux, 20 degrés de liberté).

<sup>a</sup> G1, G2 et G3 dénotent le groupe programme en garderie, le groupe témoin en garderie et le groupe témoin hors garderie, respectivement. Les détails de spécification du modèle sont présentés à la section 5.1.2.

## Sommaire

Cette section a présenté les résultats de l'impact du programme testé sur un ensemble de résultantes reliées à la préparation à la scolarisation. Le devis expérimental employé prévoyait des analyses centrées sur deux principales comparaisons. La première visait à comparer la trajectoire développementale du groupe programme en garderie à celle du groupe témoin en garderie, tandis que la seconde comparaison opposait le groupe programme en garderie au groupe témoin hors garderie.

La première comparaison visait à déterminer si l'inscription dans un nouveau programme préscolaire de service de garde entraîne de meilleurs résultats que l'inscription dans une garderie dont le programme est typique à ceux habituellement disponibles dans un établissement approuvé par le gouvernement. Pour cette série de comparaisons, un avantage du nouveau programme préscolaire de service de garde a été observé quant aux variables liées aux résultantes attendues (p. ex. Communication, Conscience de soi et Habilités cognitives). De plus, tel que prévu, aucun effet n'a été observé sur l'échelle d'Habilité physique. Ce patron de résultats donne de la crédibilité aux résultats rapportés car il démontre une validité convergente et une validité discriminante.<sup>67</sup> Les analyses des échelles de vocabulaire suggèrent également que les résultats découlent principalement d'items évaluant le Vocabulaire expressif en français. Nous avons obtenu de solides indications de l'impact du programme sur la préparation des enfants à la scolarisation en français.

La question abordée par la deuxième comparaison cherchait à déterminer si l'inscription dans un nouveau programme préscolaire de service de garde engendrait de meilleurs résultats que l'inscription dans un milieu de garderie moins formel. Les résultats des analyses présentées ont offert peu de preuves permettant de confirmer que tel était le cas. Les trajectoires développementales des deux groupes expérimentaux se sont avérées équivalentes à chacune des périodes de temps. Cette équivalence était apparente avant et après avoir exercé un contrôle sur une vaste gamme de variables.

La comparaison la plus intéressante pour évaluer l'impact du nouveau programme préscolaire de service de garde est celle opposant la trajectoire de développement du groupe programme en garderie à celle du groupe témoin en garderie. L'examen du profil de résultats indique que l'avantage du nouveau programme préscolaire de service de garde sur les programmes témoins déjà existants en garderie n'est que transitoire en ce qui concerne les résultantes à l'étude. L'effet programme est plus prononcé, et statistiquement significatif, pour le premier post-test (quatre mois suivant le niveau de base), tandis qu'il est plus faible et non significatif pour les périodes post-test suivantes.

---

<sup>67</sup> Le terme validité convergente fait ici référence au cas où l'attente d'une association entre deux variables est confirmée. À l'opposé, le terme validité discriminante fait référence au cas où l'attente d'une absence d'association (ou association négligeable) est confirmée. Dans cette étude, ni les chercheurs ni les participants n'étaient aveugles au traitement. Dans cette situation, la validité des résultats est renforcée si les effets programme sont spécifiques aux résultantes ciblées (c.-à-d. quand la validité convergente et la validité discriminante sont toutes deux présentes). Cette affirmation repose sur le postulat que les différentes sources de biais affecteraient toutes les variables de façon semblable.<sup>68</sup> La modération et la médiation sont des concepts apparentés. L'idée derrière le concept de modération est qu'on utilise une variable pour contrôler l'effet d'une autre variable, soit en augmentant ou en diminuant l'importance de son effet. On a recours à la médiation lorsqu'une variable donnée, dans ce cas-ci le groupe programme, exerce indirectement son effet sur une résultante par l'entremise d'une autre variable, dans ce cas-ci la fidélité/qualité du programme.

### 5.1.3 Modérateurs et médiateurs de l'effet d'intervention

Dans la section précédente, nous avons présenté les résultats des principales analyses qui évaluaient l'impact de l'appartenance aux groupes expérimentaux pour quatre résultantes liées à la préparation à la scolarisation et deux mesures improvisées de vocabulaire. Nous avons enrichi cet ensemble de résultats au moyen d'analyses supplémentaires incluant, cette fois, des variables pouvant modérer ou « interagir avec » l'effet programme. Autrement dit, nous avons examiné des prédicteurs de l'importance de l'effet programme dans une série d'analyses de régression. Cette approche permet d'aborder la question de recherche initiale, c'est-à-dire : « le programme testé a-t-il eu un effet positif sur l'échantillon actuel d'enfants » sous un angle légèrement différent, soit : « sur qui l'impact du programme testé a-t-il été le plus important ? » Cette approche plus flexible permet de découvrir des effets programme importants qui n'étaient pas mis en évidence par les principales analyses d'impact. Si nous réussissons à prédire cet aspect des résultats, nous renforcerons les preuves appuyant notre affirmation voulant que le programme testé ait un impact réel sur le développement des enfants.

Ces analyses supplémentaires ont été effectuées au moyen de trois types de variables modératrices : l'exposition au programme ou le dosage, la fidélité/qualité de la mise en œuvre du programme préscolaire de service de garde, et les caractéristiques linguistiques de l'échantillon. Le premier type de variable cherche à évaluer si le temps passé en garderie exerce une influence quant à l'impact du programme. Le deuxième type de variable a été traité lors des analyses comme « médiateur », un concept apparenté mais distinct.<sup>68</sup> Enfin, les analyses effectuées quant aux caractéristiques linguistiques de l'échantillon sont pertinentes à la question de la validité externe. Si le nouveau programme préscolaire de service de garde doit être implanté dans d'autres contextes, il est important de savoir pour quels groupes il sera le plus efficace.

#### ***Définitions précises du dosage : Moyenne d'heures passées en garderie***

Selon l'hypothèse sous-tendant les principales analyses d'impact rapportées à la section précédente, l'inscription dans un nouveau programme préscolaire de service de garde favorisera le développement quant aux résultantes associées à la préparation à la scolarisation. Cette hypothèse de recherche suppose une exposition suffisante au programme pour que l'intervention ait l'effet désiré. Selon cette hypothèse, le rapport entre le degré d'exposition à la garderie et les variables résultantes dépendra également du type de programme en garderie auquel est exposé l'enfant. Plus particulièrement, nous prévoyons que l'exposition au nouveau programme préscolaire de service de garde produira de meilleures résultantes développementales que les programmes offerts dans le groupe témoin en garderie. Autrement dit, une heure d'exposition au nouveau programme préscolaire de service de garde devrait valoir davantage en terme de résultantes développementales que le même temps d'exposition aux programmes actuels des groupes témoins en garderie. Nous portons maintenant notre attention sur une série d'analyses conçues pour vérifier ces hypothèses.

---

<sup>68</sup> La modulation et la médiation sont des concepts apparentés. L'idée derrière le concept de modulation est qu'on utilise une variable pour contrôler l'effet d'une autre variable, soit en augmentant ou en diminuant l'importance de son effet. On a recours à la médiation lorsqu'une variable donnée, dans ce cas-ci le groupe programme, exerce indirectement son effet sur une résultante par l'entremise d'une autre variable, dans ce cas-ci la fidélité/qualité du programme.



## Détails de spécification

L'effet du dosage a été évalué en modifiant la spécification utilisée dans les principales analyses d'impact pour le dosage, soit en insérant des effets fixes pour le dosage. La variable dosage a été opérationnalisée comme la moyenne d'heures par semaine qu'un enfant a passé en garderie durant la période de quatre mois précédant immédiatement l'évaluation.<sup>69</sup> Les enfants qui n'avaient pas passé de temps en garderie se sont vus attribuer la valeur zéro pour cette variable, qui a été centrée à la moyenne avant son insertion dans le modèle pour éviter une dégradation inutile des résultats due à un problème de multicollinéarité (Cohen et al., 2003, pp. 261-266). Nous avons décidé de centrer cette variable selon les moyennes obtenues à des temps spécifiques plutôt que selon la moyenne générale afin de prendre en compte les variations saisonnières de l'exposition à la garderie.

Les cinq spécifications du modèle de régression décrites à la section 5.1.2 ont été modifiées pour inclure: a) l'effet direct du dosage, b) l'interaction entre le dosage et le temps, et c) l'interaction triple entre le dosage, le temps et l'appartenance aux groupes expérimentaux. L'effet qui nous intéresse plus particulièrement est celui de l'interaction triple, car nous tenons à estimer l'effet du dosage sur les gains développementaux des enfants pour chaque groupe de garderie et à établir la différence entre les deux estimations. En termes techniques, l'estimation de cet effet consiste d'abord à estimer les effets DD du dosage (temps x dosage), séparément pour les deux groupes en garderie. L'étape suivante consiste à établir la différence entre les estimations DD obtenues. Ce second calcul permet d'obtenir ce qu'on appelle l'estimation par triple différence de différences (DDD; pour un exemple de l'emploi des estimateurs DDD aux fins de l'évaluation d'un programme, voir Monheit & Steinburg Schone, 2004). À l'instar des estimations DD rapportées aux sections précédentes, une estimation DDD est rapportée pour la comparaison entre le groupe programme en garderie et le groupe témoin en garderie à chaque période post-test. Un test global de l'effet combiné des estimations DDD est également rapporté: le test F analyse l'interaction triple entre le temps, le groupe expérimental et le dosage. Nous étions particulièrement intéressés à la performance des deux groupes en garderie et, par conséquent, nous avons spécifiquement examiné leur contribution à l'effet d'interaction combiné. Autrement dit, les modèles de régression ont été appliqués à tout l'échantillon des participants, mais les comparaisons ayant trait au groupe témoin hors garderie n'ont pas été prises en compte.<sup>70</sup>

---

<sup>69</sup> L'autre option envisagée consistait à définir le dosage en fonction du nombre total d'heures d'exposition au milieu de garderie au cours de l'étude. Cette seconde option présente certaines complications, car l'insertion d'une telle variable dans l'analyse de régression nécessiterait une transformation (p. ex. logarithmique) afin de tenir compte de sa distribution asymétrique et de sa relation potentiellement non linéaire avec les mesures de résultantes (c.-à-d. une diminution de l'importance du dosage en fonction d'une exposition croissante). De plus, les analyses corrélationnelles préliminaires ont suggéré que les deux définitions de l'exposition prises en compte ici étaient corrélées au-delà de 0,76 pour les quatre premières périodes d'évaluation (Spearman Rho). Étant donné la quasi-équivalence des deux définitions, nous avons opté pour la conceptualisation la plus simple, qui dans tous les cas évalue le dosage durant un intervalle raisonnablement prolongé (quatre mois).<sup>70</sup> L'échantillon total a été utilisé pour cette analyse afin d'augmenter l'efficacité des tests de signification. Les estimations rapportées sont néanmoins propres à la comparaison G1 v. G2. Les comparaisons ayant trait au groupe témoin hors garderie (G3) étaient difficiles à interpréter, en raison du fait que la distribution de la variable dosage pour ce groupe différait énormément de celle des deux autres groupes. Les estimations découlant de ces comparaisons n'ont par conséquent pas été rapportées.<sup>71</sup> Les détails du calcul de ces indices se trouvent à la section 2.5.6 ainsi que dans le document d'accompagnement le *Rapport de mise en œuvre du projet* (Bérubé et al., 2014).

<sup>70</sup> L'échantillon total a été utilisé pour cette analyse afin d'augmenter l'efficacité des tests de signification. Les estimations rapportées sont néanmoins propres à la comparaison G1 v. G2. Les comparaisons ayant trait au groupe témoin hors garderie (G3) étaient difficiles à interpréter, en raison du fait que la distribution de la variable dosage pour ce groupe différait énormément de celle des deux autres groupes. Les estimations découlant de ces comparaisons n'ont par conséquent pas été

La distribution de la variable dosage était statistiquement équivalente pour les deux groupes en garderie à l'intérieur d'une même période de temps ainsi que d'une période de temps à l'autre. Cette équivalence des distributions est une condition essentielle pour assurer la validité des estimations DDD. Des différences entre les distributions des groupes pourraient avoir un effet significatif, mais celui-ci serait difficile à interpréter (d'où la décision de ne pas interpréter les effets ayant trait au groupe témoin hors garderie). Précisément, l'effet observé pourrait découler de cette différence initiale dans la distribution des scores plutôt que d'une cause d'intérêt significatif. Par conséquent, les estimations DDD associées ne pourraient pas être interprétées de manière raisonnable. En dernier lieu, il est à noter que la puissance statistique des tests d'interaction triple (p. ex. les estimations DDD) est moins grande que celle des tests bilatéraux (c.-à-d. les estimations DD) rapportés précédemment, en raison du plus grand nombre de degrés de liberté et de l'erreur incluse dans la mesure de la variable dosage.

### **Impact du dosage sur les scores de Communication standardisés (échantillon total)**

Les estimations DDD des effets du dosage sont rapportées dans le tableau 5.3. Pour la période de base, l'effet estimé du dosage était négatif pour l'échantillon total, mais faible et statistiquement non significatif, soit  $b=-0,006$ ,  $ET=0,007$  (effet principal du dosage non rapporté au tableau 5.3). Pour cette évaluation, le degré d'exposition au nouveau programme préscolaire de service de garde n'a pas semblé être lié à la performance sur l'échelle de Communication.

En portant notre attention sur les résultats présentés dans le tableau, nous observons que les estimations DDD de l'effet du dosage sont constamment plus positives pour le groupe programme que pour le groupe témoin en garderie, qui présente des estimations dont les valeurs sont négatives. Les différences obtenues de cette comparaison ne s'avèrent statistiquement significatives que pour la première évaluation post-test, mais cet effet est robuste à l'inclusion des covariables. Le patron des résultats est le même ici que celui obtenu antérieurement lorsque le dosage était défini en fonction de l'appartenance aux groupes expérimentaux.

Les estimations DDD présentées au tableau 5.3 doivent être situées dans un contexte plus large pour être bien comprises. Nous fournissons ce contexte dans ce qui suit en présentant des estimations de l'effet dosage à différents niveaux d'analyse. Les estimations en question découlent du modèle 5, qui inclut toute la liste des covariables.

La taille de l'effet dosage en fonction du temps et du groupe expérimental est d'un intérêt particulier. Les estimations DD de l'effet dosage pour le groupe programme en garderie étaient positives pour toutes les périodes post-test. Par contre, à aucun moment l'effet dosage DD n'est devenu statistiquement significatif pour le groupe programme en garderie. Les estimations DD étaient de  $b=0,013$ ,  $ET=0,010$ ;  $b=0,009$ ,  $ET=0,009$ ; et  $b=0,022$ ,  $ET=0,013$  pour les première, deuxième et troisième évaluations post-test, respectivement (les résultats de temps x dosage sont non rapportés au tableau 5.3). Ces estimations nous permettent de conclure que les divers degrés d'exposition au nouveau programme en garderie préscolaire ont peu d'effet sur les scores de Communication.

---

rapportées.<sup>71</sup> Les détails du calcul de ces indices se trouvent à la section 2.5.6 ainsi que dans le document d'accompagnement le *Rapport de mise en œuvre du projet* (Bérubé et al., 2014).

Ce qui semble entraîner l'effet DDD rapporté au tableau 5.3 est l'effet négatif du dosage pour le groupe témoin en garderie par rapport au groupe programme en garderie. Tel qu'indiqué plus tôt, cet effet est propre à l'estimation DDD pour la première période post-test.

À titre d'exemple concret, l'estimation DD de l'effet dosage pour la première période post-test démontre que chaque dix heures supplémentaires par semaine passées dans une garderie programme est associée à une augmentation attendue du score de Communication de 0,13 écart-type ( $0,13=10 \text{ heures} \times 0,013$ ). Cet effet n'est pas statistiquement significatif, mais le score équivalent attendu pour le groupe témoin en garderie est beaucoup plus bas et négatif (-0,014). Il importe de souligner que les estimations du modèle suggèrent une augmentation de la taille d'effet en fonction de l'augmentation du degré d'exposition à la garderie. À 10 heures d'exposition au-dessus de la moyenne de l'échantillon, le score de Communication attendu d'un enfant inscrit dans le groupe témoin en garderie est de 0,27 écart-type plus bas ( $0,27=10 \text{ heures} \times 0,027$ ) que celui d'un enfant inscrit au groupe programme en garderie. À 20 heures d'exposition au-dessus de la moyenne, la différence augmente à 0,54 écart-type ( $0,54=20 \text{ heures} \times 0,027$ ), soit un effet dit moyen. En somme, l'ampleur de l'effet programme sur le développement augmente en fonction de la moyenne d'heures passées en milieu de garde. Pour les degrés d'exposition les plus élevés, l'ampleur de l'effet estimé par le modèle est substantielle. La distribution des effets résiduels ne suggère pas la présence d'une relation non linéaire, ce qui est cohérent avec la littérature qui indique que les effets du dosage sont linéaires (National Institute of Child Health and Human Development Early Child Care Research Network, 2003).

**Tableau 5.3 : Première année – Effets programme basés sur l’estimateur DDD (temps x groupe x dosage) et sur les scores de Communication de l’ÉPE–AD (version française, échantillon total)**

	<i>Inclusion séquentielle des covariables – Erreurs groupées par garderie</i>									
	(1)		(2)		(3)		(4)		(5)	
Effets programme	b	ET robuste	b	ET robuste	b	ET robuste	b	ET robuste	b	ET robuste
G1 v. G2 : Différence de la taille de l’effet du dosage au niveau de base	-,002	,017	,001	,017	,002	,012	,003	,012	,002	,011
DDD 1 <sup>er</sup> post-test	-,031 <sup>†</sup>	,010	-,033*	,012	-,026 <sup>m</sup>	,013	-,029*	,012	-,027*	,013
DDD 2 <sup>e</sup> post-test	-,005	,015	-,003	,014	-,003	,013	-,003	,012	-,002	,012
DDD 3 <sup>e</sup> post-test	-,013	,018	-,011	,019	-,017	,018	-,018	,018	-,019	,017
Tests F de Wald pour interaction triple :	F (3, 18) = 4,45*		F (3, 18) = 3,91*		F (3, 18) = 1,74		F (3, 18) = 1,91		F (3, 18) = 2,36	
Nombre de participants :	243		240		239		239		232	

**Note :** La résultante a été standardisée, c’est-à-dire que les effets peuvent être interprétés en unités d’écart-type. Le groupe programme en garderie (G1) est mis en référence, ce qui signifie que les valeurs négatives des estimations DDD indiquent un effet programme positif. Les erreurs-types ont été estimées grâce au programme SPSS pour le calcul de l’estimateur robuste à l’hétérogénéité de White (White, 1980). De plus, les erreurs-types ont été groupées par garderie pour les ajuster à l’auto-corrélation des résidus dans le temps. Les détails de spécification du modèle sont présentés à la section 5.1.3.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests bilatéraux, 20 degrés de liberté).

<sup>a</sup> Ici, l’exposition est définie comme la moyenne d’heures passées en garderie par semaine durant les 4 mois précédant l’évaluation. <sup>b</sup> G1 dénote ici le groupe programme en garderie et G2 dénote le groupe témoin en garderie. Les enfants du G3 (groupe témoin hors garderie) ont participé à l’estimation du modèle, mais les comparaisons associées ne sont pas rapportées. Le test F d’interaction triple repose sur les groupes G1 et G2 seulement.

### Impact du dosage sur les échelles standardisées de l'ÉPE-AD (sous-échantillon)

Nous avons appliqué la stratégie d'analyse de la section précédente pour évaluer l'importance du dosage en tant que modérateur de l'effet programme sur les six échelles de l'ÉPE-AD. Les analyses reposent sur le sous-échantillon de participants qui avaient complété l'évaluation en français un nombre suffisant de fois. Nous ne rapportons que les résultats des analyses de régression effectuées avec la série complète de covariables. Les estimations DDD des différences d'effet entre les groupes de garderie apparaissent au tableau 5.4.

L'examen des estimations DDD révèle que l'effet observé pour l'échelle de Communication se répète pour ce sous-échantillon de participants. Pour les autres résultats, nous avons vérifié si le signe des coefficients correspondait à un effet programme positif. Les effets se sont avérés positifs pour toutes les échelles, sauf pour celles d'Habilité physique et de Vocabulaire réceptif. Particulièrement, l'estimation de l'effet programme pour la sous-échelle de Vocabulaire expressif est de taille comparable à celle obtenue avec l'échelle de Communication, soit d'un gain estimé de 0,026 écart-type par heure d'exposition à la garderie. Par contre, cet effet ne s'est pas avéré statistiquement significatif. Comme pour les analyses précédentes, les résultats des deuxième et troisième évaluations post-test ne sont pas significatifs sur le plan statistique. Nous ne discuterons pas plus en détails de l'effet associé à l'échelle de Communication parce que les analyses détaillées apparaissent à la section précédente. De surcroît, les autres estimations DDD ne sont pas décrites plus en détails, car elles ne sont pas statistiquement significatives.

Arrêtons-nous quelques instants pour examiner l'effet de l'exposition à la garderie indépendamment de l'appartenance aux groupes. Pour toutes les échelles, l'effet direct du dosage au niveau de base était faible et statistiquement non significatif. L'estimation de cet effet allait comme suit pour chaque échelle: Conscience de soi :  $b = -0,014$ ,  $ET = 0,008$ ; Habiletés cognitives :  $b = -0,013$ ,  $ET = 0,011$ ; Habileté physique :  $b < 0,001$ ,  $ET = 0,016$ ; sous-échelle de Vocabulaire réceptif :  $b = 0,003$ ,  $ET = 0,006$ ; et enfin, sous-échelle de Vocabulaire expressif :  $b = -0,014$ ,  $ET = 0,008$  (les résultats de l'effet principal sont non rapportés au tableau 5.4). Nous avons réévalué le modèle de régression, cette fois en ignorant le terme d'interaction triple incluant les variables groupe, temps et dosage. Le but de ces analyses secondaires était de vérifier si l'importance de l'exposition à la garderie variait en fonction du temps. Ce ne fut pas le cas avec l'échelle de Conscience de soi, car le test F d'interaction bilatérale entre le temps et le dosage n'était pas statistiquement significatif,  $F \text{ Wald } (3,18) = 0,797$ . Dans le même ordre d'idées, l'interaction entre le temps et le dosage n'était pas statistiquement significative pour l'échelle d'Habilités cognitives,  $F \text{ Wald } (3,18) = 1,06$ , ou l'échelle d'Habilité physique,  $F \text{ Wald } (3,18) = 0,733$ . On a observé qu'avec le temps, l'effet du dosage était également invariable pour l'échelle de Vocabulaire réceptif,  $F \text{ Wald } (3,18) = 0,147$ , et l'échelle de Vocabulaire expressif,  $F \text{ Wald } (3,18) = 1,46$  (les résultats de l'interaction temps x dosage sont non rapportés au tableau 5.4). En somme, ces analyses secondaires offrent peu d'indices suggérant que la moyenne d'heures passées en garderie au cours des quatre mois précédents l'évaluation représentait un prédicteur utile des résultantes associées à la préparation à la scolarisation ou des scores de vocabulaire. Les seules exceptions correspondent à l'effet DDD observé pour l'échelle de Communication et, dans une moindre mesure, pour la sous-échelle de Vocabulaire expressif (c.-à-d. la tendance non significative associée).

**Tableau 5.4 : Première année – Effets programme basés sur l’estimateur DDD des effets programme (temps x groupe x dosage) pour les scores des sous-échelles standardisées de l’ÉPE–AD (participants ayant répondu aux échelles en français seulement)**

	<i>Analyses par sous-échelle – Erreurs groupées par garderie</i>											
	Communication		Conscience de soi		Habiletés cognitives		Habileté physique		Vocabulaire réceptif		Vocabulaire expressif	
Effets programme	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
G1v. G2 : Différence de la taille de l’effet du dosage au niveau de base	,001	,011	,004	,013	,012	,014	-,001	,017	-,020	,011	,006	,014
DDD 1 <sup>er</sup> post-test	-,026*	,013	-,015	,014	-,019	,013	,001	,024	,001	,015	-,027	,016
DDD 2 <sup>e</sup> post-test	-,006	,012	-,006	,016	,002	,016	,007	,030	,004	,016	,008	,019
DDD 3 <sup>e</sup> post-test	-,029 <sup>m</sup>	,015	-,038 <sup>m</sup>	,021	-,053	,018	-,028	,024	-,001	,016	-,035	,021
Test F composite de Wald :	F (3, 18) = 2,67 <sup>m</sup>		F (3, 18) = 2,92 <sup>m</sup>		F (3, 18) = 5,38 <sup>†</sup>		F (3, 18) = 6,19 <sup>†</sup>		F (3, 18) = ,113		F (3, 18) = 4,95*	
Nombre de participants :	219		217		219		218		218		217	

*Note* : La résultante a été standardisée, c’est-à-dire que les effets peuvent être interprétés en unités d’écart-type. **Le groupe programme en garderie (G1) est mis en référence, ce qui signifie que les valeurs négatives des estimations DDD indiquent un avantage pour ce groupe.** Les erreurs-types ont été estimées grâce au programme SPSS pour le calcul de l’estimateur robuste à l’hétérogénéité de White (White, 1980). De plus, les erreurs-types ont été groupées par garderie pour les ajuster à l’auto-corrélation des résidus dans le temps. Les détails de spécification du modèle sont présentés à la section 5.1.3.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests bilatéraux, 20 degrés de liberté).

<sup>a</sup> Ici, l’exposition est définie comme la moyenne d’heures passées en garderie par semaine durant les 4 mois précédant l’évaluation. G1 dénote ici le groupe programme en garderie et G2 dénote le groupe témoin en garderie

## Sommaire

En ce qui concerne le degré d'exposition à la garderie, le patron de résultats obtenu est compatible avec l'idée que le nouveau programme préscolaire de service de garde a eu un effet protecteur. Toutefois, cet *effet était propre aux scores de Communication standardisés*. Pour cette échelle, de plus hauts degrés d'exposition au groupe témoin en garderie ont été associés à des changements développementaux qui étaient moins positifs et, en fait, légèrement négatifs. À l'opposé, on estime que le degré d'exposition au nouveau programme préscolaire de service de garde est un facteur neutre, occasionnant un effet légèrement positif et non significatif. Les résultats n'indiquent pas un effet programme extrêmement positif. Cependant, ils donnent à penser que pour la période de quatre mois précédant le premier post-test, le programme préscolaire de service de garde testé a eu un effet positif par rapport aux programmes existants des garderies témoin.

Certes, les analyses du dosage n'ont pas corroboré tous les effets qui ont été observés avec les comparaisons simples de groupes. Alors que l'estimation DDD pour la sous-échelle de Vocabulaire expressif était appréciable (bien que non significative), aucune des autres estimations associées à la première période post-test n'a suggéré la présence de l'effet prévu. Puisque ces tests n'étaient pas assez puissants pour détecter les petits effets, l'incapacité de répliquer de façon exacte le patron de résultats observé lors des principales analyses d'impact ne doit pas servir de preuve contre la validité de la première série de résultats. Néanmoins, pour les raisons énumérées ci-dessus, de tels résultats auraient donné du poids à leur validité. Au point où nous en sommes, nous n'avons obtenu ces preuves positives que pour l'échelle de Communication.

### ***Définitions précises de l'intégrité du programme: Fidélité et qualité du programme en garderie***

L'hypothèse qui sous-tend les principales analyses d'impact rapportées dans la première section de ce chapitre veut que l'inscription au nouveau programme préscolaire de service de garde favorise le développement des résultantes associées à la préparation à la scolarisation. Cette formulation de l'hypothèse de recherche suppose une exposition suffisante pour que le programme ait les effets désirés. Dans cette section, l'appartenance à l'un des trois groupes expérimentaux a été remplacée, dans le modèle de régression, par des indices liés au niveau et à la qualité du programme testé. Ces indices font partie d'un modèle développé par Dane et Schneider (1998) pour analyser l'intégrité d'un programme ou d'une intervention (pour connaître les détails de ces mesures, voir la section 2.5.6). Les indices de fidélité font référence à la proportion des éléments de programme qui sont correctement mis en place. Les indices de qualité permettent de faire des distinctions plus fines. Alors que les indices de fidélité examinent la présence ou l'absence des volets du programme, les indices de qualité reflètent la mesure dans laquelle les volets centraux du programme étaient appliqués avec soins et précision sur une échelle de sept points, où 1 indique un niveau de soins inadéquat et 7 indique un niveau de soins excellent.

Nous prévoyons que les garderies programme et les garderies témoin se distingueront les unes des autres tant au niveau de la fidélité qu'au niveau de la qualité du programme. Cette hypothèse est testée au moyen d'une première série d'analyses. De plus, nous croyons que le

degré de fluctuation entre les garderies quant aux aspects mesurés de la fidélité et de la qualité du programme sera un prédicteur des résultantes développementales. Cette hypothèse est testée dans une deuxième série d'analyses. Si elle est confirmée, nous pourrions appuyer l'affirmation selon laquelle les dimensions de fidélité et de qualité ciblées par le programme ont un impact sur le développement des enfants. Comme nous l'avons vu à la section 4, ces analyses peuvent être vues comme une version plus sensible des analyses DD par groupe expérimental. Les résultats des analyses menées sur la fidélité et la qualité du programme sont présentés dans des sections séparées.

### Détails de spécification et présentation des résultats

Deux séries d'analyses, de type qualitativement différent, ont été menées. La *première série d'analyses* a employé les sites de garderie comme unité d'analyse. Son but est d'évaluer dans quelle mesure les analyses menées en fonction des groupes expérimentaux représentent une comparaison de garderies qui diffèrent véritablement en ce qui a trait aux dimensions de fidélité et de qualité examinées ici. En d'autres mots, cette première série d'analyses permet une certaine vérification de la manipulation employée pour le programme testé, soit un test formel de différenciation (voir aussi les analyses de différenciation décrites dans le *Rapport de mise en œuvre du projet*, 2014). La *deuxième série d'analyses* tire ses observations des enfants pris individuellement et peut être considérée comme une reproduction des analyses par groupe expérimental. Le caractère distinctif de la deuxième série d'analyses est que les calculs de l'estimateur DD sont effectués à l'aide des estimations de la fidélité et de la qualité du programme, plutôt qu'à l'aide de l'appartenance aux groupes expérimentaux. Nous croyons que cette manière alternative de conceptualiser l'effet programme produira le même patron de résultats que les analyses menées en fonction des groupes expérimentaux. Enfin, nous prévoyons que les différences entre les deux groupes en garderie seront considérablement réduites une fois que la fidélité et la qualité du programme auront été contrôlées, validant ainsi le résultat.

Nous avons donc mené une *première série d'analyses* pour évaluer: a) la fidélité et la qualité avec lesquelles le programme testé a été mis en œuvre dans les garderies programme; et b) la mesure dans laquelle les garderies du groupe programme se distinguaient des garderies témoins quant aux indices de fidélité et de qualité. Tel qu'indiqué à la section 2.5.6, nous faisons une distinction entre deux types d'indices de fidélité et quatre types d'indices de qualité. Les dimensions de fidélité sélectionnées étaient celles de la fidélité structurelle et la fidélité de contenu, tandis que les dimensions de qualité sélectionnées étaient celles de la qualité structurelle, la qualité éducative, la sensibilité des éducatrices et la qualité de lecture.<sup>71</sup> En ce qui concerne la fidélité, nous étions en mesure d'estimer le degré de changement pour les groupes programmes en garderie au fil du temps. Dans tous les cas, nous avons pu estimer une différence entre les garderies programmes et les garderies témoins.

Comme les questions d'échantillonnage par groupes ne sont pas pertinentes à ces analyses, nous avons mesuré les effets au moyen de tests de signification statistique non ajustés. Les tests de rang paramétriques et non paramétriques sont rapportés comme une vérification contre la violation des postulats de distribution pour ce petit échantillon. Une caractéristique distinctive des tests non paramétriques est qu'ils sont sensibles au rang relatif des scores de qualité, mais

---

<sup>71</sup> Les détails du calcul de ces indices se trouvent à la section 2.5.6 ainsi que dans le document d'accompagnement le *Rapport de mise en œuvre du projet* (Bérubé et al., 2014).



non à leurs différences absolues.<sup>72</sup> Les tests t corrélés (paramétriques) et les tests de la somme des rangs de Wilcoxon (non paramétriques) sont rapportés en tant que tests du changement en fonction du temps. Le test de Welch (un test t qui ne suppose pas l'homogénéité de la variance et la présence d'échantillons de taille égale) et la statistique U de l'ordre des rangs de Mann-Whitney (non paramétrique) sont rapportés pour présenter les différences qui existent entre des groupes indépendants.<sup>73</sup> Le test r de Pearson (paramétrique) et le rang tau de Kendall (non paramétrique) sont rapportés comme mesures d'association.

Le petit nombre de garderies (N=15) signifie que les tests de signification ne sont pas assez puissants pour détecter de manière fiable les effets réels.<sup>74</sup> Pour pallier ce manque de puissance, nous avons fixé le niveau de probabilité nominale de signification statistique à 0,10 plutôt qu'à 0,05 pour ces analyses. Quoi qu'il en soit, on devrait interpréter avec soins les effets nuls dans cette situation. Enfin, nous soulignons que les résultats apparaissant dans cette sous-section peuvent différer de ceux qui sont présentés ailleurs.<sup>75</sup> Cette divergence découle du fait que ces analyses reposent sur le sous-échantillon des communautés incluses dans les analyses d'impact.

Dans une *seconde série d'analyses*, nous avons estimé l'impact des indices de fidélité et de qualité sur les résultantes en matière de développement pour les enfants. Pour ces analyses, nous avons employé les mêmes spécifications et la même stratégie d'analyse que celles qui ont servi

---

<sup>72</sup> D'importantes différences dans la taille de l'effet estimé par ces deux méthodes indiqueraient que les différences relatives entre les scores sont pertinentes à la détermination de l'importance de l'effet. Un tel résultat est instructif, mais n'indique pas nécessairement que le résultat du test paramétrique pose problème. Une inspection visuelle a démontré que la distribution des scores dans chaque cellule était raisonnable compte tenu du petit nombre d'observations.<sup>73</sup> Des tests indépendants ont été menés pour chaque période de temps. Nous avons décidé de ne pas effectuer une analyse plus synthétique (p. ex. ANOVA avec mesures répétées) dû au fait que les scores de fidélité pour les garderies du groupe témoin sont invariables pour le pré-test et la première période post-test en raison de la stratégie d'imputation adoptée ici.<sup>74</sup> Dans le *Rapport de mise en œuvre du projet* (Bérubé et al., 2014), des analyses descriptives ont été effectuées au niveau des classes prises individuellement plutôt que des garderies. Une autre stratégie était nécessaire pour les analyses quantitatives rapportées dans le présent document en raison du fait que les classes étaient emboîtées au sein des lieux des garderies ce qui, si négligé, pourrait conduire à une surestimation de la précision des estimations d'effet. Nous avons réglé le problème en regroupant les données des classes pour obtenir un seul score pour chaque lieu de garderie.<sup>75</sup> *Rapport de mise en œuvre du projet* (Bérubé et al., 2014). Dans le présent rapport, nous avons employé les moyens « flexibles » décrits dans le rapport sous-mentionné et regroupé les données de classe au niveau du lieu de garderie.<sup>76</sup> On trouvera une description plus détaillée des indices de qualité à la section 5.6.

<sup>73</sup> Des tests indépendants ont été menés pour chaque période de temps. Nous avons décidé de ne pas effectuer une analyse plus synthétique (p. ex. ANOVA avec mesures répétées) dû au fait que les scores de fidélité pour les garderies du groupe témoin sont invariables pour le pré-test et la première période post-test en raison de la stratégie d'imputation adoptée ici.<sup>74</sup> Dans le *Rapport de mise en œuvre du projet* (Bérubé et al., 2014), des analyses descriptives ont été effectuées au niveau des classes prises individuellement plutôt que des garderies. Une autre stratégie était nécessaire pour les analyses quantitatives rapportées dans le présent document en raison du fait que les classes étaient emboîtées au sein des lieux des garderies ce qui, si négligé, pourrait conduire à une surestimation de la précision des estimations d'effet. Nous avons réglé le problème en regroupant les données des classes pour obtenir un seul score pour chaque lieu de garderie.<sup>75</sup> *Rapport de mise en œuvre du projet* (Bérubé et al., 2014). Dans le présent rapport, nous avons employé les moyens « flexibles » décrits dans le rapport sous-mentionné et regroupé les données de classe au niveau du lieu de garderie.<sup>76</sup> On trouvera une description plus détaillée des indices de qualité à la section 5.6.

<sup>74</sup> Dans le *Rapport de mise en œuvre du projet* (Bérubé et al., 2014), des analyses descriptives ont été effectuées au niveau des classes prises individuellement plutôt que des garderies. Une autre stratégie était nécessaire pour les analyses quantitatives rapportées dans le présent document en raison du fait que les classes étaient emboîtées au sein des lieux des garderies ce qui, si négligé, pourrait conduire à une surestimation de la précision des estimations d'effet. Nous avons réglé le problème en regroupant les données des classes pour obtenir un seul score pour chaque lieu de garderie.<sup>75</sup> *Rapport de mise en œuvre du projet* (Bérubé et al., 2014). Dans le présent rapport, nous avons employé les moyens « flexibles » décrits dans le rapport sous-mentionné et regroupé les données de classe au niveau du lieu de garderie.<sup>76</sup> On trouvera une description plus détaillée des indices de qualité à la section 5.6.

<sup>75</sup> *Rapport de mise en œuvre du projet* (Bérubé et al., 2014). Dans le présent rapport, nous avons employé les moyens « flexibles » décrits dans le rapport sous-mentionné et regroupé les données de classe au niveau du lieu de garderie.<sup>76</sup> On trouvera une description plus détaillée des indices de qualité à la section 5.6.

aux analyses par groupe expérimental, à la différence près que les indices de fidélité et de qualité remplaçaient le facteur «groupe expérimental » dans l'estimation des effets. L'impact de chaque indice de qualité/fidélité a été estimé dans le cadre d'analyses indépendantes. Ainsi, nous n'avons pas vérifié les effets communs de ces indices. Nous voulions surtout déterminer si l'un des indices pouvait se démarquer en prédisant les résultantes développementales. Pour cette section, nous avons également rapporté les résultats d'une deuxième série d'analyses, où nous avons simultanément intégré les effets DD pour l'effet du groupe expérimentaux et pour les indices de fidélité ou de qualité. En contrôlant pour la variance prédictive partagée par les groupes expérimentaux et la qualité/fidélité, nous désirions vérifier à quel point les effets du groupe expérimental rapportés précédemment diminueraient avec l'inclusion des indices de fidélité/qualité. Si l'effet du groupe expérimental résulte de la manipulation des dimensions évaluées par les indices de fidélité et de qualité, nous devrions alors observer une diminution importante ou totale des effets.

### Inter-corrélations des indices de fidélité et de qualité

Avant de présenter les résultats des analyses d'impact préliminaire et secondaire, un examen des corrélations observées entre ces indices est instructif (voir le tableau 5.5). Nous observons des corrélations positives entre tous les indices de qualité, ce qui est compatible avec l'idée qu'ils sont tous des aspects d'un même construit. Notons que plusieurs des corrélations sont faibles et qu'aucune d'entre elles n'approche un niveau d'unité ( $r > 0,70$ ), ce qui supporte notre stratégie initiale voulant que l'on considère ces indices séparément comme des prédicteurs du développement.

**Tableau 5.5 : Corrélations entre les indices de fidélité et de qualité**

	1	2	3	4	5	6
Fidélité						
1-Structurelle	1	,64*	,39	,64*	,23	,66*
2-Contenu	,55*	1	,16	,59*	,36	,44
Qualité						
3-Structurelle	,39	,16	1	,41	,51	,41
4-Éducative	,64*	,59*	,42	1	,61*	,21
5-Sensibilité des éducatrices	,26	,36	,51	,61*	1	,09
6-Lecture	,66*	,44*	,41	,21	,09	1

*Note* : Les garderies sont l'unité d'analyse (N =15). Les estimations de fidélité sont basées sur les données recueillies entre la seconde et la troisième évaluation. Les corrélations r de Pearson sont rapportées dans la diagonale inférieure, alors que les corrélations tau non paramétriques de Kendall sont rapportées dans la diagonale supérieure.

### Première série d'analyses – I : Fidélité structurelle et de contenu – comparaison entre les garderies programme et les garderies témoins

Les statistiques descriptives et les tests de signification pour la comparaison des indices de fidélité pour les groupes en garderie apparaissent au tableau 5.6. Tel qu'indiqué précédemment, les tests d'association paramétrique et non paramétrique sont rapportés.

L'examen du tableau 5.6 révèle la faible puissance statistique des analyses testant le changement, relativement à la période de base, pour les garderies programme (voir la portion inférieure du tableau). Nous soulignons que le changement en question reflète l'évolution depuis la mise en œuvre initiale à la mise en œuvre subséquente, et non la différence réelle entre la pré-évaluation et la post-évaluation. Les estimations quantitatives de ces changements revêtant un intérêt intrinsèque, nous rapportons donc les résultats de ces analyses malgré leur faible puissance.

Pour commencer, nous considérons les changements survenus dans les garderies programme d'une période d'évaluation à l'autre. Dans ce cas, nous observons, au premier post-test, que les garderies programme ont en moyenne augmenté leur fidélité de contenu de 0,05 et leur fidélité structurelle de 0,15 par rapport au niveau de base. Ces effets indiquent des gains de 5 et 15 %, respectivement, pour ce qui est du nombre total d'éléments du programme mis en place dans les garderies programme. Au second post-test, le gain estimé pour la fidélité de contenu a enregistré une baisse négligeable de 0,01, tandis que les gains pour la fidélité structurelle ont diminué de 0,08. En somme, nous pouvons observer que les garderies programme enregistrent à travers le temps des gains pour les indices de fidélité que ces gains diminuent suivant la première période post-test.

En examinant les contrastes comparant les deux types de garderie, nous remarquons que les tests paramétriques indiquent que les garderies programme différaient de façon importante des garderies témoins seulement quelques mois après le début du programme. La différence était de 0,12 pour la fidélité de contenu et de 0,39 pour la fidélité structurelle. Ce résultat démontre que, peu après qu'on ait commencé à mettre en œuvre le programme, les garderies programme avaient déjà mis en place beaucoup plus d'éléments centraux du programme testé que ce qui aurait été attendu selon les estimations provenant des garderies témoins.

La différence entre les groupes en garderie a fait l'objet d'une autre estimation effectuée à l'aide d'indices calculés à partir des observations faites pour les deux groupes durant la première période post-test. Pour les deux indices de fidélité, la première période post-test a démontré la plus grande différence entre les groupes. Pendant cet intervalle de temps, la différence entre les groupes a augmenté à 0,15 pour la fidélité de contenu et à 0,54 pour la fidélité structurelle. Selon les tests paramétriques, les deux différences étaient statistiquement significatives, à un niveau alpha de 0,10. En particulier, l'effet de la fidélité structurelle a été confirmé par la signification statistique du test non paramétrique. L'effet est donc le plus robuste pour la première période post-test.

Pour la deuxième période post-test, la différence moyenne entre les groupes en garderie reste inchangée à 0,15 pour l'indice de fidélité de contenu. Dans le même ordre d'idées, la différence moyenne entre les groupes pour l'indice de fidélité structurelle diminue légèrement à 0,53. Toutefois, comparativement à la première période post-test, le profil de signification statistique change. En particulier, le test fondé sur la fidélité de contenu indique que la différence n'est plus statistiquement significative. L'examen des erreurs-types donne à penser que ce changement est dû à une augmentation de la variabilité au niveau de cet indice, pour les groupes programme en garderie. La différence entre les groupes quant à la fidélité structurelle demeure statistiquement significative tant pour le test paramétrique que pour le test non paramétrique.

**Tableau 5.6 : Fidélité du programme, comparaison entre les groupes et à travers le temps**

	<i>Dimension de fidélité</i>			
	<i>Contenu</i>		<i>Structurelle</i>	
<i>Entre les groupes en garderie</i>	<b>Moyenne</b>	<b>ET</b>	<b>Moyenne</b>	<b>ET</b>
<b>Niveau de base<sup>c</sup></b>				
Garderies programme (N=4)	,67	,01	,78	,10
Garderies témoins (N=11)	,55	,14	,39	,15
Welch <sup>a</sup>	F (1, 10) = 7,94*		F (1, 8) = 31,63 <sup>†</sup>	
U de Mann-Whitney <sup>b</sup>	Z=1,31		Z=2,76	
<b>Premier post-test</b>				
Garderies programme (N=4)	,72	,10	,93	,05
Garderies témoins (N=11)	,57	,14	,39	,15
Welch <sup>a</sup>	F (1, 7) = 5,30 <sup>m</sup>		F (1, 12) = 102,10 <sup>†</sup>	
U de Mann-Whitney	Z = -1,63		Z = -2,89 <sup>†</sup>	
<b>Second post-test</b>				
Garderies programme (N=4)	,71	,13	,86	,14
Garderies témoins (N=11)	,56	,12	,33	,14
Welch <sup>a</sup>	F (1, 5) = 3,95		F (1, 5) = 40,24 <sup>†</sup>	
U de Mann-Whitney	Z = -1,64		Z = -2,89 <sup>†</sup>	
<b>Troisième post-test</b>	S/O	S/O	S/O	S/O
<b>Au sein du groupe programme</b>				
Groupe programme : Niveau de base v. 1 <sup>er</sup> post-test	t pairé (3) = ,93 Wilcoxon = 2,00, Z = ,45		t pairé (3) = 2,10 Wilcoxon = 9,00, Z = 1,46	
Groupe programme : Niveau de base v. 2 <sup>e</sup> post-test	t (3) = ,56 Wilcoxon = 7,00, Z = ,73		t (3) = 1,18 Wilcoxon = 2,00, Z = ,45	

**Note :** Les données des garderies témoins n'étaient disponibles que pour la première période post-test. En l'absence d'une intervention dans ces garderies, nous avons présumé une stabilité à travers le temps et imputé les valeurs pré-test en fonction de celles du premier post-test. Pour les deux groupes garderies, les données de fidélité étaient seulement disponibles jusqu'à la seconde période post-test.

<sup>a</sup> Le test de Welch est robuste au biais qui peut découler de la combinaison de tailles d'échantillons inégales et d'une variance hétérogène.

<sup>b</sup> Le test compare les rangs moyens des scores pour les deux groupes et il n'émet aucun postulat quant à la distribution de ces scores.

<sup>c</sup> Pour cette série de comparaisons seulement, les scores de la période de base ne constituent pas un véritable pré-test. Les éléments du programme testé étaient déjà en place à cette période. Même s'ils étaient en place, il y aurait eu peu de chance qu'un effet soit observé un effet chez les enfants à ce stade. La validité de l'évaluation pré-test n'est donc pas remise en cause.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , <sup>†</sup>  $p < 0,01$  (tous les tests sont bilatéraux).

En somme, ces analyses préliminaires des estimations de fidélité révèlent que 70 à 80 % des éléments centraux du programme étaient en place dans les garderies programme au niveau de base et que cette proportion a augmenté de 5 à 15 % au premier post-test, avec quelques pertes observées au deuxième post-test. La différence entre les garderies programme et les garderies témoins s'est accrue pour les deux mesures à la période de base et est demeurée stable jusqu'à la deuxième évaluation. Les conclusions que l'on peut tirer de cette observation sont limitées, car le nombre de garderies ayant reçu le programme testé était insuffisant pour effectuer un test équitable. Quoiqu'il en soit, les résultats indiquent que les garderies programme avaient mis en place beaucoup plus d'éléments centraux du programme testé que les garderies témoins.

### **Première séries d'analyses – II : Qualité – comparaison entre les garderies programme et les garderies témoins**

Alors que les indices de fidélité présentent une image grossière de la qualité des programmes en garderie, en termes de la mise en place ou non de certains éléments clés du programme testé, les estimations de la qualité permettent des distinctions plus fines entre les groupes en garderie en fonction de la façon dont les éléments ciblés du programme ont été mis en œuvre. Les estimations de qualité reflètent la qualité moyenne de la mise en œuvre des différents volets centraux du programme sur une échelle à sept points, où 1 indique un programme de garde d'enfants de qualité inadéquate et 7 indique un programme de garde d'enfants d'excellente qualité. Les quatre dimensions abordées sont: a) la qualité structurelle; b) la qualité éducative; et c) la sensibilité des éducatrices et d) la qualité de lecture.<sup>76</sup> Nous disposons d'une estimation de chaque dimension pour la première période post-intervention de 12 mois. Les statistiques descriptives et les tests de signification de la comparaison des groupes en garderie pour ces indices apparaissent au tableau 5.7.

L'examen des moyennes affichées par les groupes indique que le groupe programme a un avantage quant à toutes les dimensions de la qualité. Cette différence apparente de la qualité moyenne s'est vue confirmée par les tests paramétriques menés quant aux dimensions de qualité structurelle et éducative, à un niveau de signification statistique de 0,10. L'effet le plus important et le plus statistiquement significatif était celui associé à la qualité de lecture. Le test n'était pas assez sensible pour détecter les différences entre les groupes pour ce qui est de la sensibilité des éducatrices, dimensions qui a démontré beaucoup plus de variations intra-groupes que les autres (voir les estimations des écarts-types). Les résultats sont compatibles avec l'idée voulant qu'en moyenne, les enfants inscrits aux garderies offrant le nouveau programme préscolaire aient bénéficié d'un programme de qualité supérieure, plus particulièrement en ce qui concerne les activités de lecture, comparativement à ceux des garderies du groupe témoin.

---

<sup>76</sup>On trouvera une description plus détaillée des indices de qualité à la section 5.6.

**Tableau 5.7 : Comparaison de la qualité entre les garderies programme et les garderies témoins**

	<i>Dimensions de qualité</i>							
	Structurale		Éducative		Sensibilité de l'éducatrice		Lecture	
	Moyenne	ET	Moyenne	ET	Moyenne	ET	Moyenne	ET
Garderies programme (N=4)	6,42	,42	5,71	,82	5,88	2,25	5,88	1,44
Garderies témoins (N=11)	5,73	,81	4,38	1,42	4,55	2,45	1,27	,90
Welch <sup>a</sup>	F (1, 10) = 4,56 <sup>m</sup>		F (1, 9) = 5,03 <sup>m</sup>		F (1, 5) = ,975		F (1, 3) = 35,90	
U de Mann-Whitney	Z = 1,46		Z = 1,60		Z = 1,04		Z = -3,35	

<sup>a</sup> Le test de Welch est robuste au biais qui peut découler de la combinaison de tailles d'échantillons inégales et d'une variance hétérogène.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux).

### Deuxième série d'analyses – I : Impact de la fidélité du programme sur les échelles standardisées de l'ÉPE-AD

Comme dans les analyses par groupes expérimentaux, nous souhaitons évaluer l'impact de la fidélité du programme sur les trajectoires développementales de nos groupes en garderie. En termes techniques, cela consiste à estimer un effet DD pour les deux dimensions de la fidélité, soit la fidélité structurelle et la fidélité de contenu. Nous voulons aussi vérifier si les effets DD observés de fidélité sont redondants avec les effets associés aux groupes expérimentaux. Sur le plan technique, cela consiste à estimer de nouveau l'impact des groupes expérimentaux dans un modèle qui inclut simultanément les effets associés à la fidélité.

Les estimations des effets du programme préscolaire de services de garde testé, basées sur les indicateurs de fidélité, sont rapportées séparément pour chaque variable associée aux résultantes. Les analyses de l'échelle de Communication reposent sur les participants faisant partie de l'un des groupes expérimentaux en garderie. L'échantillon utilisé pour les analyses des autres échelles était légèrement plus petit, puisqu'il excluait les enfants qui n'avaient pas complété l'ÉPE-AD en français au moins deux fois durant la première année. Pour raisons de parcimonie, nous rapportons seulement les résultats obtenus en utilisant l'ensemble complet de covariables (modèle 5 tel que décrit à la section 5.1.3).

**Échelle de Communication (standardisée):** Les indices de fidélité ont été insérés comme prédicteurs des scores de Communication. Au contraire des analyses précédentes, les estimations des scores de Communication standardisés ne sont présentées qu'une seule fois, soit pour les analyses de l'échantillon total (voir le tableau 5.8). Comme nous pouvions nous y attendre étant donné que la différence entre les échantillons n'est que de six enfants, le profil de résultats est très similaire pour les deux séries d'analyses.

L'effet estimé des deux types de fidélité est négatif pour la période de base, mais n'est statistiquement significatif que pour la fidélité structurelle. Les estimations négatives de la fidélité pourraient s'expliquer par un biais de sélection où les enfants ayant de moins bonnes compétences de communication sont inscrits de façon disproportionnée dans les garderies caractérisées par une grande fidélité structurelle. De toute évidence, les covariables insérées dans les analyses n'ont pas suffi pour neutraliser ce biais de sélection dans le cas de la fidélité

structurelle. Quoiqu'il en soit, la direction de l'association s'inverse à la première période post-test, alors que les estimations DD démontrent que les gains des scores de Communication observés de la période de base au premier post-test sont positivement associés aux deux indices de fidélité.<sup>77</sup> Comme pour les analyses principales d'impact (rapportées à la section 5.1.2), les estimations DD pour les périodes de temps subséquentes ne sont pas statistiquement significatives.

L'ampleur de l'effet peut être mieux appréciée si nous résolvons l'équation de régression pour une différence donnée au niveau de la fidélité structurelle. Pour y arriver, il suffit de multiplier la différence désirée quant à la fidélité structurelle par l'estimation DD rapportée au sein du tableau. En effectuant ce calcul pour l'effet de la fidélité structurelle (premier post-test), nous obtenons un effet de 0,068 écart-type pour une différence de 0,10 au niveau la fidélité structurelle (c.-à-d.  $0,068 = 0,68 \times 0,10$ ). Dans le même ordre d'idées, une différence de 0,30 au niveau de l'indice de fidélité structurelle correspond à un impact estimé de traitement d'environ 0,20 écart-type, considéré comme un petit effet selon les points de références conventionnels. En faisant le même calcul à partir de l'estimation de fidélité de contenu, on constate que la taille des effets est jugée, en moyenne, beaucoup plus importante. Une différence de seulement 0,20 au niveau de l'indice de fidélité de contenu est associée à un gain estimé équivalent de 0,20 écart-type au niveau des scores de Communication.

**Tableau 5.8 : Estimation DD de l'effet de la fidélité sur les scores de Communication standardisés (temps x indices de fidélité) – Version française**

	<i>Dimensions de fidélité – Erreurs groupées par garderie</i>			
	Fidélité structurelle		Fidélité de contenu	
	Effet	ET robuste	Effet	ET robuste
Effet au niveau de base	-,482*	,184	-,367	,518
DD 1 <sup>er</sup> post-test	,680*	,243	1,04 <sup>m</sup>	,554
DD 2 <sup>e</sup> post-test	,368	,224	-,575	,598
DD 3 <sup>e</sup> post-test	,175	,208	-,400	,753
F de Wald	F (3, 12) = 2,28		F (3, 12) = 3,57 <sup>m</sup>	
Nombre de participants	154		154	

**Note :** La résultante a été standardisée, c'est-à-dire que les effets peuvent être interprétés en unités d'écart-type.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux, 14 degrés de liberté).

Après avoir établi que les principaux résultats d'impact étaient reproduits en employant les estimations de fidélité comme prédicteurs, nous nous sommes demandés si un contrôle statistique de la fidélité éliminerait l'effet statistiquement significatif observé lors de la comparaison des groupes en garderie pour la première période post-test (c.-à-d.  $DD_{1er\text{post-test}} = -0,314^*$ ,  $ET = 0,151$ ). Si le programme testé produit son effet de cette manière, nous nous attendons alors à ce que cela soit le cas. Pour ces analyses secondaires, nous avons employé un indice global de fidélité (fidélité structurelle et fidélité de contenu sont fusionnées en un seul indice). Comme prévu, l'estimation DD de l'effet de temps par groupe expérimental avait

<sup>77</sup> Dans le cas présent, nous interprétons l'effet de fidélité de contenu comme étant statistiquement significatif, car il est marginal à 0,10 pour les analyses effectuées avec l'échantillon total et significatif à 0,05 pour les analyses effectuées avec le sous-échantillon.

considérablement diminué et n'était pas statistiquement significatif lorsque nous avons contrôlé pour l'estimation globale de fidélité,  $DD_{ajustée1erpost-test} = -0,101$ ,  $ET = 0,233$ .

**Échelle de Conscience de soi (standardisée):** À l'instar des analyses précédentes, nous avons inséré les indices de fidélité comme prédicteurs des scores de Conscience de soi. Le patron de résultats observé reflète celui des analyses des scores de Communication. La seule exception notable est le fait que, pour ces analyses, les mesures de fidélité n'ont pas d'effet statistiquement significatif au niveau de base. Les effets de fidélité estimés issus de cette analyse sont présentés au tableau 5.9.

**Tableau 5.9 : Estimation DD de l'effet de la fidélité sur les scores de Conscience de soi standardisés (temps x indices de fidélité) – (participants ayant répondu au test en français seulement)**

	<i>Dimensions de fidélité – Erreurs groupées par garderie</i>			
	Fidélité structurelle		Fidélité de contenu	
	Effet	ET robuste	Effet	ET robuste
Effet au niveau de base	-,410 <sup>m</sup>	,202	-,292	,541
DD 1 <sup>er</sup> post-test	,674 <sup>†</sup>	,197	1,017*	,463
DD 2 <sup>e</sup> post-test	,423 <sup>m</sup>	,200	-,605	,605
DD 3 <sup>e</sup> post-test	,040	,225	-,383	,879
F de Wald	F (3, 12) = 5,91 <sup>†</sup>		F (3, 12) = 5,73*	
Nombre de participants	147		147	

*Note :* La résultante a été standardisée, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type.

<sup>m</sup> Indique la signification à  $p < 0,10$ , \*  $p < 0,05$ , <sup>†</sup>  $p < 0,01$  (tous tests sont bilatéraux, 14 degrés de liberté).

Étant donné que les effets DD importants qui ont été observés dans les analyses par groupe expérimental ont été reproduits ici, une extension logique de ces analyses consiste à vérifier si l'effet de fidélité est redondant à l'effet des groupes expérimentaux rapporté précédemment,  $DD_{1erPost-test} = -0,337^{\dagger}$ ,  $ET = 0,084$ . Un indice global de fidélité a été utilisé dans ces analyses secondaires. Les résultats indiquent qu'un contrôle statistique de la fidélité entraîne une réduction de la taille de l'effet des groupes expérimentaux qui devient alors non significatif,  $DD_{ajustée1erPost-test} = 0,169$ ,  $ET = 0,169$ . Ceci est le résultat qui serait attendu si l'appartenance aux groupes expérimentaux exerçait son effet par la fidélité de la mise en œuvre du programme.

**Échelle des Habiletés cognitives (standardisée).** Les indices de fidélité ont ensuite été entrés comme prédicteurs des scores d'Habiletés cognitives. Les effets estimés de la fidélité qui ont résulté de ces analyses sont rapportés dans le tableau 5.10. Les résultats reproduisent partiellement ceux qui ont été obtenus dans les analyses principales d'impact rapportées dans le tableau 5.2.



**Tableau 5.10 : Estimation DD de l'effet de la fidélité sur les scores d'Habilités cognitives standardisés (temps x indices de fidélité) – (participants ayant répondu au test en français seulement)**

	<i>Dimensions de fidélité – Erreurs groupées par garderie</i>			
	Fidélité structurelle		Fidélité de contenu	
	Effet	ET robuste	Effet	ET robuste
Effet au niveau de base	-,429	,251	,108	,521
DD 1 <sup>er</sup> post-test	,480*	,210	,064	,535
DD 2 <sup>e</sup> post-test	,264	,236	-,344	,797
DD 3 <sup>e</sup> post-test	,221	,230	-,447	,808
F de Wald	F (3, 12) = 1,99		F (3, 12) = ,184	
Nombre de participants	148		148	

*Note* : La résultante a été standardisée, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux, 14 degrés de liberté).

Pour ces analyses, l'effet est propre à la fidélité structurelle. Ainsi, nous avons vérifié si l'effet observé lors des analyses principales d'impact pour la première période post-test,  $DD_{1er\text{post-test}} = -0,279^{\dagger}$ ,  $ET = 0,293$ , demeurerait statistiquement significatif une fois que nous avons contrôlé pour la fidélité structurelle. L'estimation DD ajustée suggère un effet de suppression (la taille estimée de l'effet est plus grande), mais demeure statistiquement non significatif,  $DD_{ajustée\ 1er\text{post-test}} = -0,496$ ,  $ET = 0,293$ . Un effet de suppression survient lorsqu'une variable indépendante (dans ce cas, la fidélité structurelle) supprime la variance d'erreur d'une autre variable indépendante de façon à améliorer l'ajustement global du modèle de régression. Nous n'irons pas plus loin dans l'interprétation de cette question secondaire. Toutefois, nous reconnaissons que le test ne corrobore pas notre hypothèse de départ.

***Échelle d'Habilité physique (standardisée)***: Nous avons aussi inséré les indices de fidélité comme prédicteurs des scores d'Habilité physique. Les effets estimés de fidélité résultant de ces analyses sont présentés au tableau 5.11.

Comme dans les analyses précédentes, il n'y a aucun impact sur l'Habilité physique pour la première période post-test. Cette absence d'effet correspond aux attentes et aux principaux résultats provenant des analyses par groupe expérimental. Contrairement à nos attentes, nous observons un effet de la fidélité structurelle pour le troisième post-test. Cet effet pourrait être écarté comme une aberration, mais il ne faut pas ignorer que les estimations DD de fidélité structurelle augmentent systématiquement avec le passage du temps. Toutefois, la fidélité structurelle ne présente pas de corrélat évident avec le développement de l'Habilité physique.

**Tableau 5.11 : Estimation DD de l'effet de la fidélité sur les scores d'Habilité physique standardisés (temps x indices de fidélité) – (participants ayant répondu au test en français seulement)**

	<i>Dimensions de fidélité – Erreurs groupées par garderie</i>			
	Fidélité structurelle		Fidélité de contenu	
	Effet	ET robuste	Effet	ET robuste
Effet au niveau de base	-,213	,335	,315	,692
DD 1 <sup>er</sup> post-test	,074	,361	-,491	,623
DD 2 <sup>e</sup> post-test	,404	,325	-,374	,773
DD 3 <sup>e</sup> post-test	,640*	,245	,212	1,02
F de Wald	F (3, 12) = 1,98		F (3, 12) = 2,85	
Nombre de participants	148		148	

*Note* : La résultante a été standardisée, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux, 14 degrés de liberté).

**Échelles de Vocabulaire réceptif et de Vocabulaire expressif (standardisées)** : En dernier lieu, nous avons inséré les indices de fidélité comme prédicteurs des scores de Vocabulaire réceptif et de Vocabulaire expressif. Les effets estimés de fidélité résultant de ces analyses sont présentés aux tableaux 5.12 et 5.13, respectivement.

Les résultats ne permettent pas de conclure que la fidélité était associée aux scores de Vocabulaire réceptif. L'effet au niveau de base et les estimations DD ne sont pas statistiquement significatifs. Ce résultat est compatible avec la conclusion tirée des analyses principales d'impact selon laquelle on ne peut pas de distinguer le groupe programme en garderie du groupe témoin en garderie en utilisant cette variable.

**Tableau 5.12 : Estimation DD de l'effet de la fidélité sur les scores de Vocabulaire réceptif standardisés (temps x indices de fidélité) – (participants ayant répondu au test en français seulement)**

	<i>Dimensions de fidélité – Erreurs groupées par garderie</i>			
	Fidélité structurelle		Fidélité de contenu	
	Effet	ET robuste	Effet	ET robuste
Effet au niveau de base	,180	,431	-1,02	,681
DD 1 <sup>er</sup> post-test	-,384	,695	,315	,789
DD 2 <sup>e</sup> post-test	,349	,537	,659	,830
DD 3 <sup>e</sup> post-test	,061	,350	1,019	,668
F de Wald	F (3, 12) = 2,88 <sup>m</sup>		F (3, 12) = ,882	
Nombre de participants	148		148	

*Note* : La résultante a été standardisée, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux).

À l'opposé, les analyses des scores de Vocabulaire expressif indiquent un avantage important et statistiquement significatif associé à des niveaux croissants de fidélité pour la première estimation DD post-test. Une différence de 0,20 au niveau de la fidélité correspond

approximativement à une augmentation légère (0,20 écart-type) et modérée (0,40 écart-type) des gains développementaux chez les enfants dans les mesures de Vocabulaire expressif. Comme avec les analyses principales d'impact rapportées à la section 5.1.2, on observe des gains initiaux à la première période post-test, mais aucun gain à la deuxième et troisième évaluation post-test, pour l'une ou l'autre des estimations de fidélité.

**Tableau 5.13 : Estimation DD de l'effet de la fidélité sur les scores de Vocabulaire expressif standardisés (temps x indices de fidélité) – (participants ayant répondu au test en français seulement)**

	<i>Dimensions de fidélité – Erreurs groupées par garderie</i>			
	<b>Fidélité structurelle</b>		<b>Fidélité de contenu</b>	
	Effet	ET robuste	Effet	ET robuste
Effet au niveau de base	-,697*	,247	-,454	,712
DD 1 <sup>er</sup> post-test	1,17 <sup>†</sup>	,299	2,07*	,767
DD 2 <sup>e</sup> post-test	,454	,333	-,941	,763
DD 3 <sup>e</sup> post-test	,124	,247	-,582	,905
F de Wald	F (3, 12) = 9,06 <sup>†</sup>		F (3, 12) = 9,71 <sup>†</sup>	
Nombre de participants	147		147	

*Note* : La résultante a été standardisée, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , <sup>†</sup>  $p < 0,01$  (tous les tests sont bilatéraux).

Nous avons ensuite vérifié si le contrôle statistique des effets observés de la fidélité réduisait substantiellement ou éliminait l'effet DD rapporté dans les analyses précédentes par groupe expérimental,  $DD_{1erpost-test} = -0,543^{\dagger}$ ,  $ET = 0,144$ . Pour cette analyse secondaire, nous avons employé l'indice global de fidélité. Les résultats indiquaient que le contrôle de la fidélité entraîne une réduction de la taille de l'effet sur les mesures de Vocabulaire expressif, effet qui devient non significatif,  $DD_{ajustée_{1erpost-test}} = -0,060$ ,  $ET = 0,238$ .

### **Impact des indices de qualité sur les échelles standardisées de l'ÉPE-AD**

Dans cette section, nous répétons les analyses rapportées à la section précédente, cette fois en estimant l'impact des divers indices de qualité. Les estimations dérivées des modèles de régression pour chaque variable de résultantes sont présentées à tour de rôle. Nous avons accordé une attention particulière à la qualité de lecture en tant que prédicteur et celle-ci sera donc abordée en dernier.

**Échelle de Communication (standardisée)**: L'impact estimé des trois indices de qualité, pour l'échantillon total de participants inscrits dans une garderie francophone au cours d'une période de temps donnée, apparaît au tableau 5.14.<sup>78</sup> L'ampleur de l'impact est invariablement la plus grande pour la première période post-test et elle est statistiquement significative pour les indices de qualité éducative et de sensibilité de l'éducatrice. Ce résultat est compatible avec notre attente selon laquelle ces indices seraient ceux liés de façon la plus constante au développement des compétences de Communication. Dans les deux cas, une différence de deux unités quant au

<sup>78</sup> Comme pour les analyses menées quant à la fidélité, nous ne rapportons pas les analyses reposant sur le sous-échantillon qui a complété les autres échelles de l'ÉPE-AD en français, car les deux échantillons ne diffèrent ici que d'un seul cas.

score de qualité est requise pour produire un effet de taille minimale (c.-à-d. un effet de 0,20 écart-type). Les résultats de ces analyses secondaires suggèrent que le contrôle de la qualité globale neutralise l'effet DD observé pour la première période post-test dans le cadre des analyses principales d'impact comparant les groupes expérimentaux, DD ajustée<sub>1erpost-test</sub> = -0,166, ET = 0,097.

**Tableau 5.14 : Estimation DD de l'effet de la qualité sur les scores de Communication standardisés (temps x indices de qualité) – (version française)**

	<i>Dimensions de qualité – Erreurs groupées par garderie</i>					
	Structurelle		Éducative		Sensibilité de l'éducatrice	
	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
Effet au niveau de base	-,032	,102	-,075	,052	-,026	,028
DD 1 <sup>er</sup> post-test	,218 <sup>m</sup>	,115	,148 <sup>†</sup>	,048	,098 <sup>†</sup>	,024
DD 2 <sup>e</sup> post-test	-,025	,114	,014	,073	,004	,034
DD 3 <sup>e</sup> post-test	-,141 <sup>m</sup>	,056	,038	,081	,011	,035
F de Wald	F (3, 12) = 6,18 <sup>†</sup>		F (3, 12) = 4,12 <sup>*</sup>		F (3, 12) = 10,62 <sup>†</sup>	
Nombre de participants	154		154		154	

*Note* : La résultante a été standardisée, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type. Modèle estimé avec la série complète de covariables, mais excluant l'appartenance au groupe et l'interaction de groupe par temps.

<sup>m</sup> dénote une signification à  $p < 0,10$ , <sup>\*</sup>  $p < 0,05$ , <sup>†</sup>  $p < 0,01$  (tous les tests sont bilatéraux, 14 degrés de liberté).

**Échelle de Conscience de soi (standardisée)** : Nous avons inséré les indices de qualité comme prédicteurs du développement des scores de Conscience de soi. Les estimations résultant de ces analyses sont rapportées au tableau 5.15. Encore une fois, nous avons observé un impact statistiquement significatif de tous les indices de qualité pour la première période post-test. Les résultats des analyses secondaires suggèrent qu'un contrôle de la qualité globale du programme occasionne une diminution de 50% de l'ampleur de l'effet estimé du traitement, effet qui demeure toutefois statistiquement significatif à un niveau alpha de 0,05, DD ajustée<sub>1erpost-test</sub> = -0,271, ET = 0,085. En d'autres termes, approximativement la moitié de l'effet du traitement est attribuable à la qualité globale du programme, mais cet effet ne peut être réduit à cette seule variable.

**Tableau 5.15 : Estimation DD de l'effet de la qualité sur les scores de Conscience de soi standardisés (temps x indices de qualité) – (participants ayant répondu au test en français seulement)**

	<i>Dimensions de qualité</i>					
	Structurelle		Éducative		Sensibilité de l'éducatrice	
	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
Effet au niveau de base	,057	,168	-,082	,061	-,010	,032
DD 1 <sup>er</sup> post-test	,192*	,072	,110 <sup>†</sup>	,036	,061*	,028
DD 2 <sup>e</sup> post-test	-,013	,090	-,015	,061	,001	,027
DD 3 <sup>e</sup> post-test	-,150 <sup>m</sup>	,076	,015	,088	,015	,033
F de Wald	F (3, 12) = 9,09 <sup>†</sup>		F (3, 12) = 4,20*		F (3, 12) = 1,71	
Nombre de participants	147		147		147	

*Note* : La résultante a été standardisée, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , <sup>†</sup>  $p < 0,01$  (tous les tests sont bilatéraux).

**Échelle d'Habilités Cognitives (standardisée)** : Comme pour les analyses précédentes, les indices de qualité ont été insérés comme prédicteurs du développement des scores d'Habilités cognitives. Les estimations provenant de ces analyses apparaissent au tableau 5.16. Nous n'avons observé aucune relation statistiquement significative. Les effets observés en comparant les groupes expérimentaux ne peuvent donc pas être expliqués par les mesures de qualité. Nous n'avons pas tenté d'estimer de nouveau l'effet du traitement en contrôlant pour la qualité globale, car la qualité n'était pas un prédicteur significatif des Habiletés cognitives.

**Tableau 5.16 : Estimation DD de l'effet de la qualité sur les scores d'Habilités cognitives standardisés (temps x indices de qualité) – (participants ayant répondu au test en français seulement)**

	<i>Dimensions de qualité – Erreurs groupées par garderie</i>					
	Structurelle		Éducative		Sensibilité de l'éducatrice	
	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
Effet au niveau de base	,044	,151	-,034	,056	,019	,021
DD 1 <sup>er</sup> post-test	-,058	,080	,003	,042	-,037	,028
DD 2 <sup>e</sup> post-test	-,119	,107	,042	,076	-,023	,034
DD 3 <sup>e</sup> post-test	-,125	,092	-,017	,049	-,041	,031
F de Wald	F (3, 12) = ,582		F (3, 12) = ,913		F (3, 12) = ,778	
Nombre de participants	148		148		148	

*Note* : La résultante a été standardisée, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , <sup>†</sup>  $p < 0,01$  (tous les tests sont bilatéraux).

**Échelle d'Habilité physique (standardisée)** : Comme pour les analyses précédentes, les indices de qualité ont été insérés comme prédicteurs du développement des scores d'Habilité physique. Les estimations provenant de ces analyses apparaissent au tableau 5.17. À l'instar de la plupart des analyses rapportées précédemment pour cette résultante, nous n'avons observé aucune relation.

**Tableau 5.17 : Estimation DD de l'effet de la qualité sur les scores d'Habilité physique standardisés (temps x indices de qualité) – (participants ayant répondu au test en français seulement)**

	<i>Dimensions de qualité – Erreurs groupées par garderie</i>					
	Structurelle		Éducative		Sensibilité de l'éducatrice	
	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
Effet au niveau de base	,160	,161	,031	,054	,002	,042
DD 1 <sup>er</sup> post-test	-,132	,118	,018	,064	,014	,035
DD 2 <sup>e</sup> post-test	,056	,126	,026	,078	,003	,039
DD 3 <sup>e</sup> post-test	,000	,130	,063	,051	,010	,046
F de Wald	F (3, 12) = 1,24		F (3, 12) = ,523		F (3, 12) = ,145	
Nombre de participants	148		148		148	

*Note* : La résultante a été standardisée, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux).

**Échelles de Vocabulaire réceptif et de Vocabulaire expressif (standardisées)**: Enfin, nous avons inséré les indices de qualité comme prédicteurs du développement des scores standardisés de Vocabulaire réceptif et expressif. Les impacts estimés résultant de ces analyses sont rapportés aux tableaux 5.18 et 5.19, respectivement. Les résultats indiquent que la qualité structurelle et la qualité éducative sont liées au Vocabulaire expressif. Il est intéressant de noter que la sensibilité de l'éducatrice s'est avérée être un prédicteur significatif des scores de Communication, mais non des scores de Vocabulaire expressif. Inversement, la qualité structurelle ne permettait pas de prédire les scores de Communication, mais elle était fortement liée aux scores de Vocabulaire expressif dans la présente analyse. Aucun des indices de qualité n'était associé de façon significative au Vocabulaire réceptif.

**Tableau 5.18 : Estimation DD de l'effet de la qualité sur les scores de Vocabulaire réceptif standardisés (temps x indices de qualité) – (participants ayant répondu au test en français seulement)**

	<i>Dimensions de qualité – Erreurs groupées par garderie</i>					
	Structurelle		Éducative		Sensibilité de l'éducatrice	
	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
Effet au niveau de base	,016	,109	-,136	,087	-,056	,037
DD 1 <sup>er</sup> post-test	,060	,099	,089	,099	,121 <sup>m</sup>	,065
DD 2 <sup>e</sup> post-test	,266 <sup>m</sup>	,142	,145	,109	,097 <sup>m</sup>	,046
DD 3 <sup>e</sup> post-test	,054	,080	,097	,087	,055	,032
Wald	F (3, 12) = 1,43		F (3, 12) = ,600		F (3, 12) = 1,52	
Nombre de participants	148		148		148	

*Note* : La résultante a été standardisée, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux).

**Tableau 5.19 : Estimation DD de l'effet de la qualité sur les scores de Vocabulaire expressif standardisés (temps x indices de qualité) – (participants ayant répondu au test en français seulement)**

	<i>Dimensions de qualité – Erreurs groupées par garderie</i>					
	Structurelle		Éducative		Sensibilité de l'éducatrice	
	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
Effet au niveau de base	-,173	,169	-,095	,055	-,006	,034
DD 1 <sup>er</sup> post-test	,419 <sup>†</sup>	,119	,203 <sup>†</sup>	,052	,094	,058
DD 2 <sup>e</sup> post-test	,033	,140	,016	,083	,017	,037
DD 3 <sup>e</sup> post-test	-,033	,078	,057	,088	,040	,032
F de Wald	F (3, 12) = 4,83*		F (3, 12) = 4,41*		F (3, 12) = ,955	
Nombre de participants	147		147		147	

*Note* : La résultante a été standardisée, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux).

Un deuxième ensemble d'analyses a été effectué pour vérifier si le contrôle statistique de la qualité globale du programme pouvait réussir à neutraliser l'effet DD sur le Vocabulaire expressif observé dans les analyses principales d'impact menées en fonction du groupe expérimental. Les résultats ont révélé un effet du traitement plus faible, mais qui est demeuré statistiquement significatif à un niveau de 0,05, DD ajusté<sub>1erPost-test</sub> = -0,437, ET = 0,177. Néanmoins, cet effet présente une réduction d'ampleur de 36 % comparativement à l'effet original. Comme pour l'échelle de Conscience de soi, nous concluons que la qualité rend compte d'environ un tiers de l'effet observé du traitement, mais ne supprime pas cet effet complètement.

***Communication, Cognition, et Vocabulaire expressif (standardisé) en fonction de la qualité de lecture.*** Une place particulière a été accordée à la qualité de lecture au sein des analyses en raison de son importance théorique et pratique pour le programme testé (voir tableau 5.7 et le *Rapport de mise en œuvre du projet*, 2014). Nous rapportons dans le tableau 5.20 les estimations de son effet sur les résultantes de l'ÉPE-AD qu'elle était la plus susceptible d'influencer, soit: les échelles de Communication, d'Habilités cognitives et de Vocabulaire expressif.

On n'a pas constaté que la qualité de lecture influençait les scores de Communication des enfants fréquentant une garderie. Toutefois, la qualité de lecture a influencé de façon significative les Habiletés cognitives et le Vocabulaire expressif de ces enfants. Comme dans les analyses précédentes, l'effet est propre au premier post-test. Les deux résultantes touchées se composent d'items qui exploitent les habiletés précoces de littératie et, bien sûr, le Vocabulaire expressif.

**Table 5.20 : Scores de Communication, d'Habilités cognitives et de Vocabulaire expressif en fonction de la qualité de lecture (temps x qualité) – (participants ayant répondu au test en français seulement)**

	<i>Résultantes de l'ÉPE-AD</i>					
	Communication		Habilités cognitives		Vocabulaire expressif	
	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
Effet au niveau de base	-,028	,020	-,034	,029	-,062	,029
DD 1 <sup>er</sup> post-test	,041	,030	,057*	,026	,113 <sup>†</sup>	,026
DD 2 <sup>e</sup> post-test	,006	,035	,015	,039	,011	,040
DD 3 <sup>e</sup> post-test	,000	,031	,023	,040	,005	,031
F de Wald	F (3, 12) = ,55		F (3, 12) = 4,21*		F (3, 12) = 7,29 <sup>†</sup>	
N de participants	146		146		146	

*Note* : La résultante a été standardisée, ce qui signifie que les effets peuvent être interprétés en tant qu'unités d'écart-type.

<sup>m</sup> Indique la signification à  $p < 0,10$ , \*  $p < 0,05$ , <sup>†</sup>  $p < 0,01$  (tous tests sont bilatéraux).

Comme dans les analyses précédentes, nous avons réévalué les différences entre les groupes en garderie en utilisant la qualité comme covariable. Les effets ajustés du programme pour le premier post-test n'ont pas atteint la signification statistique, tant pour l'échelle des Habiletés cognitives (DD ajusté<sub>1erPost-test</sub> = -0,176, ET = 0,160) que pour la sous-échelle du Vocabulaire expressif (DD ajusté<sub>1erPost-test</sub> = -0,397, ET = 0,432). L'effet du programme testé sur les résultantes en question est par conséquent modulé par la qualité des activités de littératie qu'il offre.

## Sommaire

Plusieurs mesures de fidélité et de qualité du programme ont été prises et on a constaté qu'elles étaient faiblement à modérément corrélées entre elles. Une première série d'analyses a indiqué que les garderies programme se différencient des garderies témoins pour la plupart des dimensions mesurées. La qualité supérieure des activités de lecture offertes dans les garderies programme était peut-être l'aspect le plus notable des résultats. Pris dans leur ensemble, les résultats indiquent que si les enfants fréquentant les deux types de garderie connaissent des résultantes statistiquement différentes, alors il est possible que ces différences soient attribuables à la mise en œuvre du programme testé plutôt qu'à un autre facteur. Cette supposition a été confirmée dans une deuxième série d'analyses où (a) les résultats principaux issus des comparaisons entre les groupes expérimentaux ont été reproduits, puis (b) ultérieurement éliminés ou substantiellement réduits à l'aide d'un contrôle statistique pour la fidélité et pour la qualité du programme. Dans les cas où une telle confirmation a été obtenue, nous pouvons être confiants que les résultats observés sont attribuables aux dimensions mesurées du programme testé et non à un autre facteur. De façon intéressante, certaines dimensions de la qualité et de la fidélité ont été démontrées comme étant plus prédictives de certaines résultantes chez les enfants que d'autres. Cette information peut servir à éclairer les changements à apporter au programme ainsi que les secteurs sur lesquels insister lorsque nous envisageons l'endroit où concentrer l'investissement des ressources.



### **Analyses détaillées: Caractéristiques linguistiques de l'échantillon**

Comme pour toute intervention, il est intéressant de se demander si l'efficacité du traitement dépend de certaines caractéristiques de l'enfant. Un postulat très pertinent au projet Capacité d'apprentissage veut que l'exposition accrue à un programme en garderie de grande qualité administré en français ait un impact positif sur les résultantes développementales chez les enfants, notamment en ce qui concerne les compétences langagières. Il est raisonnable de croire que le programme testé aura son plus grand effet auprès des enfants pour lesquels le programme aura entraîné, par rapport à la période de base, le changement le plus important quant à l'exposition au français/la qualité de l'exposition au français. Nous présumons que les enfants présentant un haut degré d'exposition à l'anglais au niveau de base correspondent à cette description. En effet, nous nous attendons à ce que les enfants les moins exposés au français au niveau de base répondent le plus fortement au programme testé (d'autres évaluations de programme ont permis d'obtenir des résultats similaires quant aux résultantes associées au langage, Maltais, 2007).

Les caractéristiques linguistiques de l'enfant peuvent être captées sous divers aspects au moyen d'indices relatifs aux caractéristiques linguistiques des familles. Cette liste d'indices comprend le continuum de français parlé par l'enfant avec les gens de son environnement immédiat, la langue des activités de littératie, la composition linguistique du foyer (endogame francophone v. autres), le continuum de français parlé par la mère à son enfant et le continuum de français parlé par le père à son enfant.<sup>79</sup> Nous rapportons au tableau 5.21 les corrélations invariablement élevées observées entre toutes les variables linguistiques. La composition linguistique du foyer (appelé « type de foyer ») est une variable qui fait une distinction entre les foyers homogènes où les deux parents sont francophones (c.-à-d. endogame francophone) et les familles où les parents sont plus diversifiés du point de vue linguistique (c.-à-d. autres). Deux variables « types de foyer » ont été prises en considération dans les analyses qui suivent. La première était le « type de foyer » fondé sur la PLOP (première langue officielle parlée), laquelle se définit par les langues officielles connues des parents, par leur langue maternelle et par la langue utilisée à la maison. La deuxième est une version alternative du « type de foyer » fondée exclusivement sur la langue que les parents déclarent utiliser avec l'enfant (relatif à l'enfant). L'avantage de cette seconde définition tient au fait que l'accent porte exclusivement sur les conversations auxquelles participe l'enfant. Ces variables ont été réduites à des dichotomies (p. ex., endogame francophone v. autres francophones) afin de simplifier les analyses et de tenir compte de la petite taille d'échantillon pour certaines catégories (p. ex. familles endogames anglophones). Pour les deux variables binaires, les foyers endogames francophones ont reçu le code 1 tandis que les autres types de familles ont reçu le code 0.

Même si les inter-corrélations observées étaient fortes, les variables linguistiques mesurées peuvent être classées en fonction d'une importante dimension: la participation active de l'enfant. En tant que construit, l'exposition au français peut être perçue comme un continuum allant de complètement passif (p. ex., une conversation entendue par l'enfant, mais à laquelle il ne participe pas) à complètement actif (p. ex., une conversation qui implique l'enfant en tant que participant actif et où le français est la langue à la fois comprise et produite). Nous prévoyons

---

<sup>79</sup> Une définition alternative serait : « le continuum de la langue parlée à l'enfant par le parent francophone ». Les deux définitions de cette variable sont corrélées au-delà de 0,90. En fait, un changement de définition n'influerait pas sur les données de sept familles du présent échantillon. Nous avons opté pour la définition la plus simple et pour celle qui correspond le plus à la littérature indiquant que les caractéristiques de la mère sont les meilleurs prédicteurs des résultantes chez les enfants.

que les variables linguistiques qui reflètent le plus la participation active de l'enfant interagiront le plus fortement avec l'effet du programme. Cette prédiction a été fondée sur le fait que l'exposition active (plutôt que passive) au matériel linguistique est reconnue comme étant la méthode la plus efficace pour transférer de telles connaissances et compétences (Whitehurst et al., 1988; Hargrave & Sénéchal, 2000). Les enfants qui emploient le français de façon plus constante invitent les autres à les exposer davantage à la langue française, phénomène qui a été décrit en termes de réaction en chaîne (Pearson, 2007). Cette conceptualisation de la participation « active » suggère que le meilleur candidat en tant que modérateur de l'effet du programme est le continuum linguistique parlé par l'enfant.

La distinction entre l'exposition passive et l'exposition active est pertinente pour évaluer les répercussions du programme testé en raison du fait que la participation active des enfants constitue un volet important à la fois du programme préscolaire de service de garde testé et des Ateliers d'alphabétisation familiale. En fait, selon le *Rapport de mise en œuvre du projet* (Bérubé et al., 2014), c'est pour la qualité des activités de lecture en classe (qualité de lecture) que les plus grandes différences ont été observées entre les garderies programme et celles du groupe témoin. Les scores les plus élevés sont obtenus pour cette dimension lorsque les enfants s'engagent à participer activement en français lors des activités en classe. Si la participation active en français est un élément clé du programme testé, alors les enfants qui étaient les plus faibles dans ce domaine avant de recevoir l'intervention devraient connaître le plus grand effet. La force de cette relation devrait être plus élevée pour les variables qui reflètent le mieux la composante « active » de l'exposition et plus faible pour les variables linguistiques qui sont fondées sur un critère « passif », tel que la variable « type de foyer » basée sur la PLOP.

Enfin, nous prévoyons que la langue des activités de littératie à la maison sera particulièrement utile pour distinguer le groupe programme des groupes témoins. Cette prédiction est justifiée par le fait que le volet d'alphabétisation familiale de l'intervention a ciblé directement cette dimension et que les activités de littératie contribuent de façon importante à la préparation à la scolarisation (Neuman & Celano, 2001).<sup>80</sup> Cette mesure est aussi la seule qui soit liée à des comportements précis et concrets, propriété de mesure qui est reconnue pour provoquer des réponses plus précises et justes (pour une discussion, voir Shrigley, 1990). Enfin, dans d'autres interventions, on a constaté qu'elle était un modérateur de l'efficacité de programmes similaires (Maltais, 2007).

---

<sup>80</sup> Les analyses décrites dans la section 5.2 ne révèlent aucun effet général du programme en ce qui concerne la résultante liée à la langue des activités de littératie à la maison. Il est néanmoins possible que l'instrument utilisé n'ait pas été suffisamment sensible pour détecter un effet. Dans ce cas, des effets indirects sur les résultantes chez les enfants sont encore possibles.

**Table 5.21 : Corrélations d'ordre zéro entre les indicateurs de profils linguistiques**

Nom de la variable	1	2	3	4	5	6
1 – Continuum linguistique (enfant)	1	,63	,56	,71	,66	,69
2 – Langue des activités de littératie	,86	1	,49	,65	,59	,64
3 – Type de foyer (PLOP)	,63	,57	1	,72	,62	,59
4 – Type de foyer (Relatif à l'enfant)	,80	,74	,72	1	,82	,80
5 – Continuum linguistique (mère à l'enfant)	,78	,72	,63	,81	1	,64
6 – Continuum linguistique (père à l'enfant)	,83	,82	,59	,76	,61	1

*Note* : Fondé sur le sondage de bas, N = 232 à 243. Les corrélations r de Pearson sont rapportées dans la diagonale inférieure. Les corrélations de rang tau de Kendall sont rapportées dans la diagonale supérieure. Toutes les corrélations sont statistiquement significatives au niveau alpha 0,05. PLOP = Première langue officielle parlée.

### **Impact conditionnel (à l'exposition linguistique) du programme sur les échelles standardisées de l'ÉPE-AD**

Nous prévoyons que l'impact du programme sera le plus grand pour les enfants dont l'exposition à la langue française était la plus faible. Le test le plus fort de cette idée est fourni par une variable résultante qui représente l'éventail le plus large de compétences en langue française. Pour le projet Capacité d'apprentissage, le candidat idéal est l'échelle de Communication parce qu'elle comprend des observations valides pour l'échantillon au complet, y compris les enfants présentant un faible niveau d'habileté en français (c.-à-d. tous les enfants ont été évalués pour ce domaine, peu importe leurs antécédents linguistiques ou leur niveau de compétence en français). Les autres résultantes de l'ÉPE-AD ne démontreront peut-être pas aussi clairement l'importance de l'exposition à la langue au niveau de base. Ce manque de puissance attendu est une conséquence de l'exclusion des enfants les moins compétents en français, en raison du fait qu'ils ont complété l'ÉPE-AD en anglais. Ces enfants sont plus susceptibles d'être représentés par l'extrémité inférieure des indicateurs d'exposition au français, rapportés dans le tableau 5.21.

Nous avons examiné cette hypothèse en calculant des interactions triples en fonction des groupes expérimentaux, du temps et des caractéristiques linguistiques examinées, et en ajoutant ces interactions aux spécifications du modèle.<sup>81</sup> Les estimations découlant de ces analyses de régression ont été utilisées comme point de départ pour générer des estimations conditionnelles de l'effet du programme pour les enfants à « haute exposition au français » et pour ceux à « faible exposition au français »; elles définissent une équation qui peut être résolue pour divers

<sup>81</sup> Pour tester adéquatement l'interaction triple, nous devons aussi ajouter au modèle des effets d'interaction dits de «second-ordre». Par conséquent, toutes les interactions bilatérales possibles, incluant les caractéristiques linguistiques examinées, les groupes expérimentaux et le temps, ont été incluses dans les spécifications du modèle de régression.<sup>82</sup> Il est à noter que dans le tableau 5.23, certaines des estimations conditionnelles indiquent un effet significativement négatif du programme pour les enfants à faible exposition au français (c.-à-d. que les coefficients DD sont positifs et significatifs). Ces estimations sont spécifiquement celles associées aux variables de type de ménage. Ces estimations ne doivent pas être interprétées avec trop de force étant donné qu'elles ne sont pas corroborées par les autres modérateurs (continus).

niveaux d'exposition linguistique. Les estimations conditionnelles de l'effet du programme rapportées dans le présent rapport ont été calculées de cette manière.

Il est important de noter que le rôle modérateur de chaque caractéristique linguistique a été vérifié dans des modèles de régression distincts. Autrement dit, le patron des résultats est libre de varier pour chaque caractéristique linguistique, mais il nous faut garder à l'esprit que les variables sont hautement corrélées. Des résultats constants pour tous les indicateurs suggéreraient qu'ils sont tous associés à un construit commun (p. ex., l'exposition au français) et que leur variance commune est à la base des résultats obtenus. À l'inverse, des résultats qui se distinguent de manière marquée selon la variable linguistique considérée suggéreraient que les variables font chacune référence un aspect unique du construit (p. ex., exposition linguistique « active » v. « passive »). Nous avons déjà mentionné que les résultats obtenus pourraient varier si l'exposition captée par une variable donnée est du type passif ou actif.

Dans ce qui suit, nous rapportons les estimations de l'effet du programme conditionnel à l'exposition au français pour chaque résultante de l'ÉPE-AD. Les estimations DD de l'effet du programme sont interprétées de la même manière qu'elles l'ont été dans les analyses précédentes. La seule différence tient au fait que, dans le présent cas, les effets sont conditionnels à une caractéristique particulière de l'individu. Ces effets conditionnels sont les plus faciles à interpréter lorsque le test de l'interaction trilatérale entre le temps, le groupe à l'étude et la caractéristique linguistique est significatif du point de vue statistique. Une telle interaction significative peut être interprétée comme un test de la question de savoir si l'ampleur de l'effet du programme varie de manière importante en fonction de la caractéristique linguistique.

Deux types de tests d'interaction triple sont rapportés. Le premier type évalue si l'ampleur des effets du programme (c.-à-d., les estimations DD) varie selon les différents niveaux d'exposition linguistique, pour toutes les périodes post-test prises ensemble. Cette hypothèse générale est vérifiée à l'aide du test F de Wald d'interaction globale, lequel est rapporté à l'avant-dernière rangée des tableaux. Le deuxième type d'interaction triple évalue si les estimations DD varient en fonction de l'exposition linguistique. Les résultats de ce deuxième type de test, qui se résume à un test d'interaction simple ou à un test DDD (comme exemple, voir les analyses par dosage rapportées dans le tableau 5.4), sont rapportés dans les tableaux et le symbole  $\Delta$  indique lorsque le test est statistiquement significatif ( $p < 0,10$ ).

***Échelle de Communication (standardisée)***. L'hypothèse selon laquelle l'ampleur de l'effet du programme varie de façon importante en fonction des caractéristiques linguistiques de l'échantillon a été vérifiée formellement au moyen d'une interaction triple. Selon le test F de Wald, l'interaction était statistiquement significative à  $p = 0,10$  pour cinq des six caractéristiques linguistiques considérées ici. Les estimations conditionnelles de l'effet du programme sont rapportées dans le tableau 5.22. Les tests d'interaction « simple » significatifs sont indiqués par le symbole  $\Delta$  dans le tableau.

Les résultats qui ont été obtenus pour chaque modérateur convergent pour former une image relativement claire de la relation existant entre les caractéristiques linguistiques et l'ampleur de l'effet du programme. Dans l'ensemble, les estimations de l'effet du programme étaient bien plus faibles pour les enfants hautement exposés au français que pour ceux qui y étaient moins exposés. Ce résultat correspond à nos attentes. Aucun des tests associés aux enfants à forte exposition n'était statistiquement significatif. Nous allons maintenant regarder de plus près les effets estimés du traitement pour la condition de faible exposition.

Pour les enfants à faible exposition, nous observons que l'effet du programme basé sur la comparaison des groupes en garderies est semblable à celui qui a été rapporté à la section 5.1.2. Pour tous les modérateurs, l'estimation DD pour le premier post-test est significative du point de vue statistique. Un effet du programme émerge au troisième post-test lorsque le continuum linguistique de l'enfant et la langue des activités de littératie sont considérés. Ce résultat est intéressant étant donné que ces deux comportements sont directement ciblés par le programme testé et impliquent la participation active de l'enfant.

Il est possible que la question la plus cruciale est celle répondue par les tests DDD («simple» test d'interaction triple qui, lorsque statistiquement significatif, est indiqué dans le tableau 5.22 par le symbole <sup>Δ</sup>): Est-ce que l'ampleur des estimations DD varie de manière importante en fonction des caractéristiques linguistiques de l'enfant? Le patron exact a varié selon les analyses, mais en général, les effets significatifs du programme relatifs au groupe témoin hors garderie pour les deuxième et troisième périodes post-test étaient significativement plus grands que les effets non significatifs pour la condition de «forte exposition au français». Un seul effet issu de la comparaison du groupe programme en garderie et du groupe témoin en garderie a interagi avec le profil linguistique: l'effet DD du troisième post-test mettant en cause le continuum linguistique de l'enfant. Tous les autres effets rapportés dans le tableau 5.22 peuvent être interprétés comme preuves positives seulement parce qu'ils vont dans la direction attendue.

Contrairement aux résultats rapportés dans la section 5.1.2, nous observons dans ces analyses des effets du programme sur l'échelle de Communication relativement au groupe témoin hors garderie. La taille des effets est plus grande et plus constante dans le temps dans ces analyses que dans les précédentes. Pour tous les indicateurs, les effets du programme sont significatifs pour le premier et pour le troisième post-test. De plus, on observe, pour la première fois dans ces analyses, des effets significatifs du programme pour le deuxième post-test lorsque les modérateurs suivants sont pris en compte: continuum linguistique de l'enfant, type de foyer (relatif à l'enfant), et continuum linguistique parlé par la mère à l'enfant. Pris dans leur ensemble, les résultats tendent à confirmer l'attente selon laquelle le programme produit une plus grande différence auprès des enfants les moins exposés au français au niveau de base. Ils confirment également que ce phénomène est le mieux capté par les variables linguistiques qui reflètent une exposition « active » plutôt que « passive ».

**Tableau 5.22 : Effets estimés du programme sur les scores de Communication (version française) conditionnels au profil linguistique (échantillon total)**

	<i>Modérateurs de l'effet du programme</i>												
	Continuum linguistique (enfant)		Langue des activités de littératie		Type de foyer (PLOP)		Type de foyer (relatif à l'enfant)		Continuum linguistique (mère-enfant)		Continuum linguistique (père-enfant)		
	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET	
<b>Forte exposition au français</b>													
<b>G1 v. G2</b>													
Niveau de base	,131	,129	,150	,169	,162	,162	,087	,149	,159	,153	,249	,114	
1 <sup>er</sup> post-test DD	-,292	,223	-,216	,190	-,313	,230	-,336	,215	-,299	,188	-,315	,190	
2 <sup>e</sup> post-test DD	-,189	,163	-,135	,177	-,190	,177	-,192	,149	-,143	,208	-,158	,117	
3 <sup>e</sup> post-test DD	-,017 <sup>Δ</sup>	,182	,012	,222	-,090	,199	-,056	,176	-,081	,181	-,030	,157	
<b>G1 v. G3</b>													
Niveau de base	-,077	,173	-,0311	,089	-,100	,099	-,124	,106	-,088	,090	-,019	,060	
1 <sup>er</sup> post-test DD	-,131	,195	-,101 <sup>Δ</sup>	,148	-,126	,210	-,127 <sup>Δ</sup>	,173	-,136 <sup>Δ</sup>	,164	-,209	,164	
2 <sup>e</sup> post-test DD	,058	,144	,062 <sup>Δ</sup>	,154	,039	,163	,106 <sup>Δ</sup>	,126	,119 <sup>Δ</sup>	,197	-,039	,144	
3 <sup>e</sup> post-test DD	,088 <sup>Δ</sup>	,150	,103 <sup>Δ</sup>	,197	,115 <sup>Δ</sup>	,165	,123 <sup>Δ</sup>	,131	,085 <sup>Δ</sup>	,174	-,012 <sup>Δ</sup>	,167	

	Continuum linguistique (enfant)		Langue des activités de littératie		Type de foyer (PLOP)		Type de foyer (relatif à l'enfant)		Continuum linguistique (mère-enfant)		Continuum linguistique (père-enfant)	
	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET
<b>Faible exposition au français</b>												
<b>G1 v. G2</b>												
Niveau de base	,266*	,119	,343*	,113	,416*	,169	,420*	,152	,348**	,119	,247	,158
1 <sup>er</sup> post-test DD	-,331**	,100	-,398**	,115	-,316*	,097	-,290*	,109	-,319*	,118	-,356*	,141
2 <sup>e</sup> post-test DD	-,218	,187	-,251	,185	-,153	,220	-,161	,228	-,214	,160	-,168	,245
3 <sup>e</sup> post-test DD	-,319* <sup>Δ</sup>	,112	-,295*	,140	-,166	,190	-,204	,174	-,173	,172	-,178	,151
<b>G1 v. G3</b>												
Niveau de base	-,338 <sup>†</sup>	,075	-,285*	,113	-,072	,145	-,046	,108	-,224*	,096	-,299*	,108
1 <sup>er</sup> post-test DD	-,405 <sup>†</sup>	,095	-,394** <sup>Δ</sup>	,127	-,388*	,101	-,440* <sup>Δ</sup>	,098	-,389* <sup>Δ</sup>	,111	-,324*	,130
2 <sup>e</sup> post-test DD	-,325 <sup>m</sup>	,170	-,298 <sup>Δ</sup>	,186	-,212	,181	-,455 <sup>mΔ</sup>	,220	-,442* <sup>Δ</sup>	,125	-,226	,226
3 <sup>e</sup> post-test DD	-,499 <sup>†Δ</sup>	,110	-,472** <sup>Δ</sup>	,119	-,443* <sup>Δ</sup>	,069	-,625* <sup>Δ</sup>	,131	-,527* <sup>Δ</sup>	,113	-,463* <sup>Δ</sup>	,133
Interaction triple (F de Wald)	F (5, 15) = 1,79		F (5, 15) = 3,88*		F (5, 15) = 2,38 <sup>m</sup>		F (5, 15) = 5,25**		F (5, 15) = 3,46*		F (5, 15) = 3,57*	
Nombre de participants	230		230		230		230		230		223	

**Note :** Les erreurs-types, robustes à l'hétérogénéité, ont été regroupées par garderie. Les variables de foyer sont dichotomiques. Par conséquent, les estimations représentent le contraste entre les familles endogames francophones et les autres types de familles. Pour les quatre autres variables, le tableau rapporte les estimations, selon modèle de régression, des effets du programme à des niveaux d'utilisation de la langue française choisis arbitrairement. L'estimation de la borne supérieure a été calculée pour le score maximal sur l'échelle, tandis que l'estimation de la borne inférieure a été évaluée en fonction de la valeur se situant un écart type sous la moyenne. Le groupe programme en garderie (G1) est mis en référence, donc les effets négatifs indiquent un avantage pour ce groupe. Les variables continues ont été centrées à la moyenne avant les analyses. L'évaluation au niveau de base et le groupe programme en garderie sont les catégories de référence pour tous les effets DD et DDD.

<sup>m</sup>  $p < 0,10$ , \* $p < 0,05$ , \*\* $p < 0,01$ , <sup>†</sup>  $p < 0,001$  ; <sup>Δ</sup> indique les effets qui varient significativement en fonction des divers degrés d'exposition au français  $p < 0,10$ .

**Échelle de Conscience de soi (standardisée).** Les analyses des scores standardisés de Conscience de soi ont révélé des interactions statistiquement significatives pour seulement une des six caractéristiques linguistiques, selon le test global F de Wald. Le seul modérateur significatif de l'effet du programme était le continuum linguistique parlé par le père, test d'interaction triple: F de Wald (5, 16) = 3,68\*. Aucun des tests de l'effet DDD des interactions simples n'était significatif du point de vue statistique, ce qui signifie qu'aucune des estimations DD de l'effet du programme n'a varié de manière significative lorsque les niveaux d'exposition de base au français étaient manipulés. L'ensemble des analyses n'est par conséquent pas rapporté étant donné qu'il serait redondant avec les résultats présentés dans la section 5.1.2 (offerts sur demande).

**Échelle des Habiletés cognitives (standardisée).** Les scores standardisés d'Habiletés cognitives ont été analysés de la même manière que ci-dessus. Les analyses ont révélé des interactions triples statistiquement significatives pour cinq des six modérateurs, selon les tests globaux F de Wald. Toutefois, le patron de ces interactions est inverse à celui qui a été rapporté pour l'échelle de Communication (tableau 5.21). Dans le présent cas, les enfants à forte exposition au français du groupe programme en garderie ont présenté des gains supérieurs comparativement à leurs homologues des groupes témoins, pour le troisième post-test en particulier (comme le prouve le résultat des tests DDD, indiqué par le symbole  $\Delta$ ). Peu importe si les enfants ont été admis dans le programme avec un fort ou un faible degré d'exposition, l'effet initial positif du programme associé au premier post-test était statistiquement significatif.<sup>82</sup> Toutes les estimations conditionnelles (au modérateur linguistique) de l'effet du programme sont rapportées dans le tableau 5.23.

Les simples tests d'interaction triple produisent des résultats cohérents pour tous les modérateurs linguistiques, pour les effets du troisième post-test impliquant les groupes en garderie (tel qu'indiqué par le symbole  $\Delta$ ). Des effets parallèles ont été observés relativement au groupe témoin hors garderie pour les deux modérateurs de type de foyer seulement. Étant donné qu'il s'agit de la période où la plupart des enfants de l'échantillon ont été inscrits pour la première fois à l'école (52 %), ces effets sont particulièrement intéressants. Pris dans leur ensemble, les résultats indiquent qu'au moment où les enfants entrent à l'école, ceux qui ont *la plus forte exposition* au français au niveau de base bénéficient le plus du programme, particulièrement pour ce qui est de leur développement cognitif.

---

<sup>82</sup> Il est à noter que dans le tableau 5.23, certaines des estimations conditionnelles indiquent un effet significativement négatif du programme pour les enfants à faible exposition au français (c.-à-d. que les coefficients DD sont positifs et significatifs). Ces estimations sont spécifiquement celles associées aux variables de type de ménage. Ces estimations ne doivent pas être interprétées avec trop de force étant donné qu'elles ne sont pas corroborées par les autres modérateurs (continus).



**Tableau 5.23 – Effets estimés du programme sur les scores standardisés d’Habilités cognitives (version française) conditionnels au profil linguistique**

	<i>Modérateurs de l’effet du programme</i>											
	Continuum linguistique (enfant)		Langue des activités de littératie		Type de foyer (PLOP)		Type de foyer (relatif à l’enfant)		Continuum linguistique (mère-enfant)		Continuum linguistique (père-enfant)	
	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET
<b>Forte exposition au français</b>												
<b>G1 v. G2</b>												
Niveau de base	,246	,147	,277	,153	,113	,166	,139	,131	,265	,163	,301	,153
1 <sup>er</sup> post-test DD	-363**	,126	-,366**	,155	-,210	,146	-,353*	,138	-,282	,172	-,387**	,132
2 <sup>e</sup> post-test DD	-230	,177	-,200	,169	-,187	,147	-,230	,173	-,133	,213	-,191	,162
3 <sup>e</sup> post-test DD	-,390* <sup>Δ</sup>	,145	-,413** <sup>Δ</sup>	,154	-,336* <sup>Δ</sup>	,158	-,380* <sup>Δ</sup>	,147	-,265 <sup>Δ</sup>	,210	-,309 <sup>mΔ</sup>	,150
<b>G1v. G3</b>												
Niveau de base	,283	,156	,236	,095	,242	,159	,267	,142	,237 <sup>m</sup>	,112	,243	,150
1 <sup>er</sup> post-test DD	-,342**	,118	-,296**	,112	-,301*	,123	-,425** <sup>Δ</sup>	,117	-,256 <sup>m</sup>	,133	-,337**	,109
2 <sup>e</sup> post-test DD	-,260	,147	-,180	,124	-,314* <sup>Δ</sup>	,133	-,327* <sup>Δ</sup>	,145	-,155	,172	-,275 <sup>m</sup>	,132
3 <sup>e</sup> post-test DD	-,451 <sup>†</sup>	,093	-,404 <sup>†</sup>	,097	-,451** <sup>Δ</sup>	,129	-,527 <sup>†Δ</sup>	,115	-,330 <sup>m</sup>	,170	-,414 <sup>†</sup>	,099

Faible exposition au français	Continuum linguistique (enfant)		Langue des activités de littératie		Type de foyer (PLOP)		Type de foyer (relatif à l'enfant)		Continuum linguistique (mère-enfant)		Continuum linguistique (père-enfant)	
	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET
<b>G1v. G2</b>												
Niveau de base	,204	,160	,189	,169	,332	,213	,319	,207	,217	,212	,140	,181
1er post-test DD	-,376 <sup>†</sup>	,077	-,331**	,085	-,417*	,133	-,253*	,108	-,413*	,112	-,302*	,116
2e post-test DD	,023	,177	-,039	,155	,078	,200	,106	,171	-,030	,169	-,010	,225
3e post-test DD	,205 <sup>Δ</sup>	,141	,239 <sup>Δ</sup>	,141	,303* <sup>Δ</sup>	,141	,307 <sup>mΔ</sup>	,148	,201	,148	,141	,135
<b>G1 v. G3</b>												
Niveau de base	-,262	,185	,147	,459	-,024	,265	-,077	,274	-,007	,243	,059	,246
1er post-test DD	-,262**	,084	-,451**	,148	-,011	,174	,214 <sup>Δ</sup>	,152	-,055	,113	-,118	,304
2e post-test DD	-,157	,167	-,635 <sup>m</sup>	,319	,135 <sup>Δ</sup>	,164	,069 <sup>Δ</sup>	,117	-,320 <sup>m</sup>	,155	-,203	,207
3e post-test DD	-,319	,318	-,520	,330	,064 <sup>Δ</sup>	,149	,104 <sup>Δ</sup>	,238	-,156	,280	-,633**	,168
Interaction triple (F de Wald)	F (5, 15) = 1,54		F (5, 15) = 3,46*		F (5, 15) = 4,21*		F (5, 15) = 7,22**		F (5, 15) = 2,87*		F (5, 15) = 6,49*	
Nombre de participants	211		211		211		211		211		210	

**Note :** Les erreurs-types, robustes à l'hétérogénéité, ont été regroupées par garderie. Les variables de foyer sont dichotomiques. Par conséquent, les estimations représentent le contraste entre les familles endogames francophones et les autres types de famille. Pour les quatre autres variables, le tableau rapporte les estimations, selon le modèle de régression, des effets du programme à des niveaux d'utilisation de la langue française choisis arbitrairement. L'estimation de la borne supérieure a été calculée pour le score maximal sur l'échelle, tandis que l'estimation de la borne inférieure a été évaluée en fonction de la valeur se situant à un écart type sous la moyenne. **Le groupe programme en garderie(G1) est mis en référence, donc les effets négatifs indiquent un avantage pour ce groupe.** Les variables continues ont été centrées à la moyenne avant les analyses. L'évaluation au niveau de base et le groupe programme en garderie sont les catégories de référence pour tous les effets DD et DDD.

<sup>m</sup>  $p < 0,10$ , \* $p < 0,05$ , \*\* $p < 0,01$ , †  $p < 0,001$ ; Δ indique des effets qui varient significativement en fonction du profil linguistique.

**Échelle d'Habilité physique (standardisée).** Comme dans les analyses rapportées dans la section 5.1.2, nous n'avons pas prévu que le programme testé ait une incidence sur le développement des compétences motrices étant donné que le programme ciblait principalement le développement de compétences en langue française et le développement culturel. En ce sens, il est surprenant que des interactions triples statistiquement significatives ( $p < 0,10$ ) aient été observées pour toutes les variables modératrices, selon le test F de Wald. L'interaction provient d'un effet programme relatif au groupe témoin en garderie pour le deuxième et, en particulier, le troisième post-test (où les interactions simples sont significatives pour tous les modérateurs, comme l'indique le symbole  $\Delta$  dans le tableau 5.24).

Ce résultat inattendu a été examiné plus en profondeur dans une série d'analyses de suivi. Premièrement, nous avons divisé l'échelle en éléments mesurant le développement des compétences motrices fines et globales. Ces analyses ont révélé que les éléments relatifs aux compétences motrices globales étaient à l'origine de l'effet. Aucun élément du programme testé ne se répercute toutefois sur cette résultante particulière. Dans des analyses ultérieures, le modèle de régression a été réestimé, séparément pour chacune des communautés (non rapportée). Ces analyses ont indiqué que les échantillons de Cornwall et de Durham étaient à l'origine de l'effet. Des analyses de suivi ont été effectuées pour vérifier si cette combinaison de facteurs (forte exposition, situé à Cornwall ou à Durham) est associée à un groupe d'enfants qui varient systématiquement entre les deux groupes expérimentaux en garderie. Diverses variables confondantes potentielles (pour une liste des facteurs affectant le développement physique, voir Venetsanou & Kambas, 2010) ont été examinées comme explications possibles de cet effet, tel que la présence ou non d'un problème de santé au cours de la période en question, le nombre de maladies chroniques, le revenu familial, le nombre de frères et de sœurs, ou le nombre et la fréquence des activités sportives. Aucun de ces facteurs n'expliquait le résultat obtenu, qui doit être interprété avec prudence étant donné son caractère imprévu. Nous préférons éviter d'avancer des interprétations spéculatives avant de vérifier s'il se reproduit avec la deuxième cohorte.

**Échelle de Vocabulaire expressif (standardisée).** Les analyses des scores de Vocabulaire expressif ont permis d'identifier des interactions triples statistiquement significatives selon le test F de Wald pour deux des cinq variables modératrices : langue des activités de littératie et continuum linguistique parlé par le père à l'enfant. Toutefois, aucun des tests d'interaction simple (c.-à-d. les tests DDD indiqués par le symbole  $\Delta$ ) pour les estimations DD d'effet particulier du programme n'était statistiquement significatif. Autrement dit, nous ne pouvons pas prétendre que la distinction entre les estimations conditionnelles pour les enfants à forte et à faible exposition au français, présentées dans le tableau 5.25, est statistiquement fiable. En raison du fait que cette résultante est d'un intérêt particulier, les estimations conditionnelles sont néanmoins rapportées comme une analyse descriptive par sous-échantillon.

Dans les comparaisons entre groupes en garderie, nous observons que l'effet du programme pour le *premier post-test* est plus grand pour les enfants fortement exposés au français que pour les enfants qui y sont faiblement exposés, bien que ce ne soit pas significatif. L'association avec l'exposition linguistique va ici dans le sens contraire de ce qui avait été prédit. Toutefois, l'effet a été observé pour les conditions de forte et de faible exposition au français, ce qui signifie que ces gains précoces ne varient pas selon les degrés d'exposition linguistique. L'exposition linguistique a modéré l'effet du programme de la manière prévue pour le troisième post-test,

pour les variables linguistiques suivantes: le continuum linguistique parlé par l'enfant, la langue des activités de littératie et le continuum linguistique parlé par le père à l'enfant.

Relativement au groupe témoin hors garderie, un effet est observé pour le premier post-test, pour les enfants à faible exposition au français. Cet effet n'est pas observé pour les enfants à forte exposition. Par conséquent, il semble y avoir ici un avantage conditionnel, mais cet avantage n'est pas présent au troisième post-test lorsque la majorité des enfants sont admis à l'école.

**Tableau 5.24 : Effets estimés du programme sur les scores standardisés d'Habilité physique (version française) conditionnels au profil linguistique**

	<i>Modérateurs de l'effet du programme</i>											
	Continuum linguistique (enfant)		Langue des activités de littératie		Type de foyer (PLOP)		Type de foyer (relatif à l'enfant)		Continuum linguistique (mère-enfant)		Continuum linguistique (père-enfant)	
<b>Fortes exposition au français</b>	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET
<b>G1 v. G2</b>												
Niveau de base	,325*	,165	,360*	,176	,184	,140	,196	,135	,277*	,123	,267	,150
1 <sup>er</sup> post-test DD	,106 <sup>Δ</sup>	,209	,016	,245	,128	,270	,070	,240	,054	,219	,074	,203
2 <sup>e</sup> post-test DD	-,401*	,169	-,432*	,200	-,359 <sup>mΔ</sup>	,189	-,341 <sup>mΔ</sup>	,195	-,331 <sup>m</sup>	,177	-,295 <sup>m</sup>	,159
3 <sup>e</sup> post-test DD	-,458 <sup>**Δ</sup>	,150	-,518 <sup>**Δ</sup>	,171	-,502 <sup>**Δ</sup>	,187	-,499 <sup>*Δ</sup>	,179	-,398 <sup>*Δ</sup>	,148	-,487 <sup>*Δ</sup>	,175
<b>G1 v. G3</b>												
Niveau de base	,059	,155	-,022	,156	-,138	,093	-,100	,114	-,060	,134	-,118	,134
1 <sup>er</sup> post-test DD	,094	,228	,144	,237	,188	,223	,158	,226	,147	,205	,164 <sup>Δ</sup>	,222
2 <sup>e</sup> post-test DD	-,160	,192	-,119	,207	-,110	,169	-,043	,193	-,014	,182	-,007	,197
3 <sup>e</sup> post-test DD	,006	,185	,089	,240	,088	,225	,099	,221	,144	,225	,082	,260

	Continuum linguistique (enfant)		Langue des activités de littératie		Type de foyer (PLOP)		Type de foyer (relatif à l'enfant)		Continuum linguistique (mère-enfant)		Continuum linguistique (père-enfant)	
	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET
<b>Faible exposition au français</b>												
<b>G1 v. G2</b>												
Niveau de base	-,071	,174	-,063	,164	,085	,242	,080	,228	-,015	,250	,037	,181
1 <sup>er</sup> post-test DD	-,299 <sup>Δ</sup>	,163	-,129	,166	-,265	,141	-,183	,173	-,220	,161	-,245	,204
2 <sup>e</sup> post-test DD	-,136	,234	-,111	,187	-,152 <sup>Δ</sup>	,240	-,178	,216	-,169	,206	-,224	,264
3 <sup>e</sup> post-test DD	,119 <sup>Δ</sup>	,253	,183 <sup>Δ</sup>	,246	,214 <sup>Δ</sup>	,246	,170	,226	,133	,270	,087	,174
<b>G1 v. G3</b>												
Niveau de base	-,460*	,217	,151	,514	-,153	,323	-,178	,270	-,205	,198	-,010	,233
1 <sup>er</sup> post-test DD	-,109	,629	-,428	,555	-,033	,469	-,068	,481	-,122	,561	-,392 <sup>Δ</sup>	,399
2 <sup>e</sup> post-test DD	,162	,393	-,235	,377	,187	,345	,068	,367	-,055	,378	-,034	,167
3 <sup>e</sup> post-test DD	-,047	,477	-,479	,419	,121	,386	-,094	,329	-,204	,455	-,202	,151
Interaction triple (F de Wald)	F (5, 15) = 2,76 <sup>m</sup>		F (5, 15) = 2,33 <sup>m</sup>		F (5, 15) = 4,88**		F (5, 15) = 7,92**		F (5, 15) = 3,74*		F (5, 15) = 6,27*	
Nombre de participants	211		211		211		211		211		205	

Note: Les erreurs-types, robustes à l'hétérogénéité, sont été regroupées par garderie. Les variables de foyer sont dichotomiques. Par conséquent, les estimations représentent le contraste entre les familles endogames francophones et les autres types de familles. Pour les quatre autres variables, le tableau rapporte les estimations, selon le modèle de régression, des effets du programme à des niveaux d'utilisation de la langue française choisis arbitrairement. L'estimation de la borne supérieure a été calculée pour le score maximal sur l'échelle, tandis que l'estimation de la borne inférieure a été évaluée en fonction de la valeur se situant à un écart type sous la moyenne. **Le groupe programme en garderie (G1) est mis en référence, donc les effets négatifs indiquent un avantage pour ce groupe.** Les variables continues ont été centrées à la moyenne avant les analyses. L'évaluation au niveau de base et le groupe programme en garderie sont les catégories de référence pour tous les effets DD et DDD.

<sup>m</sup>  $p < 0,10$ , \* $p < 0,05$ , \*\* $p < 0,01$ , †  $p < 0,001$ ; Δ indique des effets qui varient significativement en fonction des divers degrés d'exposition au français  $p < 0,10$ .

**Tableau 5.25 : Effets estimés du programme sur les scores standardisés de Vocabulaire expressif (version française) conditionnels au profil linguistique**

	<i>Modérateurs de l'effet du programme</i>											
	Continuum linguistique (enfant)		Langue des activités de littératie		Type de foyer (PLOP)		Type de foyer (relatif à l'enfant)		Continuum linguistique (mère-enfant)		Continuum linguistique (père-enfant)	
<b>Forte exposition au français</b>	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET
<b>G1v. G2</b>												
Niveau de base	,420*	,179	,500*	,224	,468*	,189	,358*	,218	,412*	,199	,423*	,196
1 <sup>er</sup> post-test DD	-,644**	,181	-,623**	,204	-,632**	,158	-,669**	,195	-,644**	,188	-,576**	,180
2 <sup>e</sup> post-test DD	-,227	,169	-,277	,184	-,311*	,137	-,245	,175	-,271	,201	-,154	,180
3 <sup>e</sup> post-test DD	-,051	,209	-,030	,226	-,172	,169	-,073	,215	-,126	,196	,014	,209
<b>G1v. G3</b>												
Niveau de base	,081	,110	,146	,157	,059	,095	,005	,133	,042	,132	,060	,115
1 <sup>er</sup> post-test DD	-,145	,178	-,143	,156	-,056	,161	-,116	,168	-,157	,193	-,160	,158
2 <sup>e</sup> post-test DD	,067	,199	-,037	,215	,098	,153	,067	,179	,063	,229	,052	,218
3 <sup>e</sup> post-test DD	,129	,133	,140	,158	,193	,097	,167	,112	,227	,128	,169	,145

	Continuum linguistique (enfant)		Langue des activités de littératie		Type de foyer (PLOP)		Type de foyer (relatif à l'enfant)		Continuum linguistique (mère-enfant)		Continuum linguistique (père-enfant)	
	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET
<b>Faible exposition au français</b>												
<b>G1 v. G2</b>												
Niveau de base	,236	,133	,262	,166	,333	,182	,333	,182	,354	,164	,311	,163
1er post-test DD	-,384*	,110	-,432*	,081	-,454*	,157	-,454*	,157	-,370*	,086	-,556*	,147
2e post-test DD	-,272	,292	-,211	,287	-,106	,268	-,106	,268	-,135	,263	-,336	,317
3e post-test DD	-,360*	,137	-,355*	,164	-,116	,217	-,116	,217	-,153	,166	-,335 <sup>m</sup>	,179
<b>G1 v. G3</b>												
Niveau de base	-,279	,378	-,320	,528	,188	,288	,188	,288	,161	,229	-,129	,378
1er post-test DD	-,464*	,400	-,175	,338	-,443*	,160	-,443*	,160	-,317	,272	-,353 <sup>m</sup>	,193
2e post-test DD	-,198	,405	,216	,344	-,176	,295	-,176	,295	-,204	,320	-,125	,438
3e post-test DD	-,159	,419	-,245	,330	-,184	,203	-,184	,203	-,337	,426	,141	,150
Interaction triple (F de Wald)	F (5, 15) = 1,32		F (5, 15) = 4,51**		F (5, 15) = 2,01		F (5, 15) = 1,28		F (5, 15) = 1,55		F (5, 15) = 3,68**	
Nombre de participants	211		211		211		211		211		211	

**Note :** Les variables de foyer sont dichotomiques. Par conséquent, les estimations représentent le contraste entre les familles endogames francophones et les autres types de familles. Pour les deux variables restantes, le tableau rapporte des estimations, selon modèle de régression, des effets du programme à des niveaux d'utilisation de la langue française choisis arbitrairement. L'estimation de la borne supérieure a été calculée pour le score maximal sur l'échelle, tandis que l'estimation de la borne inférieure a été évaluée en fonction de la valeur se situant à un écart-type sous la moyenne. **Le groupe programme en garderie (G1) est en référence, donc les effets négatifs indiquent un avantage pour ce groupe.** Les variables continues ont été centrées à la moyenne avant les analyses. L'évaluation au niveau de base et le groupe programme en garderie sont les catégories de référence pour tous les effets DD et DDD.

<sup>m</sup>  $p < 0,10$ , \*  $p < 0,05$ , \*\*  $p < 0,01$ , †  $p < 0,001$ ; Aucun des effets rapportés n'a varié de manière significative en fonction du profil linguistique.



## Sommaire

Les résultats rapportés dans la présente section ont confirmé que les variables linguistiques d'«exposition active» étaient les plus utiles pour expliquer les effets hétérogènes du traitement. Les effets modérateurs clés étaient presque invariablement obtenus en présence de ces variables (p. ex., continuum linguistique parlé par l'enfant). Le patron exact des résultats et les caractéristiques linguistiques qui se sont avérées pertinentes étaient fonction de la résultante et la période d'évaluation considérées.

Comme prévu, les enfants ayant la plus forte exposition au français au début du projet ont été le moins influencés dans le développement de leurs compétences linguistiques (c.-à-d. échelles de Communication et de Vocabulaire expressif), tandis que les gains associés au programme étaient positifs et statistiquement significatifs pour ces résultantes lorsque l'exposition au français était faible (pour certaines caractéristiques linguistiques). Ce patron de résultats était partiellement inversé dans le cas de l'échelle d'Habiletés cognitives où ce sont les enfants à forte exposition qui ont montré les effets positifs du programme les plus importants. Ces effets étaient significatifs tant pour le premier que pour le troisième post-test, moment où les enfants commençaient à être inscrits à la prématernelle. L'interprétation de cette apparente dissociation est abordée dans le résumé du chapitre et dans le chapitre 7 (discussion générale).

## 5.2 IMPACTS SUR LES PARENTS

Le programme évalué a été conçu pour fonctionner à deux niveaux: enfants et parents. Les estimations de l'impact direct du programme sur les résultantes liées au développement des enfants ont été rapportées à la section précédente. Nous tenterons maintenant d'évaluer si les Ateliers familles ont eu un impact significatif sur les connaissances, croyances, attitudes et comportements déclarés par les parents.

Les analyses sont présentées en deux sections. La première fournit les résultats associés aux comportements des parents lors des activités de littératie et à la langue dans laquelle ces activités se déroulent. La deuxième partie porte sur les analyses d'autres échelles conçues pour évaluer les connaissances, l'auto-efficacité, les comportements de modélisation et les croyances déclarées par les parents (pour plus d'information, voir le chapitre 2 et les descriptions plus détaillées ci-dessous). Comme dans les analyses des résultantes pour les enfants, nous évaluons le rôle que jouent le dosage (c.-à-d. l'assiduité des parents) et la qualité (c.-à-d. les indices de qualité des ateliers) pour modérer l'impact des ateliers.

### 5.2.1 Fréquence et langue des activités de littératie

Un volet important des ateliers d'activités de littératie était l'emphase qu'ils plaçaient sur l'importance de mener fréquemment, et en français, de telles activités à la maison afin de favoriser le développement linguistique des enfants francophones. Dans une première série d'analyses, nous rapportons les impacts estimés du programme en ce qui a trait à la fréquence déclarée et à la langue des activités de littératie. Ces analyses sont basées sur toutes les périodes de temps inclusivement et elles ne portent que sur les quatre communautés où l'on retrouve les trois groupes expérimentaux (Cornwall, Orléans, Durham et Edmundston). Nous rappelons au

lecteur que les Ateliers d'alphabétisation familiale étaient offerts seulement durant la première année du programme testé.

Avant d'examiner les résultats, nous associons d'abord l'animation des différents ateliers aux périodes d'évaluation afin d'offrir un support à l'interprétation des résultats. Les résultantes ont été recueillies à partir de huit sondages: le sondage de base et sept sondages de suivi (le premier sondage de suivi, effectué en octobre 2007, n'a été administré qu'aux parents inscrits au projet avant le 1<sup>er</sup> septembre 2007). Les données des deux premiers sondages ont été combinés en un seul niveau de base, lequel a servi de mesure de pré-test (équivalente en terme de période à la première évaluation des enfants). Les autres sondages de suivi, commençant au troisième, correspondent à la deuxième jusqu'à la septième période d'évaluation des enfants, et donc au premier jusqu'au sixième post-test. Les termes «niveau de base» et «post-test» font par conséquent références aux sondages, et non aux évaluations, dans les analyses rapportées dans la présente sous-section.

Quatre Ateliers d'alphabétisation familiales ont en tout été animés entre le mois de novembre 2007 et le début du mois de décembre 2008, soit la période correspondant à ce que nous considérons comme la première période d'évaluation post-test (ou ici, au deuxième sondage de suivi mené en février 2008). Les six autres Ateliers d'alphabétisation familiales ont été animés entre la fin du mois de janvier 2008 et le mois de mars 2008, ce qui correspond aux première et deuxième périodes d'évaluation post-test (ou ici, aux deuxième et troisième sondages de suivi menés en février 2008 et juin 2008). La première période post-atelier correspond à la troisième période d'évaluation. La relation entre les évaluations et les sondages est décrite à la section 2.5.1.

### ***Détails de spécification du modèle***

La spécification du modèle de régression utilisée pour estimer les effets de la fréquence est presque identique à celle qui est décrite à la section 5.1.2. Nous décrirons maintenant les différences entre ces deux spécifications.

Les présentes analyses se distinguent de trois façons. Tout d'abord, elles comprennent toutes les périodes d'évaluation.<sup>83</sup> À l'instar des analyses précédentes, chaque période d'évaluation était représentée par une variable fictive dont le niveau de base est en référence. Les effets du programme ont été évalués pour le premier post-test (hiver 2008), jusqu'au sixième post-test (automne 2009). Ces périodes post-test correspondent aux troisième et septième évaluations. Ces variables fictives ont été croisées avec des variables fictives représentant l'appartenance aux groupes expérimentaux pour obtenir les variables qui ont servi à produire l'estimation DD. Ensuite, des variables fictives représentant l'inscription à l'école ont été incluses dans le modèle de base (comme à la section 6.1.2). Enfin, les erreurs-types pour ces analyses ont été regroupées par enfant plutôt que par garderie (voir la section 4.3). Cette décision a mené à une forte augmentation des degrés de liberté pour les analyses. Ces degrés de liberté supplémentaires nous

---

<sup>83</sup>Les erreurs-types n'ont pas été regroupées au niveau des garderies dans ces analyses, ce qui a entraîné une très forte augmentation des degrés de liberté, éliminant ainsi l'obstacle qui avait motivé la séparation de la première et de la deuxième année dans les analyses des résultantes pour les enfants.<sup>84</sup> De grandes divergences dans l'importance de l'effet estimé avec ces deux méthodes indiqueraient que les différences relatives entre les scores sont nécessaires à la détermination de la taille de l'effet. Un tel résultat est instructif, mais n'indique pas nécessairement que le résultat du test paramétrique pose problème. Un examen visuel a montré que la distribution des scores dans chaque cellule était raisonnable, compte tenu du petit nombre d'observations.<sup>85</sup> Des tests indépendants ont été menés pour chaque comparaison.

ont permis de tester avec une grande puissance statistique l'interaction générale entre le temps et les groupes expérimentaux, malgré la perte des degrés de liberté entraînée par l'inclusion de toutes les mesures dans les mêmes analyses.

### ***Fréquence des activités de littératie***

Les résultats des analyses DD menées quant à la fréquence des activités de littératie sont présentés au tableau 5.26. Pour commencer, notons que le groupe programme en garderie est supérieur au groupe témoin en garderie au niveau de base. Notons également que l'ampleur de l'effet est petit (environ 0,30), ce qui équivaut en termes réels à un point d'item (chaque item a été évalué sur une échelle de 1 à 5). L'implication pratique associée à cette différence est négligeable. Les résultats démontrent qu'aucune estimation DD comparant le groupe programme en garderie et le groupe témoin en garderie n'est significative.

Plus précisément, les estimations DD sont pour la plupart très petites, ce qui reflète un changement parallèle pour les deux groupes concernant cette résultante. Par contre, la plupart des estimations DD comparant le groupe programme en garderie au groupe témoin hors garderie sont significatives. Le patron de résultats suggère que les familles dont les enfants sont inscrits à une garderie francophone démontrent des gains plus importants, par rapport au niveau de base, que celles dont les enfants étaient assignés au groupe hors garderie. Nous ne pouvons pas attribuer cet effet spécifiquement à l'impact des Ateliers familles ou, plus généralement, à la participation au programme évalué, car nous avons constaté que les deux groupes en garderie étaient statistiquement équivalents.

Bien que non significative, l'interprétation de ces résultats gagnerait à être nuancée. Premièrement, notons que la cohérence interne de l'échelle était relativement basse ( $\alpha$  de Cronbach  $>0,60$ ), ce qui indique que des erreurs de mesure pourraient atténuer les estimations d'impact. Essentiellement, les questions soulevées quant à cette échelle étaient associées à des fréquences très précises (1-Jamais, 2-Une fois par semaine ou moins, 3-Quelquefois par semaine, 4-Une ou deux fois par jour, 5-Trois fois ou plus par jour). Le score théorique maximal pour l'échelle était de 25, c'est-à-dire, par exemple, que même pour obtenir un score de 20, les parents devaient avoir répondu 4 pour chacun des cinq items de l'échelle. Bref, la question est de savoir s'il y a suffisamment d'heures dans une semaine pour que les parents puissent investir autant de temps à toutes ces activités. L'extrémité supérieure de l'échelle représente peut-être une gamme de fréquences qui n'est pas réaliste selon le comportement mesuré, créant un effet plafond empirique. Les moyennes observées des groupes expérimentaux reposant sur la version non standardisée de l'échelle allaient de 14,39 (groupe témoin en garderie au niveau de base) à 17,35 (groupe programme en garderie au sixième post-test). Si l'échelle avait présentée une meilleure cohérence interne ou si les choix de réponses avaient permis une discrimination plus précise quant aux fréquences faible à moyenne du spectrum de fréquences, alors la taille des estimations d'impact aurait peut-être été plus importante.

**Tableau 5.26 : Effets différence des différences (DD) du groupe pour la fréquence déclarée des activités de littératie au foyer (scores standardisés)**

	<i>Inclusion séquentielles des covariables – Erreurs groupées par enfant</i>									
	(1)		(2)		(3)		(4)		(5)	
Effets programme	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
Différences de groupe au niveau de base (G1 v. G2) <sup>a</sup>	-,380*	,151	-,376 <sup>†</sup>	,140	-,409 <sup>†</sup>	,141	-,379 <sup>†</sup>	,141	-,293*	,135
DD 1 <sup>er</sup> post-test Mi-atelier	-,018	,155	-,026	,154	-,025	,155	-,024	,155	-,040	,157
DD 2 <sup>e</sup> post-test 1 <sup>er</sup> post-atelier	-,056	,158	-,074	,156	-,084	,158	-,084	,158	-,098	,160
DD 3 <sup>e</sup> post-test 2 <sup>e</sup> post-atelier	-,209	,167	-,241	,161	-,251	,162	-,254	,161	-,270	,164
DD 4 <sup>e</sup> post-test 3 <sup>e</sup> post-atelier	,026	,170	-,015	,154	-,010	,154	-,011	,155	-,027	,153
DD 5 <sup>e</sup> post-test 4 <sup>e</sup> post-atelier	-,052	,172	-,089	,161	-,089	,161	-,103	,160	-,099	,159
DD 6 <sup>e</sup> post-test 5 <sup>e</sup> post-atelier	-,012	,193	-,066	,182	-,063	,182	-,068	,180	-,081	,178
Différences de groupe au niveau de base (G1 v. G3) <sup>a</sup>	,009	,160	,033	,165	-,023	,165	,051	,171	,018	,161
DD 1 <sup>er</sup> post-test Mi-atelier	-,340 <sup>m</sup>	,177	-,408*	,179	-,411*	,180	-,418*	,180	-,384*	,182
DD 2 <sup>e</sup> post-test 1 <sup>er</sup> post-atelier	-,244	,169	-,297 <sup>m</sup>	,171	-,302 <sup>m</sup>	,172	-,306 <sup>m</sup>	,172	-,300 <sup>m</sup>	,174
DD 3 <sup>e</sup> post-test 2 <sup>e</sup> post-atelier	-,503 <sup>†</sup>	,175	-,544 <sup>†</sup>	,178	-,551 <sup>†</sup>	,180	-,568 <sup>†</sup>	,181	-,561 <sup>†</sup>	,184
DD 4 <sup>e</sup> post-test 3 <sup>e</sup> post-atelier	-,247	,188	-,304	,195	-,323 <sup>m</sup>	,195	-,335 <sup>m</sup>	,194	-,295	,189
DD 5 <sup>e</sup> post-test 4 <sup>e</sup> post-atelier	-,400*	,202	-,439*	,205	-,463*	,204	-,480*	,204	-,414*	,200
DD 6 <sup>e</sup> post-test 5 <sup>e</sup> post-atelier	-,408 <sup>m</sup>	,208	-,467*	,213	-,495*	,210	-,480*	,210	-,395 <sup>m</sup>	,203
Test F composite de Wald (interaction temps x groupe)	F (12, 231) = 1,09		F (12, 228) = 1,14		F (12, 227) = 1,20		F (12, 227) = 1,22		F (12, 220) = ,996	
Nombre de participants	243		240		239		239		232	

**Note :** La résultante a été standardisée, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type. **Le groupe programme en garderie (G1) est le groupe de référence, ce qui signifie que les valeurs négatives des estimations DD indiquent un avantage pour ce groupe.** Les erreurs-types ont été estimées grâce au programme SPSS pour le calcul de l'estimateur robuste à l'hétérogénéité de White (White, 1980). De plus, les erreurs-types ont été groupées par enfant afin d'ajuster pour les corrélations entre les résidus dans le temps.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , <sup>†</sup>  $p < 0,01$  (tous les tests sont bilatéraux).

<sup>a</sup> G1, G2 et G3 dénotent le groupe programme en garderie, le groupe témoin en garderie et le groupe témoin hors garderie, respectivement.

**Les détails de spécification du modèle sont présentés à la section 5.1.2. Il est à noter que l'une des covariables du modèle de base est l'inscription ou non de l'enfant à l'école.**

### ***Langue des activités de littératie***

Les résultats des analyses d'impact pour la langue des activités de littératie sont rapportés au tableau 5.27. Ces analyses permettent de vérifier si les Ateliers d'alphabétisation familiale ont atteint leur but qui était d'encourager les parents à créer pour leurs enfants un environnement propice à leur développement intellectuel dans la langue française. Dans le cas présent, la langue utilisée pour mener les activités de littératie a servi à représenter ce concept.

Les estimations DD rapportées au tableau indiquent que le taux de changement pour cette variable était similaire pour les deux groupes en garderie. En revanche, nous avons observé une différence significative entre le groupe programme en garderie et le groupe témoin hors garderie, pour la première période post-atelier (deuxième post-test). Un tel effet n'a toutefois pas été observé pour la période mi-atelier (premier post-test). La spécificité de ce résultat est compatible avec le fait que la question qui nous occupe, soit la langue des activités de littératie, a été abordée dans un atelier présenté de façon concourante à l'administration du premier post-test (fin janvier-début février 2008). Il est logique que l'effet observé ne soit apparu qu'au sondage suivant, alors que tous les participants avaient reçu la formation pertinente à cette résultante. Naturellement, s'il s'agissait vraiment d'un effet propre à l'atelier d'alphabétisation familiale, nous nous serions attendus à observer le même effet en comparant les deux groupes en garderie. Un résultat moins ambigu aurait été un effet significatif tant relativement au groupe témoin en garderie qu'au groupe témoin hors garderie. Bref, les analyses ont fourni peu de preuves définitives que les familles dont les enfants sont inscrits à un groupe programme en garderie se distinguent des familles assignées aux groupes témoin.

**Tableau 5.27 : Effets différence des différences (DD) du groupe pour la langue déclarée des activités de littératie au foyer (scores standardisés)**

Effets programme	Inclusion séquentielle des covariables – Erreurs groupées par enfant									
	(1)		(2)		(3)		(4)		(5)	
	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
Différences de groupe au niveau de base (G1 v. G2) <sup>a</sup>	,064	,135	,007	,024	,008	,025	,001	,027	,003	,029
DD 1 <sup>er</sup> post-test Mi-atelier	-,029	,095	-,026	,094	-,025	,094	-,025	,094	-,030	,096
DD 2 <sup>e</sup> post-test 1 <sup>er</sup> post-atelier	-,089	,097	-,095	,095	-,097	,095	-,097	,095	-,099	,097
DD 3 <sup>e</sup> post-test 2 <sup>e</sup> post-atelier	-,124	,109	-,128	,105	-,118	,105	-,118	,105	-,125	,107
DD 4 <sup>e</sup> post-test 3 <sup>e</sup> post-atelier	,209	,135	-,005	,060	-,003	,061	-,005	,061	,002	,063
DD 5 <sup>e</sup> post-test 4 <sup>e</sup> post-atelier	,171	,138	,014	,055	,021	,054	,017	,055	,023	,056
DD 6 <sup>e</sup> post-test 5 <sup>e</sup> post-atelier	,077	,142	-,058	,078	-,048	,079	-,060	,079	-,058	,081
Différences de groupe au niveau de base (G1 v. G3) <sup>a</sup>	,264 <sup>m</sup>	,141	,021	,025	,022	,027	,009	,031	,016	,032
DD 1 <sup>er</sup> post-test Mi-atelier	-,114	,092	-,132	,088	-,129	,088	-,129	,087	-,148	,090
DD 2 <sup>e</sup> post-test 1 <sup>er</sup> post-atelier	-,330 <sup>†</sup>	,101	-,303 <sup>†</sup>	,091	-,299 <sup>†</sup>	,091	-,299 <sup>†</sup>	,091	-,313 <sup>†</sup>	,094
DD 3 <sup>e</sup> post-test 2 <sup>e</sup> post-atelier	-,146	,114	-,110	,101	-,106	,101	-,103	,101	-,099	,104
DD 4 <sup>e</sup> post-test 3 <sup>e</sup> post-atelier	-,213	,164	-,084	,072	-,076	,072	-,069	,073	-,082	,075
DD 5 <sup>e</sup> post-test 4 <sup>e</sup> post-atelier	-,313 <sup>m</sup>	,169	-,127 <sup>m</sup>	,069	-,113	,068	-,106	,069	-,112	,072
DD 6 <sup>e</sup> post-test 5 <sup>e</sup> post-atelier	-,191	,175	,003	,102	,018	,103	,053	,103	,047	,108
Test F composite de Wald (interaction temps x groupe)	F (12, 231) = 2,62 <sup>†</sup>		F (12, 228) = 1,99*		F (12, 227) = 1,87*		F (12, 227) = 1,89*		F (12, 220) = 1,96*	
Nombre de participants	243		240		239		239		232	

Note : La résultante a été standardisée, c'est-à-dire que les effets peuvent être interprétés en unités d'écart-type. **Le groupe programme en garderie (G1) est le groupe de référence, ce qui signifie que les valeurs négatives des estimations DD indiquent un avantage pour ce groupe.** Les erreurs-types ont été estimées grâce au programme SPSS pour le calcul de l'estimateur robuste à l'hétérogénéité de White (White, 1980). De plus, les erreurs-types ont été groupées par enfant afin d'ajuster pour les corrélations entre les résidus dans le temps.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux).

<sup>a</sup> G1, G2 et G3 dénotent le groupe programme en garderie, le groupe témoin en garderie et le groupe témoin hors garderie, respectivement.

Les détails de spécification sont fournis dans le corps du texte.

## 5.2.2 Effets pré-post: Connaissances, auto-efficacité et comportements de modélisation déclarés par les répondants

Les ateliers étaient également conçus pour: a) informer les parents sur des questions concernant le développement de l'enfant, les stratégies pour les aider à mieux préparer leur enfant à l'école et la disponibilité de services francophones dans leur communauté (Connaissances); b) développer leur sentiment d'auto-efficacité dans leur rôle de parents comme premiers éducateurs de leur enfant (Auto-efficacité); et c) augmenter la fréquence à laquelle ils ont modélisé les activités de littératie auprès de leur enfant (Modélisation). Les analyses ont porté sur les participants, des six communautés, assignés au groupe programme en garderie. Les résultats de ces analyses comparant les familles participantes (participant) et non participantes (non-participant) aux Ateliers familles apparaissent au tableau 5.28, alors que les tests de changement sont rapportés dans le texte. Les résultats pour chaque indice sont décrits en détails dans des sections distinctes.

### *Détails de spécification*

Les effets rapportés dans cette section ne sont pas ajustés par l'inclusion d'une longue liste de covariables à cause du trop petit nombre de participants (Green, 1991). Nous avons toutefois des raisons de croire que des différences significatives existent entre les familles participantes et non-participantes. Le test de rang paramétrique et le test de rang non paramétrique sont rapportés comme une vérification contre la violation des postulats de distribution. Un trait distinctif des tests non paramétriques est qu'ils sont sensibles au rang relatif des scores, mais pas à leurs différences absolues.<sup>84</sup> Le test F de Welch (un test-t qui ne suppose pas l'homogénéité de la variance et la taille égale des groupe) et la statistique U de l'ordre des rangs de Mann-Whitney (non paramétrique) ont été effectués afin d'évaluer les différences qui existent entre des groupes indépendants.<sup>85</sup>

---

<sup>84</sup> De grandes divergences dans l'importance de l'effet estimé avec ces deux méthodes indiqueraient que les différences relatives entre les scores sont nécessaires à la détermination de la taille de l'effet. Un tel résultat est instructif, mais n'indique pas nécessairement que le résultat du test paramétrique pose problème. Un examen visuel a montré que la distribution des scores dans chaque cellule était raisonnable, compte tenu du petit nombre d'observations.<sup>85</sup> Des tests indépendants ont été menés pour chaque comparaison.

<sup>85</sup> Des tests indépendants ont été menés pour chaque comparaison.

**Tableau 5.28 : Comparaison des mesures déclarées par les parents pré/post-ateliers et participants v. non-participants**

	Point de repère de l'estimation du parent				N
	Pré - (rétrospective)		Post -		
	Moyenne	ET	Moyenne	ET	
Efficacité					
Participant	18,75	3,34	21,20	2,34	89
Non-participant	Voir post	Voir post	20,07	1,64	14
F de Welch (participant v. non-participant)	F (1, 33) = 5,48*, Z = 1,18		F (1, 22) = 5,05*, Z = -2,05*		
Connaissances					
Participant	17,5	3,26	20,81	2,20	90
Non-participant	Voir post	Voir post	17,73	3,31	15
F de Welch (participant v. non-participant)	F (1, 18) = ,06, Z = ,32		F (1, 16) = 12,12†, Z = 3,70†		
Modélisation					
Participant	18,73	3,28	18,88	3,03	88
Non-participant	18,17	3,24	19,13	2,20	15
F de Welch (participant v. non-participant)	F (1, 14) = ,32, Z = ,49		F (1, 24) = ,16, Z = -,14		

**Note :** Les valeurs pré-test sont rapportées pour les non-participants seulement lorsqu'elles étaient disponibles (prospectives). Les détails de spécification sont rapportés dans le texte.

Z = Valeur Z du test de la somme des rangs de Wilcoxon (non paramétrique).

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests bilatéraux).



### **Connaissances déclarées par les répondants**

Le patron des résultats obtenu pour l'échelle de Connaissances est cohérent avec les conditions qui ont été décrites, dans la section 3.2.4 du rapport, pour un « véritable » effet d'atelier. Les estimations pré-test obtenues auprès des familles participantes sont équivalentes aux estimations post-test des familles non participantes. De plus, il y a des preuves constantes de la présence d'un effet positif du programme.

Premièrement, un test-t à mesures répétées a confirmé que les parents ayant participé aux ateliers ont fourni des estimations significativement plus élevées que les estimations au niveau de base, soit en moyenne de 3,31 points, ce qui représente plus de la moitié de l'étendue d'un item,  $t(89)=13,26$ ,  $ET=0,25$ ,  $p<0,001$ ; rang pour observations appariées de Wilcoxon=3,160,  $Z=7,75$ ,  $p<0,001$ . Aucun des participants n'a rapporté une estimation post-test inférieure à l'estimation pré-test, et seulement 11 n'ont rapporté aucun changement. Deuxièmement, nous observons au tableau 5.28 une équivalence entre les familles participantes et non-participantes relativement aux estimations pré-test, ainsi qu'un avantage statistiquement significatif pour les participants quand leurs estimations post-test servent d'étalon pour la comparaison. L'effet post-test est de l'ordre de trois points, ce qui correspond bien à l'effet programme estimé sur la base du score de changement. Pris dans leur ensemble, ces résultats sont compatibles avec l'idée que les deux groupes étaient semblables avant les ateliers et que les groupes qui ont assisté aux ateliers ont bénéficié d'une poussée statistiquement significative. Aucune preuve positive de biais dans les estimations n'a été observée.

### **Auto-efficacité déclarée par les répondants**

Les résultats obtenus pour l'échelle d'Auto-efficacité ne correspondent que partiellement aux conditions, décrites précédemment, reflétant la présence de l'impact réel des Ateliers familles. D'abord, un test-t à mesures répétées indiquait que le changement observé par rapport aux estimations au niveau de base était de 2,45 points et statistiquement significatif,  $t(88)=11,62$ ,  $p<0,001$ ; rang pour observations appariées de Wilcoxon=2,775,  $Z=7,52$ ,  $p<0,001$ . Aucun des participants n'a rapporté que son sentiment d'auto-efficacité avait diminué et seulement 15 participants ont fourni des estimations équivalentes pour le niveau de base et la période post-test.

L'examen du tableau 5.28 indique que les participants aux ateliers rapportaient un plus faible sentiment d'auto-efficacité que les non-participants avant les ateliers. La différence était significative avec le test paramétrique et non significative avec le test non paramétrique. Cette différence initiale peut découler d'un état pré-test sous-estimé par les participants (Taylor et coll., 2009) ou elle peut refléter une surestimation de la part des non-participants (Moore & Tananis, 2009). La différence entre les groupes pourrait aussi être réelle et due à un processus d'auto-sélection où les parents qui se sentent plus efficaces dans leur rôle sont moins susceptibles de croire que les ateliers méritent qu'ils y accordent du temps. Quoi qu'il en soit, les résultats démontrent qu'à la période post-test les participants aux ateliers rapportaient une plus grande auto-efficacité que les non-participants. Cet effet ne peut être expliqué par les types de biais qui sont habituellement associés à une estimation faite avant/en l'absence de l'intervention (Moore & Tananis, 2009) ou aux estimations post-intervention (Taylor et coll., 2009). La taille de l'effet est de l'ordre d'un seul point, une fraction du score maximal pour l'item. Pris dans leur ensemble, les résultats sont compatibles avec l'idée que les ateliers ont eu un impact positif sur le sentiment d'auto-efficacité des participants. Toutefois, a) la taille de l'effet estimé n'est pas aussi

grande que pour les connaissances déclarées par les répondants et b) les résultats ne concordent pas aussi bien avec la situation idéale où les familles participantes et non-participantes étaient statistiquement équivalentes quant aux estimations pré-test.

### ***Comportements de modélisation déclarés par les répondants***

Les analyses de l'échelle de Modélisation déclarée indiquent que les participants et les non-participants étaient semblables tant pour les estimations au niveau de base que pour les estimations post-test. L'effet des ateliers sur cette résultante était négligeable, même lorsqu'examiné à l'aide d'analyses à mesures répétées très sensibles, effet=0,01,  $t(86)=0,03$ ,  $ET=0,34$ . Un total de 37 participants ont rapporté un changement positif, 39 participants ont rapporté un changement négatif et 11 n'ont rapporté aucun changement relativement au niveau de base. Aucune preuve d'un effet programme n'a été détectée pour cette mesure, mais cela peut être dû aux points d'ancrage sélectionnés en tant que choix de réponses. Comme pour l'échelle de fréquence des activités de littératie, les questions de modélisation faisaient état de fréquences très précises (1-Jamais, 2-Une fois par semaine ou moins, 3-Quelques fois par semaine, 4-Une ou deux fois par jour, 5-Trois fois ou plus par jour), ce qui signifie que le biais de décalage des réponses (de l'anglais « response-shift bias ») n'est pas problématique (Howard, 1980). L'échelle comprend cinq items, c'est-à-dire que le score théorique maximal est de 25. Dans ce contexte, les moyennes rapportées se rapprochent de 20 (voir le tableau 5.28) pour les participants aussi bien que les non-participants au pré-test, ce qui est excellent. Pour obtenir un score supérieur à 20, il faudrait une fréquence dépassant une ou deux fois par jour pour toutes les activités. On ne peut affirmer avec certitude qu'il est raisonnable pour les parents de faire chacune des activités de modélisation trois fois ou plus par jour. Si les choix de réponses offerts aux parents avaient été mieux étalonnés (c.-à-d. particulièrement au milieu de l'échelle), on aurait peut-être pu observer un changement pour cette mesure. En dernier lieu, les données du pré-test prospectif ont été recueillies après le début des ateliers pour certains participants, ce qui peut avoir contribué à atténuer l'effet des ateliers.<sup>86</sup>

### ***Croyances et attitudes déclarées par les répondants relativement au développement de l'enfant***

En plus des autres mesures, nous avons posé quatre questions sur les croyances et les attitudes parentales. L'objectif de ces questions était d'évaluer la mesure dans laquelle la participation aux ateliers avait influencé leurs opinions (échelle à 4 points : tout à fait d'accord; d'accord; pas d'accord; et pas du tout d'accord). Nous avons observé que les questions du pré-test (prospectif, 2<sup>e</sup> sondage de suivi) et du post-test (sondage post-programme) produisaient des distributions très asymétriques, tendant vers les réponses considérées comme « idéales » par le programme d'ateliers. En interprétant ce résultat, il faut se rappeler que certaines réponses « pré-test » ont été recueillies après le début des ateliers (novembre 2007) et que le contenu des ateliers susceptible d'influencer ces opinions a été présenté dans les quelques premières séances. Autrement dit, la comparaison pré-test et la comparaison avec les non-participants peuvent toutes deux être contaminées par l'exposition à l'atelier.

---

<sup>86</sup> Selon nos dossiers, au moins 94,28 % des parents ont complété le sondage après le début des ateliers (novembre 2007). Nous n'avons pas cette information pour les six autres parents. Ce problème a été corrigé pour la seconde cohorte. Une estimation non biaisée de l'effet programme pourra être obtenue à partir de cette seconde vague de données.

***Qui est responsable de l'éducation de mon enfant ?*** On a posé deux questions à ce sujet. La première question cherchait à déterminer si le répondant croyait que les éducatrices en service de garde ou les enseignants étaient principalement responsables de l'éducation de leur enfant. Dans l'ensemble, 74,4 % de l'échantillon n'était pas d'accord avec cet énoncé avant l'atelier (28,9 % pas du tout d'accord) et 72,2 % après l'atelier (34,4 % pas du tout d'accord). Parmi ceux qui se disaient d'accord avec l'énoncé, une minorité s'est dite tout à fait d'accord avec celui-ci (6,7 % pré, 4,4 % post). Étant donné que ce sujet a été abordé dans les premiers ateliers animés en novembre 2007, il n'est pas surprenant que le pré-test «prospectif», dont la validité a été compromise, a indiqué que les parents n'étaient pas susceptibles d'endosser cette croyance avant le début de l'atelier (moyenne=2,03, ET=0,87 sur une échelle à 4 points : pas du tout d'accord à tout à fait l'accord) et que cette croyance n'ait pas beaucoup changé suite à l'atelier (moyenne=1,98, ET=0,87),  $t$  corrélé(89)=0,65, Wilcoxon=329,  $Z$ =-0,65. Des participants répondants, 17 ont démontré un changement positif alors que 21 démontraient un changement négatif (52 participants n'ont à leur avis rapporté aucun changement). Les participants étaient pour ainsi dire identiques aux non-participants pour la comparaison post-test (non-participants, moyenne=1,93, ET=0,96),  $F$  de Welch (1, 18,07)=0,028,  $U$  de Mann-Whitney =647,50,  $Z$ =-0,266.

La deuxième question cherchait à déterminer si le parent collaborait avec les éducatrices pour stimuler le développement de son enfant et pour mieux le préparer à l'école. Seulement 9 % des participants n'étaient pas d'accord avec cet énoncé lors du pré-test (6,7 % pas du tout d'accord) et seulement 10 % n'étaient pas d'accord après les ateliers (3,3 % pas du tout d'accord). Parmi ceux qui étaient d'accord, 45,6 % se sont dits tout à fait d'accord avant l'atelier et 55,6 % après. Le changement apparent des réponses est corroboré par une plus forte tendance vers un impact positif du programme quand nous comparons les participants au fil du temps, du pré-test (moyenne=3,30, ET=0,82) au post-test (moyenne=3,43, ET=0,77),  $t(88)=1,29$ , Wilcoxon=499,  $Z=1,30$ ; et quand nous les comparons aux non-participants au post-test (non-participants, moyenne =3,07, ET=1,16),  $F$  de Welch (1, 16,07)=1,31;  $U$  de Mann-Whitney =582,  $Z$ =-0,953. Au total, 25 participants démontraient un changement positif, 15 démontraient un changement négatif et 49 ne démontraient aucun changement.

***Mon enfant est-il trop jeune pour qu'on lui fasse la lecture ?*** Une autre question demandait aux parents s'ils croyaient que leur enfant était trop jeune pour qu'on lui fasse la lecture d'histoires. Pour le pré-test, 92,2 % des participants et 100 % des non-participants étaient totalement en désaccord. Pour le post-test, seul le pourcentage de participants a changé légèrement, augmentant à 95,6 %, tandis que celui des non-participants a demeuré le même. Aucun test n'est rapporté pour cette mesure en raison de l'absence de variance, mais il est intéressant de noter que quatre participants ont déclaré être «tout à fait d'accord» avec cet énoncé avant l'atelier, mais qu'aucun ne l'a fait après coup.

***Le matériel d'alphabétisation est-il en français ?*** On a demandé aux parents si les livres, jouets, pièces musicales et photos présents à la maison reflétaient la culture francophone. Chez les participants, 26,4 % n'étaient pas d'accord avant l'atelier (3,4 % pas du tout d'accord) et 27,3 % suite à l'atelier (6,8 % pas du tout d'accord). De ceux qui étaient d'accord, 31 % et 36,4 % ont endossé « tout à fait d'accord » pour les périodes pré-test et post-test, respectivement. Comme les fréquences ont très peu changé suivant la présentation des ateliers, il n'est pas étonnant que les tests formels quant au changement d'opinion des participants n'ont révélé aucun changement significatif, moyenne pré-test=3,01, ET=0,84, moyenne post-test=3,05, ET=0,94;

$t(85)=-0,035$ ; Wilcoxon=573,  $Z=0,40$ . Dans le même ordre d'idées, les participants ( $n=90$ ) ne différaient pas significativement des non-participants ( $n=15$ ),  $F$  de Welch (1, 16,07)=1,31;  $U$  de Mann-Whitney =582,  $Z=0,953$ .

### 5.2.3 Modérateurs des effets sur les parents

Dans cette section, nous abordons l'impact de facteurs qui pourraient vraisemblablement améliorer l'effet produit par les Ateliers familles. Dans les analyses qui suivent, nous explorons le rôle que jouent la qualité des ateliers et les estimations précises de dosage relativement à l'ampleur de la différence qui existe entre les estimations pré-test et post-test fournies par les parents ayant assisté aux ateliers trois fois ou plus ( $N=90$ ). Ces analyses se limitent aux résultantes qui ont fait l'objet des analyses rapportées à la section 5.2.2.<sup>87</sup> Nous rapportons les estimations de la force de l'association entre les scores de changement et les estimations de la qualité des ateliers au tableau 5.29.<sup>88</sup>

#### *Détails de spécification*

Les effets apparaissant dans cette section sont rapportés sans être ajustés par l'inclusion d'une longue liste de covariables, à cause du trop petit nombre de participants (Green, 1991). Nous avons des raisons de croire que des différences significatives existent entre les participants et les non-participants. Toutefois, les tests  $F$  de Welch, basés sur les covariables décrites à la section 4.4, n'ont pu révéler aucune différence fiable entre les groupes.

Les tests de rang d'association, paramétriques et non paramétriques, sont rapportés comme une vérification contre la violation des postulats de distribution. Un caractère distinctif des tests non paramétriques est qu'ils sont sensibles au rang relatif des scores, mais pas à leurs différences absolues.<sup>89</sup> Le  $r$  de Pearson d'ordre zéro (paramétrique), le  $r$  de Pearson partiel et le rang tau de Kendall (non paramétriques) sont rapportés comme des mesures d'association. Dans tous les cas, le  $r$  de Pearson partiel est la corrélation entre un score de changement et une mesure de qualité, après avoir ajusté statistiquement pour le score pré-test.

---

<sup>87</sup> Ces résultantes, basées sur l'échantillon complet du projet Capacité d'apprentissage, ont fourni un échantillon ( $n=90$ ) qui approchait les points de référence conventionnels pour l'obtention d'estimations stables dans les analyses de régression (approximativement  $N=(104+k)$ , où  $k$  représente le nombre de prédicteurs (Green, 1991). À l'opposé, la taille de l'échantillon de participants au programme pour les analyses associées aux activités de littératie était environ 2/3 plus petite dû à l'exclusion d'Edmonton et de Saint-Jean au sein des analyses.

<sup>88</sup> Au moins deux méthodes permettent d'estimer la relation entre la qualité des ateliers et les scores de changement (Maris, 1998). L'une consisterait simplement à corrélérer le score de changement (post-test moins pré-test) avec la variable désirée. L'autre permettrait d'enrichir les analyses en contrôlant statistiquement pour la valeur pré-test lors des analyses des scores post-test. Les deux méthodes répondent à des questions légèrement différentes. La première vérifie si, en moyenne, les scores de changement sont associés aux autres indicateurs. La deuxième vérifie si la relation serait significative si tous les participants étaient équivalents à la mesure pré-test. Chaque méthode mène à des estimations biaisées sous certaines conditions liées à la présence d'erreurs de mesure et au modèle causal reliant le score pré-test à la sélection pour le traitement (Glymour, Weuve, Berkman, Kawachi & Robins, 2005). Par conséquent, les résultats que nous rapportons sont basés sur les deux stratégies. Qui plus est, étant donné le petit nombre de communautés participant aux analyses ( $n=4$ ), nous rapportons les estimations non paramétriques d'association pour vérifier si la distribution des scores de qualité peut jouer un rôle d'une importance disproportionnée lors de la détermination de la valeur des estimations. Nous sommes d'avis qu'il faut accorder plus de crédibilité aux résultats lorsque toutes les méthodes d'estimation mènent à la même conclusion.

<sup>89</sup> D'importantes divergences dans la taille de l'effet estimé par ces deux méthodes indiqueraient que les différences relatives entre les scores sont nécessaires à la détermination de la taille de l'effet. Un tel résultat est instructif, mais n'indique pas nécessairement que le résultat du test paramétrique pose problème. Un examen visuel a démontré que la distribution des scores dans chaque cellule était raisonnable compte tenu du petit nombre d'observations.

Les données pré-test étant disponibles pour tous les participants aux ateliers, on a utilisé un estimateur DD pour estimer l'effet d'un indicateur précis de la participation des parents (c.-à-d. la proportion d'ateliers fréquentés). Dans ce cas, les estimations pré-test et post-test sont incluses dans le modèle. Le temps était représenté par une variable fictive, que l'on a croisée avec la variable de participation des parents pour obtenir l'indicateur qui a servi à estimer l'effet traitement. Les seules covariables insérées dans les analyses étaient des variables fictives représentant l'affiliation à la communauté.

### ***Qualité de l'atelier associée aux scores de changement***

Dans le tableau 5.29, les estimations de l'effet de la qualité sont surlignées en gris si elles émergent, de façon constante à travers les tests, comme étant associées aux scores de changement des participants. Les deux indicateurs de qualité qui ont le mieux prédit les gains dans les scores de changement étaient : a) la mesure dans laquelle la longueur de la séance était jugée appropriée, et b) le style d'animation de l'intervenante alphabétisation familiale (style de l'intervenante). L'estimation de la qualité globale, issue de toutes les autres variables de qualité à l'exception de celles basées sur les réactions des parents (positives et négatives), est venue confirmer cette association. L'indicateur de couverture du contenu a été associé aux scores de changement lorsqu'ils n'étaient pas ajustés pour les valeurs pré-test. La relation associée est ambiguë parce qu'elle est confondue avec les scores pré-test. En d'autres mots, l'effet du traitement dans ce cas semble dépendre de l'état estimé des participants avant l'administration du traitement. Une autre source de divergence est la différence entre les estimateurs paramétriques et non paramétriques. Lorsque les tests paramétriques et non paramétriques (basés sur l'ordre de rang) se distinguent de façon marquée, cela signale que la taille de l'intervalle entre les scores, concernant soit le score de changement ou la mesure de qualité, a exagéré ou a masqué l'association entre les variables. Nous concluons que le succès des Ateliers d'alphabétisation familiale varie selon l'aptitude de la personne qui anime les ateliers et si le contenu est présenté dans la période de temps alloué.

**Tableau 5.29 : Changement des Connaissances, de l'Auto-efficacité et de la Modélisation déclarés par les répondants en fonction de la qualité des ateliers**

Score de changement	Qualité globale <sup>a</sup>	Couverture du contenu	Style de l'intervenante	Réactions positives	Réactions négatives	Longueur des séances
ΔConnaissances (n=90)	,26* (,19*)	,21* (,18*)	,19 <sup>m</sup> (,16 <sup>m</sup> )	,03 (,05)	,06 (,20*)	,22* (,15 <sup>m</sup> )
ΔConnaissances (partiel) (n=90)	,21*	,08	,22*	-,10	-,07	,16 <sup>m</sup>
ΔAuto-efficacité (n=89)	,39 <sup>†</sup> (,30 <sup>†</sup> )	,24* (,20*)	,32** (,26**)	,17 (,16*)	,08 (,25**)	,34 <sup>†</sup> (,22**)
ΔAuto-efficacité (partiel) (n=89)	,23*	,04	,28**	,03	-,18*	,14 <sup>m</sup>
ΔModélisation (n=87)	-,05 (-,08)	,32** (-,19)	-,06 (-,14)	,14 (-,08)	,14 (,10)	-,12 (-,03)
ΔModélisation (partiel) (n=87)	-,18*	,19*	-,24*	-,07	,07	-,21*

**Note :** Les résultats qui sont convergents à travers les stratégies d'analyse sont surlignés en gris. Scores de changement en fonction des indices de qualité. Les indices de qualité ont été obtenus à partir de notes d'observations. Les corrélations *r* de Pearson sont rapportées dans chaque cellule avec le rang tau de Kendall, une estimation d'association non paramétrique apparaissant entre parenthèses. Ce dernier n'émet aucun postulat quant à la forme de la distribution de scores. Les rangées indiquant « (partiel) » rapportent les corrélations partielles de Pearson, contrôlant pour les scores pré-test. Un équivalent non paramétrique de la corrélation partielle n'est pas disponible.

<sup>a</sup> La qualité globale est la moyenne de l'ensemble des indices de qualité, excluant les indices basés sur les réactions positives et négatives.

<sup>m</sup>  $p < 0,10$ , \* $p < 0,05$ , \*\* $p < 0,01$ , †  $p < 0,001$  (tests bilatéraux, N=88 à 90).

### **Présence des parents associée aux scores de changement**

Dans les analyses précédentes, nous avons fait la distinction entre les parents qui ont assisté à trois ateliers ou plus (participants) et ceux qui ont assisté à deux ateliers ou moins (non-participants). Ces derniers ont été classés comme n'ayant « pas participé » aux Ateliers d'alphabétisation familiale. Nous avons poussé plus loin ces premières analyses en estimant la mesure dans laquelle la variation du taux de participation des parents considérés comme participants aux ateliers était liée au changement que ceux-ci avaient déclaré pour les indicateurs examinés. Il est important de noter que l'exclusion des non-participants de ce calcul a restreint l'étendue des valeurs de participation. Il est donc vraisemblable que les analyses qui suivent représenteront des sous-estimations de l'ampleur de l'effet total du dosage. Nous rapportons les résultats comme un test conservateur de l'hypothèse nulle.

Nous avons inséré dans des analyses fondées sur l'estimateur DD une variable indépendante représentant la proportion d'ateliers fréquentés par au moins un parent de l'enfant. Pour ces analyses, nous avons employé l'estimateur d'erreurs-types robuste à l'hétérogénéité, mais nous n'avons pas employé d'ajustement pour le groupement.<sup>90</sup> Les résultats de ces analyses indiquent

<sup>90</sup> Nous avons inséré des effets fixes pour la communauté afin de tenir compte de la corrélation positive entre les résidus introduits par le site des ateliers. Cette approche est plus appropriée lorsque le nombre de groupes est limité, comme dans le cas présent (groupes < 10; Snijders & Bosker, 1999).

qu'un parent avec une présence moyenne devrait montrer un changement moyen de  $b=3,08$  ( $ET=0,43$ ), entre le pré-test et le post-test. Le changement moyen représente une augmentation de 0,96 écart-type par rapport au niveau de base, ce qui représente un effet dit important.

La quantité de gains au chapitre des Connaissances, du pré-test jusqu'au post-test, a bien sûr varié selon la famille. Une partie de cette variabilité a été expliquée par la présence des parents aux ateliers,  $DD(89) = 3,40$ ,  $ET = 1,73$ , où une augmentation estimée du score de changement de 0,11 écart-type est attendue pour chaque atelier additionnel fréquenté (c.-à-d. une augmentation de 0,10 de la proportion des ateliers fréquentés). Selon ce modèle, il est attendu que les familles où au moins un parent de l'enfant a fréquenté l'ensemble des dix ateliers profiteront d'une modification de 0,63 écart-type relativement à leur niveau de base et ce, en sus des gains effectués par le parent qui a assisté à seulement quatre ateliers soit en personne ou indirectement via la participation de l'autre parent de l'enfant. Les résultats corroborent les analyses antérieures comparant les parents participants aux parents non participants.<sup>91</sup> L'augmentation de l'exposition aux ateliers est associée à des gains déclarés par les parents au chapitre des Connaissances.

Dans le même ordre d'idées, un taux moyen de participation des parents est associé à une augmentation significative relativement au niveau de base de la mesure d'Auto-efficacité. Ce résultat confirme les résultats des analyses du changement rapportés plus tôt à l'aide du test-t à mesures répétées,  $b=2,22$ ,  $ET=0,38$ . Toutefois, l'estimation  $DD$  associée n'était pas statistiquement significative,  $DD=1,80$ ,  $ET=1,41$ , indiquant que le taux de changement pour cette résultante n'avait pas varié en fonction du dosage. Comme dans les analyses antérieures, les effets associés à l'échelle de Modélisation n'étaient pas près d'une signification statistique (non rapportés). En résumé, il semble que l'augmentation des niveaux d'exposition a eu un effet spécifique sur les gains de Connaissances déclarés par les parents.

### 5.3 SOMMAIRE

Dans cette section, nous avons rendu compte des impacts du programme pendant la première année du projet Capacité d'apprentissage pour les résultantes touchant les enfants et leurs parents. Nous résumons les principaux résultats de chaque analyse à tour de rôle. Une interprétation plus détaillée apparaît au chapitre de la Discussion (voir le chapitre 7).

#### **Résultantes des enfants**

Comparativement au groupe témoin en garderie, le groupe programme en garderie présentait un *désavantage* au niveau de base de l'ordre de 0,10 à 0,30 écart-type pour toutes les échelles de l'ÉPE-AD, sauf celle relative au Vocabulaire réceptif. Toutefois, ce désavantage n'était statistiquement significatif que pour la mesure de Vocabulaire expressif.<sup>92</sup> L'intervention en

---

<sup>91</sup> Il est possible que les parents dont le taux de participation est plus élevé aient reçu des ateliers de meilleure qualité. L'idée proposée derrière cette relation est la suivante : les parents devraient être plus enclins à participer aux ateliers s'ils croient que le temps consacré en valait la peine. On pourrait calculer et rapporter les estimations d'association entre la présence des parents et la qualité des ateliers, mais les résultats seraient difficiles à interpréter. La nature des indices de qualité est telle que les parents ayant participé à un plus grand nombre d'ateliers ont peut-être contribué à offrir davantage d'information aux indices de qualité. Ainsi, une estimation d'association calculée entre ces deux variables pourrait refléter un artefact méthodologique.

<sup>92</sup> Le désavantage initial donne à penser que le groupe programme en garderie s'orientait vers une trajectoire développementale inférieure à celle du groupe témoin en garderie avant l'intervention. Si ce postulat tient, les effets traitement observés sont donc des sous-estimations des effets traitement véritables. Les résultats observés indiqueraient par conséquent que la trajectoire

garderie eu pour impact d'*inverser ce désavantage pour le reste de la première année*. Les estimations DD de l'effet programme pour la Communication, la Conscience de soi, les Habiletés cognitives et le Vocabulaire expressif démontraient que le groupe programme en garderie se développait plus rapidement que le groupe témoin en garderie pour toutes les périodes post-test. Cet avantage était statistiquement significatif quatre mois après l'intervention, mais n'était plus statistiquement détectable par la suite. En somme, bien qu'il ait débuté dans une position défavorable, le groupe programme en garderie a démontré un développement statistiquement supérieur ou équivalent à celui du groupe témoin en garderie durant la première année du projet. Les comparaisons relatives au groupe témoin hors garderie étaient équivoques.

Plusieurs sources de données probantes convergentes pour étayer l'effet programme observé quatre mois après l'intervention. Les résultats démontraient une validité discriminante, où l'avantage du groupe programme était propre à aux domaines ciblés: Communication; Conscience de soi; Habiletés cognitives; et Vocabulaire expressif. Aucun effet n'a été observé où aucun n'était prévu (c.-à-d. Habileté physique). Dans une deuxième série d'analyses, le dosage, la fidélité au programme et la qualité du programme ont été employés en tant que définitions opérationnelles alternatives pour désigner le même concept: l'intensité du traitement. Les analyses de ces indicateurs ont mené à une série convergente de résultats.<sup>93</sup> Toutes les approches de conceptualisation du traitement tendent à converger vers un ensemble cohérent de résultats qui soutient notre théorie selon laquelle les effets du groupe expérimental sont effectivement valides.

Dans la mesure où le groupe programme en garderie a surmonté un déficit initial afin d'atteindre un développement équivalent durant la première année du projet, nous pouvons affirmer que l'intervention a eu un impact positif sur les résultantes associées à la préparation à la scolarisation. Cette affirmation repose sur le postulat voulant que le désavantage initial du groupe programme en garderie est signe d'une trajectoire développementale pré-intervention inférieure à celle du groupe témoin en garderie. Nous soutenons que si le groupe programme en garderie n'avait pas reçu l'intervention, il aurait obtenu des scores significativement moins élevés que le groupe témoin en garderie à la fin de la première année pour la plupart des mesures observées. Il faut par contre reconnaître que les résultats auraient été plus importants si nous avions pu observer des impacts positifs et statistiquement significatifs du programme pour toutes les périodes post-test. Nous attribuons les estimations DD nulles des deuxième et troisième post-tests à un manque de sensibilité des mesures utilisées plutôt qu'à une atténuation réelle de l'effet du programme (voir ci-dessous).

Le fait que les gains observés dus au programme ne soient pas statistiquement significatifs huit ou douze mois après l'intervention peut s'expliquer par: a) un biais des instruments causé par un manque de sensibilité de mesure lorsque les enfants vieillissent; et b) une légère baisse de cohérence dans la mise en œuvre du programme suivant le premier intervalle de quatre mois (voir le *Rapport de mise en œuvre du projet*, 2014), ce qui signifie que les programmes offerts par

---

développementale du groupe programme en garderie a été < corrigée > par l'intervention et alignée sur celle du groupe témoin en garderie. Dans cette perspective, les effets DD statistiquement nuls observés pour le vocabulaire expressif lors des deuxième et troisième post-tests prouvent que le traitement a eu un impact positif.

<sup>93</sup> Comme ces définitions de la validité du traitement n'étaient pas explicitement disponibles aux participants ou aux expérimentateurs, elles ne sont pas aussi susceptibles d'être contaminées par diverses formes de biais. Les expérimentateurs connaissaient le statut de traitement des participants et les effets attendus. On ne peut pas en dire autant des variables de dosage et de qualité/fidélité.



les deux types de garderie sont devenus davantage comparables avec le temps. Si la première cause influence les résultats, nous nous attendrions à voir apparaître des effets programme dans les analyses des données de la deuxième année, à compter de la sixième évaluation. À partir de cette évaluation, l'ÉPE-AD a été recalibrée pour être plus sensible aux différences individuelles observées dans l'étendue moyenne à élevée des habiletés (voir le chapitre 2, section 2.5.2). En outre, des mesures normalisées de Vocabulaire réceptif (ÉVIP-R) et de Vocabulaire expressif (EOWPVT) ont été prises durant cette période. Ces mesures proviennent d'éditeurs différents de celui de l'ÉPE-AD et elles permettent de discriminer les enfants dont les capacités linguistiques sont très variées. Pour ces raisons, nous croyons que les effets programme qui existaient à la fin de la première année seront plus faciles à détecter dans les analyses de suivi menées sur les résultats de la deuxième année, notamment lors des deux dernières périodes d'évaluation. De surcroît, les nouvelles mesures devraient être plus sensibles à l'émergence des impacts différés de l'exposition au programme durant la première année du projet. Si la deuxième explication est la cause du patron de résultats observé, alors une reproduction de l'intervention auprès d'autres sites pourrait bénéficier de mesures de protection supplémentaires pour assurer que les normes de mise en œuvre soient maintenues à de haut niveau de qualité tout au long de la première année du projet.

La sous-échelle de Vocabulaire expressif tendait à montrer les effets les plus importants. Les items de cette échelle ont été glanés des échelles de Communication et de Conscience de soi de l'ÉPE-AD. Les effets observés pour ces deux dernières échelles peuvent donc découler en grande partie des items associés au Vocabulaire expressif. Nous remarquons que les deux groupes en garderie se sont avérés être significativement différents au niveau de base pour le Vocabulaire expressif.<sup>94</sup> Comme nous l'avons déjà mentionné, ces différences au niveau de base sont compatibles avec l'hypothèse de différences préexistantes entre les trajectoires développementales des deux groupes. Dans cette perspective, le groupe programme en garderie se développait moins rapidement quant à cette mesure que le groupe témoin en garderie avant l'intervention. Ce résultat donne à penser que le postulat des pentes parallèles sous-tendant l'estimateur DD ne sera vraisemblablement pas respecté ces analyses. Si nous présumons que la trajectoire développementale du groupe programme aurait demeuré inférieure à celle du groupe témoin en garderie en l'absence d'intervention, alors les estimations DD représentent des estimations conservatrices ou des « sous-estimations » de l'effet traitement. Nous pensons que ce postulat est raisonnable et que les estimations DD qui résultent peuvent être interprétées comme un test conservateur des impacts du programme. En réalité, on peut considérer la correction apparente de la trajectoire développementale du groupe programme en garderie comme la preuve que le programme a eu un impact positif.

La seule résultante de l'ÉPE-AD qui a déjoué nos prévisions initiales à toutes les périodes d'évaluation est la sous-échelle de Vocabulaire réceptif. Nous pensions observer un effet programme positif sur cette résultante parce que cette mesure évalue de façon apparente la capacité des enfants de comprendre le français. L'interprétation de ce résultat doit être nuancée par le fait que cette mesure de vocabulaire a été élaborée à partir d'items de l'ÉPE-AD et que ses propriétés psychométriques ne sont donc pas bien documentées. La corrélation entre cette mesure et une mesure largement utilisée pour évaluer le même construit est de  $r=0,43$  (voir le

---

<sup>94</sup> Cette différence initiale pourrait s'expliquer par les différences initiales qui existaient entre les deux groupes concernant le niveau de français rapporté par les pères de l'étude, différences qui n'ont pas été entièrement corrigées en contrôlant pour le type de foyer (exogame. etc.).

chapitre 2, section 2.5.2). En fait, les analyses menées quant à la version non standardisée de l'échelle indiquait que cette mesure n'était pas sensible aux changements survenant au fil du temps pour l'échantillon dans son ensemble (DD au premier post-test =1,07, ET=1,16). Ainsi, il est possible que l'effet nul rapporté ici résulte d'une absence de sensibilité plutôt que d'une réelle absence d'impact sur le Vocabulaire réceptif.<sup>95</sup>

*Les analyses selon le profil linguistique de l'enfant* ont confirmé que le programme avait le plus d'incidence sur la capacité de communiquer en français lorsque les enfants sont au départ moins exposés à la langue française (p. ex., échelle de Communication). Les effets estimés du programme pour la condition de faible exposition ont toujours été cohérents avec la présence d'un effet du programme pour cette résultante (c.-à-d. dans la bonne direction). De plus, des effets statistiquement significatifs du programme s'observent pour la condition de faible exposition relativement au groupe témoin en garderie et au groupe témoin hors garderie. Dans ces analyses, les effets conditionnels estimés sont aussi plus constants à travers le temps, où des effets ont été observés pour les première, deuxième et troisième périodes post-test. Le patron exact observé dépendait de la variable linguistique utilisée comme variable modératrice et du groupe expérimental associé à la comparaison. Néanmoins, il ne fait pas de doute que ce sont les enfants à faible exposition qui profiteraient le plus du programme au chapitre du développement de leurs habiletés de communication si le programme était mis en œuvre à une plus grande échelle.

Les effets du « nouveau » programme sur l'échelle de Communication se sont manifestés le plus fortement, dans ces analyses, au troisième post-test qui a eu lieu au début de l'année scolaire, immédiatement après les vacances d'été. C'est pour cette évaluation que les variables linguistiques se sont manifestées de la façon la plus constante en tant que variables modératrices statistiquement significatives (p. ex., pour les scores de Communication). Nous nous attendrions à des effets accrus du programme pour cette évaluation si l'intervention avait réussi à augmenter, comparativement à la condition de comparaison, la quantité et la qualité de l'exposition au français pendant les mois d'été. Ceci serait le cas, par exemple, si les parents du groupe programme avaient modifié leurs comportements linguistiques pendant cette période (p. ex., langue parlée à l'enfant, langue du matériel d'alphabétisation) ou si les prestataires du groupe programme en garderie s'étaient davantage conformés à l'exigence que la langue française devrait invariablement être la langue d'interaction, même pendant l'instabilité du programme d'été. Les deux possibilités sont en accord avec les objectifs du programme et donc, leur effet combiné peut être à l'œuvre.

Il est digne de mention que les effets positifs du programme relativement au groupe témoin en garderie sur l'échelle de Communication ont été corroborés par des effets correspondants au troisième post-test avec la sous-échelle de Vocabulaire expressif (modérateurs : continuum linguistique parlé par l'enfant, langue des activités de littératie), l'autre résultante ciblant spécifiquement la communication. La principale différence entre les deux ensembles d'analyses tient au fait qu'avec la sous-échelle de Vocabulaire expressif, les effets au premier post-test n'ont pas varié en fonction du profil linguistique. Dans ce cas, tant les enfants à forte exposition que

---

<sup>95</sup> Un examen de la distribution de cette variable a démontré une asymétrie négative distincte au niveau de base (asymétrie=-1,19) qui s'est progressivement aggravée au cours de l'étude, asymétrie au troisième post-test=-2,73. À l'opposé, la distribution de la sous-échelle de Vocabulaire expressif était plutôt normale, même au troisième post-test (asymétrie=-0,66). Nous attribuons cette asymétrie négative à la facilité relative des items de Vocabulaire réceptif et au nombre limité d'items sur cette échelle (n=6). Cette propriété de l'échelle expliquerait son incapacité à discriminer les différents participants.

les enfants à faible exposition ont obtenu des gains développementaux au premier post-test. De façon générale, le rôle modérateur de l'exposition linguistique était plus faible et invariablement non significatif pour cette variable. Ces résultats sont faibles lorsqu'on les compare à ceux qui ont été obtenus avec l'échelle de Communication, mais nous attribuons ceci à un effet dont l'étendue est restreinte étant donné les données manquantes systématiques dues à l'administration de l'ÉPE-AD en anglais pour certains des enfants.

Tel que prévu, le continuum linguistique parlé par l'enfant et la langue des activités de littératie ont été les plus forts modérateurs de l'effet du programme relativement au groupe témoin en garderie. La première de ces variables représente l'utilisation active de la langue française par l'enfant; la deuxième de ces variables est aussi une variable « active », mais elle a l'avantage supplémentaire d'être associée des comportements concrets, une caractéristique reconnue pour entraîner des réponses plus précises. Les résultats suggèrent que le programme offre un avantage particulier aux enfants présentant un déficit quant à ces dimensions, lesquelles peuvent être interprétées comme indice du bilinguisme de l'enfant (p. ex., les scores les plus élevés sont obtenus par les enfants qui parlent seulement le français et les scores les plus bas sont obtenus par ceux qui parlent une autre langue seulement, habituellement l'anglais). En faisant la promotion du développement des compétences en langue française, le programme soutient le développement du bilinguisme dans une population francophone minoritaire (voir la discussion au chapitre 7). Pour sa part, la langue des activités de littératie capte une dimension qui était la cible directe du programme d'alphabétisation familiale; il n'est pas surprenant qu'elle soit associée aux différences observées entre les groupes expérimentaux.

Le continuum linguistique parlé par la mère à l'enfant était un modérateur particulièrement important pour les comparaisons relatives au groupe témoin hors garderie. Lorsque cette dimension était prise en considération, les effets du programme sur l'échelle de Communication ont été observés pour les premier, deuxième et troisième post-tests. Chose encore plus encourageante, les effets estimés du programme dans ces analyses se sont accrus avec le temps, pour éventuellement terminer avec un effet robuste d'un demi écart-type, douze mois après la mise en œuvre. La langue parlée par la mère peut être un modérateur important lors des comparaisons avec le groupe témoin hors garderie en raison de la contribution importante que les mères apportent à l'environnement familial.

Les enfants à forte exposition au français ont aussi profité du programme, mais d'une manière différente. Ils ont commencé le programme avec a) une plus forte exposition au français, b) une meilleure habileté à communiquer en français et, par conséquent, c) les compétences fondamentales de base nécessaires pour apprendre de nouvelles informations dans cette langue (Cummins, 1979). Bien que tous les enfants aient présenté des gains initiaux (premier post-test) pour le domaine cognitif de l'échelle ÉPE-AD, les enfants à forte exposition au français ont obtenu des gains pour les items les plus avancés de l'échelle, lesquels ont été décelés par le troisième post-test. Une fois leurs compétences de communication en français en place, les éléments cognitifs plus avancés du programme (p. ex. mathématiques et habiletés de littératie) étaient dans la zone de développement proximal de ces enfants (Vygotsky, 1978). Le mécanisme proposé explique naturellement la manière dont les effets du programme se sont différenciés en fonction de l'exposition linguistique au niveau de base. En bref, les résultats indiquent qu'au moment où la plupart des enfants de l'échantillon (52 %) sont entrés en prématernelle (c.-à-d. au troisième post-test), les enfants à forte exposition ont démontré pour le domaine cognitif de l'ÉPE-AD un effet significatif du programme comparativement aux deux groupes témoins.

Dans l'ensemble, le programme a aidé à la préparation à la scolarisation pour tous les enfants, mais ses effets se différencient selon le profil linguistique. Les résultats soutiennent l'affirmation que le programme a été efficace pour augmenter la préparation à la scolarisation des enfants qui sont moins exposés au français, en accélérant le développement de leur habileté à communiquer en français. Pour leur part, les enfants à exposition relativement élevée au français au début du projet semblaient tirer des avantages cognitifs du programme, principalement au moment de l'âge de quatre ans. Ces enfants étaient manifestement mieux équipés pour acquérir de telles compétences.

### ***Résultantes des parents***

Les analyses des résultantes pour les parents n'ont pu faire la preuve que les Ateliers familles influencent la fréquence et la langue des activités de littératie rapportées par les familles participantes. Nous avons observé certains signes d'un effet à la deuxième période post-test en ce qui a trait particulièrement à la langue des activités de littératie, mais cet effet n'a pas persisté. Nous attribuons l'effet nul sur la fréquence des activités de littératie au mauvais calibrage des choix de réponses offerts aux participants (p. ex. la plus grande fréquence disponible était de trois fois ou plus par jour, ce qui pourrait être considéré comme excessif). Cependant, nous avons obtenu des effets statistiquement significatifs avec d'autres mesures auto-rapportées. Les effets les plus crédibles observés ont trait aux gains déclarés par les parents quant à leurs connaissances du développement de l'enfant, des stratégies pour préparer leur enfant à l'école, et de la disponibilité des ressources francophones dans la communauté.

Lorsque nous avons comparé les participants et les non-participants aux Ateliers familles quant aux Connaissances déclarées, le profil observé ne démontrait aucun signe de biais dans les estimations pré-test et post-test. De plus, nous rapportons des effets où la qualité du programme Ateliers familles (pour certains aspects) et la participation des parents sont associés à des gains significativement plus importants quant aux Connaissances déclarées. En somme, toutes les analyses supportent la conclusion que le volet Ateliers familles de l'intervention a eu un impact réel sur cette résultante. Nous avons aussi observé des effets du programme statistiquement significatifs concernant l'Auto-efficacité déclarée. Cet effet a été confirmé par les analyses menées quant à la qualité du programme, mais non par les analyses sur le dosage. Aucun effet n'a été détecté en ce qui a trait à l'échelle de Modélisation, qui était conçue pour évaluer la fréquence à laquelle les parents modélisent les activités de littératie pour leur enfant (il s'agit ici de la fréquence de modélisation plutôt que de la fréquence où l'enfant est encouragé à prendre part/s'engager lui-même dans ces activités). Comme pour la mesure de la fréquence des activités de littératie, nous attribuons cet effet nul à un problème de l'échelle. Un problème supplémentaire de cette mesure et des questions d'opinion parentale ciblées étaient la contamination des mesures au pré-test par l'exposition au programme. En somme, les effets d'impact rapportés sont propres aux Connaissances et à l'Auto-efficacité déclarées, alors que les effets nuls observés peuvent s'expliquer par des problèmes méthodologiques.

## **6. Analyses d'impact : Suivi de la deuxième année**

Ce chapitre présente les analyses des résultantes développementales évaluées durant la deuxième année de l'étude. Les analyses des résultantes de la première année portaient sur l'impact immédiat du programme testé, alors que les analyses de la deuxième année visent plutôt l'impact résiduel et différé de l'exposition au programme.

À l'instar du chapitre 5, les résultats de ces analyses sont présentés en deux parties complémentaires. La première partie décrit les analyses par groupe expérimental (section 6.1.2). La seconde partie présente une série d'analyses secondaires en fonction de variables modératrices (Section 6.2), comprenant le dosage, la fidélité/qualité du programme, et le profil linguistique. Chaque série d'analyses porte sur six résultantes développementales: les quatre échelles de l'ÉPE-AD (standardisées avant les analyses) et les deux échelles normalisées de vocabulaire, soit l'EOWPVT et l'ÉVIP-R.

### **6.1 VARIABLES RETENUES POUR LES ANALYSES**

Cette section donne le détail des variables retenues pour les analyses des résultantes de la deuxième année. Nous présentons dans des sections séparées les variables utilisées comme variables indépendantes et celles employées comme résultantes.

#### **6.1.1 Prédicteurs et covariables significatifs**

L'environnement des enfants a beaucoup changé durant la deuxième année, ce qui a nécessité l'inclusion d'une autre covariable dans les spécifications du modèle de régression. Spécifiquement, au début de la deuxième année du projet, certains enfants sont entrés à l'école à temps plein, d'autres à temps partiel, alors que d'autres n'étaient pas encore inscrits à l'école. Une variable catégorielle représentant ces différents degrés d'exposition à l'école a été insérée comme covariable dans nos analyses de régression. Cette variable est codée comme 0 pour la période de base et comme 0 ou 1 à partir de la quatrième évaluation; les valeurs attribuées à chaque enfant sont statiques pour toute la deuxième année du projet, soit de la quatrième à la septième période d'évaluation.

Pour les analyses des effets du dosage, nous avons adapté les données de l'exposition à la garderie de la première année. Nous avons calculé des variables statiques à l'aide des données représentant l'exposition à la garderie durant la première année de l'étude pour les garderies du groupe programme et du groupe témoin. Ces variables représentent la moyenne d'heures passées en garderie par semaine pour la période s'étendant du début de l'intervention à la troisième évaluation. Les données sur l'exposition à la garderie durant les mois d'été n'ont pas servi à ce calcul en raison de l'instabilité a) des arrangements des services de garde pendant les mois d'été, et b) de l'animation du programme durant cette période. Les données de la deuxième année sur l'exposition à la garderie n'ont pas été incluses dans les analyses, car la majorité des enfants étaient inscrits à l'école à temps plein ou à temps partiel durant la deuxième année. Tel que discuté au chapitre 3, il n'aurait pas été pratique de faire autrement.

Les indices de fidélité et de qualité du programme étaient identiques à ceux qui ont été utilisés pour les analyses de la première année. Comme pour les heures passées en garderie, la fidélité et la qualité du nouveau programme préscolaire de service de garde pendant la deuxième année n'est plus une source importante de renseignements pour la plupart des enfants de l'échantillon de la deuxième année. Nous avons considéré plus pertinent d'utiliser les cotes de fidélité et de qualité de la première année comme descripteurs de chacune des garderies dans les analyses. Autrement dit, nous étions surtout intéressés à l'effet des programmes que les enfants ont reçus au cours de la première année du projet.

Des cotes de fidélité et de qualité ont été attribuées aux garderies pour les analyses de la deuxième année de la manière suivante. Les valeurs de fidélité associées à chaque garderie à partir de la quatrième évaluation ont été basées sur les observations recueillies au cours de l'intervalle entre la deuxième et la troisième évaluation. De même, les garderies se sont vu assigner les mêmes cotes de qualité que celles qui leur avaient été attribuées dans l'analyse des résultantes de la première année.

Pour l'analyse de la deuxième année, les enfants ont été assignés à des garderies spécifiques de la manière suivante. Si les enfants n'étaient pas inscrits à l'école ou y étaient inscrits seulement à temps partiel, la qualité et la fidélité du programme attribuées à leur garderie au moment de l'évaluation ont été utilisées. Si les enfants étaient inscrits à l'école à plein temps, la qualité et la fidélité du programme attribuées à leur ancienne garderie avant qu'ils soient inscrits à l'école ont été utilisées.<sup>96</sup> Les mêmes règles régissant les changements de groupes expérimentaux (p. ex. codage différé de changements survenus pendant l'été) étaient en vigueur pour les changements du statut d'inscription en garderie.

### **6.1.2 Mesure de résultantes**

En tout, six résultantes ont été prises en compte pour les analyses d'impact de la deuxième année : Communication, Conscience de soi, Habiletés cognitives, Vocabulaire expressif (sous-échelle dérivée de l'ÉPE-AD), EOWPVT (échelle bien validée de vocabulaire expressif), et ÉVIP-R (échelle bien validée de vocabulaire réceptif). Les deux dernières mesures ont respectivement été administrées durant la sixième et la septième évaluation. Comme pour les analyses des résultantes de la première année, les mesures de l'ÉPE-AD ont été standardisées durant la période de temps précédant les analyses (pour plus de détails, voir la section 5.1). Les données pour l'échelle d'Habilité physique et la sous-échelle de Vocabulaire réceptif de l'ÉPE-AD étaient disponibles pour les quatrième et cinquième évaluations, mais les analyses reposant sur ces mesures ne sont pas rapportées ici; les données de la quatrième évaluation ont été analysées et rapportées au chapitre 5. De plus, les résultats obtenus suite aux analyses des données d'Habilité physique et de Vocabulaire réceptif recueillies à la cinquième évaluation (non rapportées) convergent vers des estimations comparables à celles de la quatrième évaluation.

---

<sup>96</sup> Les analyses reposant sur l'inscription à la garderie à la fin de la première année de l'étude (3<sup>e</sup> et 4<sup>e</sup> évaluations) donne des résultats très similaires. On ne peut donc pas écarter les effets rapportés lorsque des changements dans la composition des groupes surviennent.<sup>97</sup> Nous avons également songé à utiliser les échelles de vocabulaire improvisées, créées à cette fin en fonction de l'ÉPE-AD. Toutefois, nous avons rejeté l'idée parce qu'il aurait fallu exclure les 26 cas qui ne répondaient pas au critère selon lequel il est nécessaire de compléter l'ÉPE-AD en français un nombre suffisant de fois durant la première année du projet. L'emploi de ces variables comme covariables aurait eu pour conséquence de limiter l'étendue des scores de vocabulaire en français, ce qui aurait fait baisser les estimations d'association.

La standardisation des mesures de résultantes a aussi été requise pour les analyses de la deuxième année en raison des divergences dans la mise à l'échelle des variables notées à la section 5.1.1. Même si ce n'était pas strictement nécessaire, nous avons adopté la même stratégie de transformation qu'avec les données de la première année. Cette décision a été prise pour maintenir la cohérence dans la présentation des résultats quant aux résultantes observées chez les enfants. Un avantage de la standardisation des mesures est qu'elle nous permet de comparer de manière significative la taille relative des effets du programme à travers les mesures de résultantes.

### **6.1.3 Résultats de la comparaison des groupes**

Les résultats des analyses d'impact de la deuxième année sont rapportés dans cette section. Nous présentons d'abord la logique sous-tendant les analyses effectuées ainsi que les détails de spécification associés à nos modèles de régression. Les résultats des analyses en fonction des groupes expérimentaux sont ensuite rapportés. Pour assurer une cohérence, les résultats sont présentés selon la même stratégie qu'à la section 5.1.2. Les seuls ajouts à la liste des résultantes sont les échelles de vocabulaire de l'EOWPVT et de l'ÉVIP-R. Nous rapportons les analyses fondées sur ces résultats séparément de celles qui reposent sur les résultantes de l'ÉPE-AD. À l'instar de toutes les analyses précédentes, les résultats apparaissant ci-dessous sont basés sur la version française de l'ÉPE-AD. Cette stratégie générale de présentation est également employée à la section 6.2.

#### ***Détails de spécification du modèle***

La stratégie de spécification des modèles employée dans les analyses de la deuxième année est presque identique à celle décrite à la section 5.1.2. Nous avons employé un estimateur DD de l'effet du traitement conjointement avec des covariables dont l'inclusion était justifiée par le fait qu'elles permettaient un ajustement pour la variabilité dans la composition des groupes observée dans le temps. Nous avons ajouté à la liste originale de covariables une série de variables fictives représentant l'inscription à l'école, qui figuraient dans la spécification du modèle de base (modèle 1). Autrement, la stratégie d'analyse est demeurée la même qu'à la section 5.1.2. Toutefois, les évaluations prises en considération dans les présentes analyses étaient bien sûr différentes.

Pour les résultats rapportés ci-dessous, la période d'évaluation de base a servi de référence pour évaluer les effets programme, du troisième post-test (automne 2008) au sixième post-test (automne 2009). Ces périodes post-test correspondent aux quatrième et septième évaluations, respectivement. Nous avons exclu les deuxième et troisième évaluations de ces analyses, car leur inclusion nous aurait empêchés de tester l'interaction entre le groupe expérimental et le temps à l'aide du test F de Wald en raison, d'un trop petit nombre de degrés de liberté. Cette exclusion n'affecte pas la validité des autres estimations DD.

La principale caractéristique qui distingue les analyses de la deuxième année est le recours à une spécification alternative du modèle de régression pour analyser l'EOWPVT et l'ÉVIP-R. Au contraire des échelles de l'ÉPE-AD, une seule observation est disponible pour chacune des échelles de vocabulaire. L'EOWPVT a été administré pendant la sixième évaluation et l'ÉVIP-R pendant la septième évaluation. Aucune observation au niveau de base n'étant disponible pour ces mesures, nous avons utilisé un estimateur ANCOVA pour obtenir l'estimation de l'effet

traitement. À cette fin, les scores de Communication standardisés de l'ÉPE-AD (évaluation au niveau de base) ont servi de covariables dans la spécification du modèle de base.<sup>97</sup> Cette mesure au niveau de base est une mesure pré-test adéquate étant donné qu'elle est corrélée avec l'ÉVIP-R et l'EOWPVT à 0,75 et 0,79 (Pearson), respectivement. Les estimations de l'effet du programme qui résultent de ces analyses sont interprétées comme la différence attendue entre les groupes pour les mesures de vocabulaire si les participants de chaque groupe se sont engagés dans le projet avec des scores de Communication équivalents.<sup>98</sup> Parce que les changements dans la composition des groupes au fil du temps demeurent problématiques dans ces analyses, nous estimons les effets traitement selon l'appartenance au groupe à plusieurs périodes de temps. Pour chaque période de temps, les enfants ont été regroupés de façon légèrement différente selon les arrangements de service de garde. Si ces changements dans la composition des groupes ne sont pas importants en ce qui concerne les résultats, les résultats obtenus devraient être essentiellement les mêmes, peu importe la façon dont la question est abordée dans les analyses.<sup>99</sup>

### **RÉSULTATS D'IMPACT: Échelles standardisées de l'ÉPE-AD (en français)**

Les résultats des analyses d'impact pour les échelles de l'ÉPE-AD sont présentés dans cette section : au tableau 6.1 pour les analyses des scores de Communication menées auprès de l'échantillon total, et au tableau 6.2 pour les autres échelles de l'ÉPE-AD. Les résultats des analyses de l'échantillon total sont rapportés pour chaque étape de l'inclusion séquentielle des covariables (modèles 1 à 5). Pour les autres échelles, seuls sont rapportés les résultats du modèle final, incluant la série complète de covariables.

Les résultats révèlent un patron constant à travers toutes les résultantes de l'ÉPE-AD. Un examen des deux tableaux montre que les effets pour toutes les variables émergent dans l'une des deux, ou dans les deux évaluations finales, lorsque les estimations sont ajustées pour les covariables. Ces évaluations correspondent au moment où la version originale de l'ÉPE-AD a été changée pour une version révisée qui excluait les items les plus faciles et incluait d'autres items conçus pour être plus difficiles. En d'autres mots, nous commençons à détecter des effets programme statistiquement significatifs quand les mesures employées deviennent plus sensibles aux différences individuelles entre les enfants présentant un rendement modéré à supérieur. Autrement dit, nous observons les effets traitement attendus relativement au groupe témoin en garderie lorsque les mesures utilisées sont mieux calibrées pour l'échantillon d'enfants évalué. L'ampleur des effets va d'environ 0,20 écart-type pour l'échelle de Communication à environ 0,45 écart-type pour le score de Vocabulaire expressif. Les effets observés se situent donc, selon les points de référence conventionnels, dans une étendue d'effet dit petit à modéré. On n'observe aucun effet programme relativement au groupe témoin hors garderie, ce qui est compatible avec les conclusions rapportées à la section 5.1.2.

---

<sup>97</sup> Nous avons également songé à utiliser les échelles de vocabulaire improvisées, créées à cette fin en fonction de l'ÉPE-AD. Toutefois, nous avons rejeté l'idée parce qu'il aurait fallu exclure les 26 cas qui ne répondaient pas au critère selon lequel il est nécessaire de compléter l'ÉPE-AD en français un nombre suffisant de fois durant la première année du projet. L'emploi de ces variables comme covariables aurait eu pour conséquence de limiter l'étendue des scores de vocabulaire en français, ce qui aurait fait baisser les estimations d'association.

<sup>98</sup> Autrement dit, les scores de Communication au niveau de base sont traités comme toute autre caractéristique statique de base incluse dans le modèle de régression. On les a inclus dans le modèle de régression de base parce qu'ils servent à identifier l'effet traitement.

<sup>99</sup> En interprétant les résultats présentés ci-dessous, il est à noter que les changements les plus importants dans la composition des groupes (dus aux passages d'un groupe à l'autre) ont été observés pour la 5e période d'évaluation (février 2009), voir le tableau 2.2. Les effets traitement observés n'apparaissent toutefois qu'aux 6e et 7e périodes d'évaluation.



**Tableau 6.1 : Deuxième année – Différence des différences (DD) : Effets programme pour les scores de Communication standardisés de l'ÉPE-AD (version française, échantillon total)**

	<i>Inclusion séquentielle des covariables – Erreurs groupées par garderie</i>									
	(1)		(2)		(3)		(4)		(5)	
Effets programme	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
Groupe programme en garderie v. groupe témoin en garderie, au niveau de base (G1 <sup>a</sup> v. G2)	,205	,184	,181	,181	,203	,124	,244	,121	,204	,121
DD 3 <sup>e</sup> post-test	-,121	,247	-,084	,251	-,073	,204	-,069	,208	-,087	,202
DD 4 <sup>e</sup> post-test	,091	,165	-,019	,135	-,038	,103	-,033	,110	-,039	,108
DD 5 <sup>e</sup> post-test	-,013	,188	-,077	,146	-,121	,128	-,112	,129	-,129	,126
DD 6 <sup>e</sup> post-test	-,205	,230	-,246	,163	-,286*	,134	-,266*	,133	-,282*	,132
Groupe programme en garderie v. groupe témoin hors garderie, au niveau de base (G1 v. G3)	-,065	,180	-,205	,168	-,153	,092	-,093	,085	-,151	,090
DD 3 <sup>e</sup> post-test	-,049	,361	-,013	,324	-,024	,249	-,026	,255	-,070	,259
DD 4 <sup>e</sup> post-test	-,052	,380	,015	,280	-,007	,208	,019	,204	-,009	,221
DD 5 <sup>e</sup> post-test	,059	,192	,193	,150	,173	,111	,178	,111	,138	,115
DD 6 <sup>e</sup> post-test	-,140	,250	,013	,188	-,006	,151	,063	,123	,048	,125
Test F composite de Wald	F (8, 16) = ,836		F (8, 16) = 1,32		F (8, 16) = 1,27		F (8, 16) = 1,24		F (8, 16) = 1,19	
Nombre de participants	239		234		233		233		226	

*Note* : La résultante a été standardisée, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type. **Le groupe programme en garderie (G1) est mis en référence, ce qui signifie que les valeurs négatives des estimations DD indiquent un effet programme positif.** Les erreurs-types ont été estimées grâce au programme SPSS pour le calcul de l'estimateur robuste à l'hétérogénéité de White (White, 1980). De plus, les erreurs-types ont été groupées par garderie afin d'ajuster pour les corrélations entre les résidus dans le temps. L'effet observé est statistiquement significatif seulement une fois que les caractéristiques sociodémographiques sont insérées dans les analyses. Les détails de spécification du modèle sont présentés à la section 6.1.3.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux).

<sup>a</sup> G1, G2 et G3 dénotent le groupe programme en garderie, le groupe témoin en garderie et le groupe témoin hors garderie, respectivement.

**Tableau 6.2 : Deuxième année – Différence des différences (DD) : Effets programme pour les scores standardisés des sous-échelles de l'ÉPE-AD (participants ayant répondu au test en français seulement)**

Effets programme	Échelles de l'ÉPE-AD, toutes covariables incluses - erreurs groupées par garderie							
	Communication		Conscience de soi		Habilités cognitives		Vocabulaire expressif	
	DD	ET robuste	DD	ET robuste	DD	ET robuste	DD	ET robuste
Groupe programme en garderie v. groupe témoin en garderie, au niveau de base (G1 <sup>a</sup> v. G2)	,209 <sup>m</sup>	,113	,198	,168	,259	,141	,347	,175
DD 3 <sup>e</sup> post-test	-,064	,193	-,084	,175	-,042	,219	-,148	,159
DD 4 <sup>e</sup> post-test	-,044	,105	-,111	,126	-,100	,191	-,209	,153
DD 5 <sup>e</sup> post-test	-,118	,121	-,359*	,134	-,255*	,099	-,479*	,147
DD 6 <sup>e</sup> post-test	-,248*	,118	-,294	,173	-,327*	,105	-,461*	,199
Groupe programme en garderie v. groupe témoin hors garderie au niveau de base (G1 v. G3)	-,095	,083	-,082	,135	,185	,138	-,033	,135
DD 3 <sup>e</sup> post-test	,050	,253	,041	,118	-,174	,229	,074	,141
DD 4 <sup>e</sup> post-test	,091	,201	,160 <sup>m</sup>	,088	-,033	,170	,155	,201
DD 5 <sup>e</sup> post-test	,106	,139	-,134	,135	-,250 <sup>m</sup>	,121	-,102	,228
DD 6 <sup>e</sup> post-test	,125	,102	-,196	,142	-,222	,138	-,169	,168
Test F composite de Wald	F (8, 16) = 1,06		F (8, 16) = 4,49 <sup>†</sup>		F (8, 16) = 3,69*		F (8, 16) = 2,50 <sup>m</sup>	
Nombre de participants	209		208		209		208	

**Note :** Les résultantes ont été standardisées, c'est-à-dire que les effets DD peuvent être interprétés en unités d'écart-type. Les analyses sont basées sur un sous-échantillon de participants qui ont complété l'ÉPE-AD en français, au moins deux fois pendant la première année de l'étude. Ainsi, ce sous-échantillon est biaisé vers les enfants qui sont plus forts en français. Le Vocabulaire expressif était la seule échelle à démontrer un effet positif avant l'inclusion des covariables. **Le groupe programme en garderie (G1) est le groupe de référence, ce qui signifie que les valeurs négatives des estimations DD indiquent un effet programme positif.** Les détails de spécification du modèle sont présentés à la section 6.1.3.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux).

<sup>a</sup> G1, G2 et G3 dénotent le groupe programme en garderie, le groupe témoin en garderie et le groupe témoin hors garderie, respectivement.

### ***Échelles de vocabulaire de l'EOWPVT et l'ÉVIP-R (non normalisées)***

Les effets programme estimés pour les résultantes de l'EOWPVT et de l'ÉVIP-R sont rapportés dans cette section. Les effets du programme ont été réévalués à plusieurs reprises sur la base d'assignations différentes aux groupes expérimentaux, reflétant l'appartenance aux groupes expérimentaux pour les évaluations quatre à sept inclusivement. Ces mesures de vocabulaire ont été prises à un point unique dans le temps. Toutefois, la répétition des analyses à l'aide de différentes versions de la variable groupe expérimental, fondée sur l'appartenance antérieure aux groupes, permet d'évaluer si les résultats observés dépendent des changements de groupe de façon critique. Si cela est le cas, alors la validité des résultats peut être remise en question. Les résultats obtenus suite aux analyses de régression, pour l'échelle EOWPVT, avec inclusion séquentielle de covariables sont rapportés au tableau 6.3. Les résultats provenant des analyses de régression, pour l'échelle ÉVIP-R, avec inclusion séquentielle de covariables apparaissent au tableau 6.4.

Les résultats des deux séries d'analyses révèlent un effet programme relatif au groupe témoin en garderie de l'ordre de deux à trois mots. L'effet cesse d'être statistiquement significatif lorsque les variables de composition de la famille sont insérées comme covariables (voir les résultats pour la quatrième et la cinquième spécification). Malgré cette baisse de signification statistique, la différence entre les groupes est toujours estimée à environ deux mots.<sup>100</sup> Autrement dit, l'inclusion des covariables a entraîné une légère diminution de l'ampleur de l'effet programme estimé, qui a tout juste suffi à éliminer l'effet.

Les analyses de l'ÉVIP-R se sont distinguées en montrant les signes d'un effet programme relatif au groupe témoin hors garderie, avant que toutes les covariables ne soient insérées dans les analyses. Dans ce cas, peu d'indices d'un effet programme sont demeurés présents quand toutes les covariables ont été insérées dans le modèle (voir la cinquième spécification). Cette dépendance à l'inclusion des covariables est compatible avec le fait que le groupe témoin hors garderie s'est révélé substantiellement différent du groupe programme en garderie quant à plusieurs caractéristiques, dans les analyses primaires présentées à la section 4.4. Dans ces circonstances, il n'est pas surprenant que l'inclusion de ces caractéristiques dans le modèle produise un changement important de l'effet estimé.

En résumé, les données laissent entrevoir un effet programme relatif au groupe témoin en garderie (EOWPVT, ÉVIP-R) et au groupe témoin hors garderie (EOWPVT), quand toutes les covariables sont incluses dans les analyses. Nous ne savons pas encore si ces tendances se manifesteront comme un effet significatif lorsque nous adopterons une approche analytique plus fine basée sur les indices d'exposition au traitement et de fidélité/qualité du programme issus de la première année. De façon réaliste, l'effet en question est de l'ordre de seulement deux à trois mots. Même si l'effet était statistiquement significatif, il serait difficile de faire valoir qu'il

---

<sup>100</sup> Il est utile de rappeler que ces deux mots représentent une différence du vocabulaire total qui est bien plus grande en termes du nombre réel de mots connus par l'enfant. Nous n'avons pas de règle de « conversion » qui permettrait aux scores de l'ÉVIP-R d'être exprimés avec précision en fonction de la taille du vocabulaire total. Néanmoins, l'effet de deux mots ne doit pas être interprété comme s'il indiquait une croissance du vocabulaire réel de seulement deux mots.<sup>101</sup> Les enfants ont été fixés à leur garderie de la première année de sorte que l'interaction entre le dosage (nombre moyen d'heures passées en garderie) et la variable d'appartenance au groupe expérimental soit interprétable. Il ne serait pas désirable, par exemple, de traiter incorrectement les heures passées dans l'une des garderies du programme comme si elles avaient été passées dans une garderie de comparaison, à cause d'un changement de groupe survenu dans la deuxième année du projet.

représente un impact important de l'intervention et ce, même si nous reconnaissons qu'il représente un gain réel dans le vocabulaire de l'enfant de beaucoup plus de trois mots.

**Tableau 6.3 : Scores de l'EOWPVT (recueillis à la sixième évaluation, juin 2009) en fonction de l'appartenance au groupe aux quatre périodes d'évaluation de la deuxième année (non normalisés)**

		<i>Inclusion séquentielle des covariables – Erreurs-types robustes groupées par garderie</i>									
		(1)		(2)		(3)		(4)		(5)	
		Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET
<b>G1 v. G2</b>	Groupes à la 4 <sup>e</sup> évaluation	-3,95*	1,39	-3,42*	1,49	-3,04*	1,43	-1,85	1,60	-1,92	1,52
	Groupes à la 5 <sup>e</sup> évaluation	-2,90 <sup>m</sup>	1,42	-3,48*	1,67	-3,24 <sup>m</sup>	1,63	-2,68	1,74	-2,84	1,74
	Groupes à la 6 <sup>e</sup> évaluation	-3,21*	1,49	-3,48*	1,66	-3,49 <sup>m</sup>	1,70	-2,69	1,85	-2,99	1,87
	Groupes à la 7 <sup>e</sup> évaluation	-	-	-	-	-	-	-	-	-	-
<b>G1 v. G3</b>	Groupes à la 4 <sup>e</sup> évaluation	-,821	1,10	-1,31	1,45	-,722	1,39	,08	1,52	-,36	1,44
	Groupes à la 5 <sup>e</sup> évaluation	-2,28	2,15	-2,28	2,15	-1,88	2,12	-1,19	2,33	-1,87	2,29
	Groupes à la 6 <sup>e</sup> évaluation	-3,06	2,20	-2,73	2,12	-2,55	2,20	-1,66	2,41	-2,42	2,35
	Groupes à la 7 <sup>e</sup> évaluation	-	-	-	-	-	-	-	-	-	-
	Nombre de participants	236		231		230		229		223	

**Note :** Les scores de l'EOWPVT dénotent le nombre de mots correctement identifiés et n'ont été recueillis qu'une seule fois (sixième évaluation). Les tests de l'effet programme ont été effectués en employant quatre alternatives de groupement : appartenance aux groupes expérimentaux à la quatrième, cinquième, sixième et septième évaluation. Les résultats obtenus avec les groupements de la septième évaluation ne sont pas rapportés parce que cette évaluation a eu lieu après l'administration de l'EOWPVT. Le but d'analyser ces tests en grande partie redondants était d'évaluer dans quelle mesure la migration des participants d'un groupe à l'autre peut avoir influencé les résultats. **Le groupe programme en garderie (G1) est mis en référence, ce qui signifie que les valeurs négatives des estimations DD indiquent un effet programme positif.** Les erreurs-types ont été estimées grâce au programme SPSS pour le calcul de l'estimateur robuste à l'hétérogénéité de White (White, 1980). De plus, elles ont été groupées par garderie afin d'ajuster pour l'effet commun du milieu de garderie. Les détails de spécification du modèle sont présentés à la section 6.1.3.

G1, G2 et G3 dénotent le groupe programme en garderie, le groupe témoin en garderie et le groupe témoin hors garderie, respectivement.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux, 20 degrés de liberté).

**Tableau 6.4 : Scores de l'ÉVIP-R (recueillis à la septième évaluation, octobre 2009) en fonction de l'appartenance au groupe aux quatre périodes d'évaluation de la deuxième année (non normalisés)**

		<i>Inclusion séquentielle des covariables – Erreurs-types robustes groupées par garderie</i>									
		(1)		(2)		(3)		(4)		(5)	
		Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET
<b>G1 v. G2</b>	Groupes à la 4 <sup>e</sup> évaluation	-4,89*	1,74	-4,46 <sup>†</sup>	1,43	-3,25*	1,41	-2,29	1,57	-2,45	1,51
	Groupes à la 5 <sup>e</sup> évaluation	-2,25	1,78	-3,18*	1,46	-2,21	1,35	-1,93	1,55	-1,75	1,55
	Groupes à la 6 <sup>e</sup> évaluation	-2,66 <sup>m</sup>	1,51	-3,26*	1,38	-2,69 <sup>m</sup>	1,40	-2,32	1,56	-2,37	1,40
	Groupes à la 7 <sup>e</sup> évaluation	-2,66 <sup>m</sup>	1,51	-3,34*	1,36	-2,75 <sup>m</sup>	1,36	-2,21	1,71	-2,21	1,48
<b>G1 v. G3</b>	Groupes à la 4 <sup>e</sup> évaluation	-1,32	1,73	-1,86	1,17	-,691	1,14	,306	1,26	,804	1,35
	Groupes à la 5 <sup>e</sup> évaluation	-2,71	2,22	-3,31*	1,42	-2,12	1,27	-,58	1,28	-,215	1,53
	Groupes à la 6 <sup>e</sup> évaluation	-3,35	2,04	-3,44*	1,25	-2,58*	1,23	-1,17	1,28	-1,01	1,34
	Groupes à la 7 <sup>e</sup> évaluation	-3,35	2,04	-3,43*	1,23	-2,60*	1,20	-,912	1,35	-,725	1,37
	Nombre de participants	236		232		231		231		224	

**Note :** Les scores de l'ÉVIP-R dénotent le nombre de mots correctement identifiés et n'ont été recueillis qu'une seule fois (septième évaluation). Les tests ont été effectués en employant quatre alternatives de groupements : l'appartenance aux groupes expérimentaux de la quatrième à la septième évaluation inclusivement. Le but d'analyser ces tests en grande partie redondants était d'évaluer dans quelle mesure la migration des participants d'un groupe à l'autre peut avoir influencé les résultats. **Le groupe programme en garderie (G1) est mis en référence, ce qui signifie que les valeurs négatives des estimations DD indiquent un effet programme positif.** Les erreurs-types ont été estimées grâce au programme SPSS pour le calcul de l'estimateur robuste à l'hétérogénéité de White (White, 1980). De plus, elles ont été groupées par garderie afin d'ajuster pour l'effet commun du milieu de garderie. Les détails de spécification du modèle sont présentés à la section 6.1.2.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , <sup>†</sup>  $p < 0,01$  (tous les tests bilatéraux, 20 degrés de liberté).

## 6.2 MODÉRATEURS ET MÉDIATEURS DE L'EFFET D'INTERVENTION

Les résultats que nous présentons dans cette section visent à corroborer les résultats rapportés précédemment avec une série d'analyses plus précises. Nous examinons ici si le degré d'exposition à la garderie fait une différence au-delà de l'appartenance de la première année aux groupes expérimentaux et si l'intensité de l'exposition aux dimensions ciblées par le nouveau programme préscolaire de service de garde, tel que répertoriée par les indices de qualité et de fidélité du programme, est liée aux résultantes de la deuxième année. Nous présentons ci-dessous les résultats de ces analyses.

### 6.2.1 Définitions précises du dosage: Moyenne d'heures passées en garderie durant la première année

Il est intéressant, à ce stade, de se demander si le degré d'exposition durant la première année affecte les résultantes de la deuxième année au-delà de ce qui peut être expliqué par l'appartenance aux groupes expérimentaux. Cette question pourrait également être formulée ainsi: L'effet du dosage varie-t-il en fonction du groupe expérimental pour les analyses des résultantes de la deuxième année ? Dans les paragraphes qui suivent, nous rapportons les résultats des analyses reposant sur la série complète de covariables (modèle 5). Ces résultats seront présentés dans des sections séparées pour les résultantes de l'ÉPE-AD et pour les deux mesures de vocabulaire normalisées (c.-à-d. l'EOWPVT et l'ÉVIP-R). Les résultats des analyses du dosage de la première année, basées sur la série complète de covariables (modèle 5), sont présentés au tableau 6.5 pour les échelles standardisées de l'ÉPE-AD et au tableau 6.6 pour les mesures de vocabulaire normalisées.

#### ***Détails de spécification***

Comme dans les analyses du dosage présentées au chapitre 5, l'effet du dosage est, dans le cas présent, estimé à l'aide d'un test d'interaction triple: indicateurs de groupes expérimentaux x dosage x temps (DDD). Les analyses ont porté sur les enfants dont l'exposition était en moyenne d'au moins 10 heures par semaine. Les autres enfants ont été exclus à titre de cas aberrants dans la distribution. Les analyses rapportées sont donc uniquement basées sur des enfants qui ont reçu une exposition minimale à la garderie (pour un exemple d'une approche plus complexe de l'examen de l'effet de l'exposition sur *les sujets traités seulement*, avec une taille d'échantillon beaucoup plus importante, voir Behrman, Cheng et Todd, 2004). En résumé, les présentes analyses ont nécessité une approche différente de celle qui a été rapportée pour les données relatives au dosage de la première année (voir la Section 5.1.3). La liste de covariables était identique à celles utilisées pour les analyses par groupes expérimentaux présentées à la section précédente.

#### ***Effets de l'exposition de la première année sur les résultantes standardisées de l'ÉPE-AD à la deuxième année***

Les effets du dosage de la première année, après ajustement pour la série complète de covariables (modèle 5), sont illustrés au tableau 6.5 pour les mesures standardisées de l'ÉPE-AD. Au contraire de la section 5.1.2, les résultats pour l'échelle de Communication basés sur

l'échantillon total de participants ne sont pas rapportés, car ils ne diffèrent pas de façon marquée de ceux rapportés ici.

En premier lieu, on a observé peu de signes d'effets DDD du dosage différenciés selon le groupe expérimental (non rapportés), ce qui est confirmé par la non-signification des valeurs F de Wald rapportées au tableau 6.5. Ce résultat nul peut être attribué au nombre relativement faible de participants dans chaque groupe ( $n < 100$ ). L'addition de la seconde cohorte à l'ensemble des données pourrait augmenter la stabilité des estimations associées et, par voie de conséquence, la probabilité d'observer un impact.

En deuxième lieu, nous observons ici des effets du groupe expérimental plus importants que dans les analyses rapportées au tableau 6.2. Les premières analyses permettaient certains changements de groupe, alors que les analyses présentées au tableau 6.5 fixent tous les participants dans la garderie à laquelle ils étaient inscrits à la première année.<sup>101</sup> Il est à noter que l'on observe un effet DD significatif pour la mesure de Vocabulaire expressif (ÉPE-AD) pour les quatrième, cinquième et sixième post-tests, alors que dans les analyses qui permettaient des changements de groupe, l'effet pour le quatrième post-test était non significatif. Cette divergence peut s'expliquer par le fait que le plus important changement entre les groupes expérimentaux est survenu avant la quatrième période post-test (voir la section 2.4). Nous concluons que les changements de groupe ont annulé cet effet DD dans les analyses précédentes. Les résultats de ces analyses confirment et renforcent ceux présentés au tableau 6.2.

---

<sup>101</sup> Les enfants ont été fixés à leur garderie de la première année de sorte que l'interaction entre le dosage (nombre moyen d'heures passées en garderie) et la variable d'appartenance au groupe expérimental soit interprétable. Il ne serait pas désirable, par exemple, de traiter incorrectement les heures passées dans l'une des garderies du programme comme si elles avaient été passées dans une garderie de comparaison, à cause d'un changement de groupe survenu dans la deuxième année du projet.



**Tableau 6.5 : Groupes expérimentaux et dosage de la première année comme prédicteurs des résultantes de la deuxième année (participants ayant répondu au test en français seulement)**

	<i>Échelles ÉPE-AD, toutes covariables incluses, erreurs groupées par garderie</i>							
	Communication		Conscience de soi		Habilités cognitives		Vocabulaire expressif	
Effets du groupe expérimental de la première année (G1 v. G2)	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
Au niveau de base	,197*	,086	,211 <sup>m</sup>	,120	,281*	,107	,365*	,126
DD au 3 <sup>e</sup> post-test	-,070	,164	-,085	,155	-,068	,177	-,195	,134
DD au 4 <sup>e</sup> post-test	-,144	,115	-,161	,123	-,060	,138	-,298*	,137
DD au 5 <sup>e</sup> post-test	-,209	,124	-,395*	,125	-,257*	,105	-,585*	,136
DD au 6 <sup>e</sup> post-test	-,272*	,129	-,302 <sup>m</sup>	,152	-,348*	,092	-,514*	,183
Interaction double : groupe x temps - DDD (F de Wald)	F (4, 17) = 1,52		F (4, 17) = ,622		F (4, 17) = 1,99		F (4, 17) = 5,01 <sup>†</sup>	
Nombre de participants	148		148		148		148	
Test d'interaction triple de l'effet du dosage - DDD (F de Wald)	F (4, 17) = 1,70		F (4, 17) = ,622		F (4, 17) = 1,99		F (4, 17) = ,339	

*Note : Le groupe programme en garderie est mis en référence, ce qui signifie que les valeurs négatives des estimations DD indiquent un effet programme positif.* On a observé des différences entre les groupes lors du pré-test, qui ont été neutralisées par l'estimateur DD. G1 représente le groupe programme en garderie et G2, le groupe témoin en garderie. Les intervalles de confiance de tous les effets G1 et G2 se chevauchent, mais le dosage G1 est un prédicteur significatif, alors que le dosage G2 ne l'est pas. Les enfants exposés à la garderie en moyenne moins de 10 heures par semaine durant la première année ont été écartés des analyses (pour un exemple des effets du dosage évalués en ne tenant compte que des sujets traités, voir Behrman, Cheng et Todd, 2004). Les détails de spécification sont fournis dans le corps du texte.

G1 et G2 dénotent le groupe programme en garderie et le groupe témoin en garderie, respectivement.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux).

### ***Effets d'exposition de la première année sur les échelles de vocabulaire de l'EOWPVT et de l'ÉVIP-R de la deuxième année (non normalisées)***

Les effets du groupe expérimental et du dosage de la première année sur les échelles de vocabulaire de l'EOWPVT et de l'ÉVIP-R sont présentés au tableau 6.6. Dans ce tableau, nous rapportons trois tests: a) l'effet des groupes expérimentaux (c.-à-d. la différence entre le groupe programme et le groupe témoin); b) les effets du dosage (une estimation propre au G1, le groupe programme en garderie); et c) l'interaction entre le dosage et le groupe (DD), qui est la différence dans l'effet du dosage pour le groupe témoin en garderie (c.-à-d. G2) par rapport à G1. Si le premier est significatif, cela veut dire qu'on observe une différence statistiquement significative entre les deux groupes. Si le deuxième est significatif, cela veut dire que l'on a observé un effet statistiquement significatif de l'exposition à la garderie. Si le dernier test est significatif, cela veut dire que l'estimation de l'effet du dosage pour G1 est statistiquement différente de celle de l'effet du dosage pour G2. À l'inverse, si ce dernier test n'est pas significatif, cela veut dire que l'effet du dosage est statistiquement équivalent pour les deux groupes. Voyons d'abord les résultats de l'ÉVIP-R avant d'aborder ceux de l'EOWPVT.

Notons que les effets du groupe expérimental de la première année, du dosage de la première année et de l'interaction entre le dosage et le groupe ne sont pas statistiquement significatifs pour l'ÉVIP-R, pour aucune spécification. Ces résultats sont compatibles avec ceux rapportés au tableau 6.4. Une tendance vers un petit effet est présente dans le modèle 1, mais disparaît complètement une fois rendu au modèle 5. En somme, les deux séries d'analyses en fonction des groupes expérimentaux s'entendent: s'il y a un impact de traitement sur cette mesure de vocabulaire normalisée, celui-ci est subtil. Si un vrai pré-test avait été disponible, nous aurions pu utiliser une estimation de l'effet du traitement basée sur l'estimateur DD, ce qui nous aurait permis de mieux contrôler les caractéristiques des participants au niveau de base et de tester l'effet du programme avec plus de puissance.

De la même manière, l'EOWPVT démontre un effet nul du groupe expérimental de la première année. Dans ce cas, toutefois, un effet du dosage en garderie émerge, mais cet effet est dans la direction opposée à celle attendue. L'estimation de l'effet du dosage de G1 est : a) significativement négative pour les modèles 4 et 5; et b) statistiquement équivalente à l'effet du dosage de G2. En résumé, nous ne relevons ici aucun signe d'un impact du programme. Comme il n'y avait aucun indice de cet effet au niveau de base, sa crédibilité est entièrement tributaire d'un modèle correctement spécifié (Maris, 1998; Glymour et coll., 2005). L'examen des résidus n'a révélé aucun problème évident, mais étant donné la petite taille d'échantillon pour chacun des groupes ( $n < 100$ ), nous avons décidé de ne pas donner trop de signification à la direction de cet effet. L'ajout des données de la seconde cohorte dans les analyses devrait rendre les estimations plus stables. Des spécifications alternatives du modèle, pouvant jeter plus de lumière sur la question, sont explorées dans la section 6.2.3, qui présente des analyses basées sur le profil linguistique des participants.

**Tableau 6.6 : Effets de l'exposition à la garderie durant la première année en fonction du type de garderie – Scores de l'ÉVIP-R et de l'EOWPVT (non normalisés)**

	<i>Inclusion séquentielle des covariables – Erreurs groupées par garderie</i>									
	(1)		(2)		(3)		(4)		(5)	
Effets programme	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
<b>Scores de l'ÉVIP-R</b>										
Groupe expérimental, 1 <sup>e</sup> année	-2,39	1,74	-2,09	1,64	-1,33	1,96	-,691	2,22	-,851	2,11
Effet du dosage (G1)	,205	,161	,138	,176	-,071	,192	-,121	,208	-,205	,225
Dosage x groupe (DD) G1 v. G2	-,543 <sup>m</sup>	,259	-,647	,293	-,398	,232	-,319	,244	-,223	,256
<b>Scores de l'EOWPVT</b>										
Groupe expérimental, 1 <sup>ère</sup> année	-2,46 <sup>m</sup>	1,27	-2,16 <sup>m</sup>	1,15	-1,77	1,35	-,828	1,44	-1,15	1,58
Effet du dosage (G1)	-,004	,177	-,123	,136	-,279	,132	-,308 <sup>*c</sup>	,119	-,344 <sup>*c</sup>	,135
Dosage x groupe (DD) G1 v. G2	-,137	,206	-,131	,183	,011	,159	,055	,160	,090	,176
Nombre de participants	160		156		155		154		152	

**Note :** Les variables de dosage de la première année ont été centrées à la moyenne générale avant les analyses. Les erreurs-types ont été estimées grâce au programme SPSS pour le calcul de l'estimateur robuste à l'hétérogénéité de White (White, 1980). De plus, elles ont été groupées par garderie afin d'ajuster pour l'effet commun du milieu de garderie. Les enfants exposés à la garderie en moyenne moins de 10 heures par semaine durant la première année ont été écartés des analyses (pour un exemple des effets du dosage évalués en ne tenant compte que des sujets traités, voir Behrman, Cheng et Todd, 2004). Les détails de spécification sont fournis dans le corps du texte.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux, 20 degrés de liberté).

<sup>a</sup> G1 et G2 dénotent les groupes programme en garderie et les groupes témoins en garderie, respectivement

<sup>b</sup> Le dosage est défini ici comme la moyenne d'heures passées en garderie par semaine durant la première année de l'étude pour les garderies G1 et G2, respectivement.

<sup>c</sup> Cet effet manque de crédibilité, car : a) aucune indication de cet effet n'a été détectée dans le modèle 1; et b) le nombre d'observations par groupe est relativement petit ( $n < 100$ ).

## 6.2.2 Définitions précises de l'intégrité du programme: Fidélité et qualité du programme en garderie

En examinant les résultantes standardisées de l'ÉPE-AD (par groupe expérimental), nous avons constaté que l'exposition lors de la première année au programme en garderie testé avait un effet positif sur le développement durant la deuxième année du projet. Dans cette section, nous vérifions si une autre façon de définir l'intensité du traitement, soit la fidélité globale et la qualité du programme, sera associée de la même manière aux résultantes de la deuxième année.<sup>102</sup> Nous présentons les résultats de ces analyses dans des sections séparées pour les résultantes standardisées de l'ÉPE-AD et pour les échelles de vocabulaire.

### **Détails de spécification du modèle**

Les données ont été analysées de la même manière qu'à la section 6.2 pour les analyses des résultantes de la première année. Les estimations de la fidélité et de la qualité du nouveau programme préscolaire de service de garde ont remplacé l'appartenance au groupe expérimental dans la spécification du modèle de régression de base pour les analyses DD (modèle 5 pour les résultantes de l'ÉPE-AD; modèles 1 à 5 pour les échelles de vocabulaire normalisées). Dans des analyses secondaires, nous avons vérifié si les effets du groupe expérimental et les effets estimés selon les indices de fidélité/qualité du programme sont redondants. Dans ce cas, le modèle de régression incluait des effets directs et des effets DD pour l'appartenance au groupe expérimental afin de contrôler pour la portion correspondante de l'effet qui est due à la fidélité/qualité du programme.

### **Effets de la fidélité/qualité globale du programme sur les résultantes de la deuxième année : Échelles de l'ÉPE-AD (standardisées)**

Nous rapportons les estimations DD de l'effet à long terme de la fidélité/qualité globale du programme sur les résultantes de la deuxième année au tableau 6.7. Nous ne rapportons pas les résultats issus des analyses des scores de Communication basées sur l'échantillon total, car ils ne diffèrent pas substantiellement de ceux rapportés ci-dessous. Comme pour les analyses menées sur les effets de l'exposition à la garderie durant la première année (dosage), les résultats sont basés sur le sous-échantillon d'enfants inscrits à la garderie seulement. Nous ne faisons part que des résultats basés sur les estimations globales de la fidélité et de la qualité du programme en garderie testé. Ce type de présentation nous permet de transmettre toute l'information pertinente de façon succincte (le chapitre 5 décrit de façon plus détaillée les analyses des variables, pour les lecteurs qui s'y intéressent).

Les résultats révèlent des effets positifs pour la qualité globale relativement au niveau de base pour les échelles de Communication et d'Habilités cognitives de l'ÉPE-AD, à la sixième et à la septième évaluation (cinquième et sixième post-tests). Pour ce qui est de la fidélité globale, nous observons un effet correspondant seulement pour les Habiletés cognitives. En somme, pour ces deux résultantes, le patron de résultats observés dans les analyses par groupe expérimental est confirmé.

---

<sup>102</sup> Nous n'incluons pas les résultats de dimensions spécifiques de la qualité du programme (par ex., qualité de lecture) puisqu'ils ne changent pas de façon substantielle la façon dont nous rapportons les résultats de la deuxième année.

L'effet de la fidélité globale sur les Habiletés cognitives indique qu'une augmentation quant à cette variable de fidélité de l'ordre de 0,20 est associée à une augmentation du score de changement de 0,24 écart-type, ce qui représente un effet dit petit. De son côté, l'effet de la qualité globale démontre qu'une augmentation de deux unités quant à l'indice de qualité est associée à une augmentation du score de changement de 0,26 écart-type, ce qui représente aussi un effet dit petit. Bien que petits, ces effets étaient statistiquement détectables. En résumé, nous observons, pour ces deux résultantes, des preuves positives de l'impact du programme, en employant trois différentes définitions opérationnelles de l'intensité du traitement: a) l'appartenance aux groupes expérimentaux; b) le dosage/exposition lors de la première année; et maintenant, c) la fidélité/qualité globale du nouveau programme préscolaire de service de garde.

Les analyses secondaires révèlent que l'effet DD du groupe expérimental pour la septième évaluation, qui a été observé dans les analyses par groupe expérimental menées sur les scores de Communication, est beaucoup moins important et devient statistiquement non significatif lorsque la qualité globale du programme fait l'objet d'un contrôle statistique, DD ajustée=-0,070, ET=0,127. On observe des réductions similaires dans les estimations DD ajustées pour l'effet du groupe expérimental associé aux Habiletés cognitives, DD ajustée (sixième évaluation)=-0,207, ET=0,151; DD ajustée (septième évaluation)=-0,287\*, ET=0,117. Les effets du groupe expérimental disparaissent presque entièrement lorsque les effets associés à la fidélité globale sont employés comme covariables, DD ajustée (sixième évaluation)= 0,045, ET=0,277; DD ajustée (septième évaluation)=0,047, ET=0,211. Les résultats soutiennent l'idée que les effets associés du groupe expérimental produisent leurs effets sur les résultantes de l'ÉPE-AD par l'entremise de la qualité et de la fidélité du programme.

**Tableau 6.7 : Qualité et fidélité globales du programme comme prédicteurs des résultantes de la deuxième année**

	<b>Échelles de l'ÉPE-AD – Erreurs groupées par garderie</b>							
	<b>Communication</b>		<b>Conscience de soi</b>		<b>Habilités cognitives</b>		<b>Vocabulaire expressif</b>	
	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
<b>Fidélité globale</b>								
Au niveau de base	-,607	,314	-,677	,429	-1,03 <sup>†</sup>	,329	1,00	,460
DD au 3 <sup>e</sup> post-test	-,087	,526	-,060	,465	,276	,461	-,006	,405
DD au 4 <sup>e</sup> post-test	,106	,371	,370	,471	,576	,483	,466	,480
DD au 5 <sup>e</sup> post-test	,213	,353	,637	,441	,948 <sup>†</sup>	,273	977 <sup>m</sup>	,535
DD au 6 <sup>e</sup> post-test	,658	,459	,658	,459	1,18 <sup>†</sup>	,255	1,15 <sup>m</sup>	,541
Test F composite de Wald	F (4, 11) = ,772		F (4, 11) = 2,57 <sup>m</sup>		F (4, 11) = 5,32*		F (4, 11) = 3,64*	
Nombre de participants	148		148		148		148	
<b>Qualité globale</b>								
Au niveau de base	-,023	,051	-,024	,068	-,012	,052	-,030	,063
DD au 3 <sup>e</sup> post-test	-,040	,079	-,004	,071	-,101	,055	,059	,075
DD au 4 <sup>e</sup> post-test	-,019	,032	,065	,041	,037	,065	,096 <sup>m</sup>	,048
DD au 5 <sup>e</sup> post-test	,099*	,042	,049	,073	,132*	,049	,073	,087
DD au 6 <sup>e</sup> post-test	,143*	,051	,073	,072	,133*	,044	,118	,093
Test F composite de Wald	F (4, 11) = 8,23 <sup>†</sup>		F (4, 11) = 1,32		F (4, 11) = 7,72 <sup>†</sup>		F (4, 11) = 1,64	
Nombre de participants	148		148		148		148	

**Note :** L'effet de la qualité et de la fidélité du nouveau programme préscolaire de service de garde au niveau de base est une estimation de l'association au pré-test et ne constitue pas un effet du programme. Cette association est neutralisée par l'estimateur DD. *Des estimations DD positives dénotent des effets positifs pour la qualité et la fidélité du programme en garderie.* Les indices de qualité et de fidélité du programme en garderie sont dérivés d'une analyse qualitative. Les scores de fidélité vont de 0 à 1, tandis que les scores de qualité vont de 0 à 7. Il faut adapter l'interprétation des coefficients en fonction de ces différences d'échelle. Les résultats sont basés sur l'inscription à la garderie au moment du test, mais ils sont comparables aux analyses reposant sur l'appartenance au groupe à la fin de la première année. Effets de la fidélité/qualité globale du programme sur les résultantes de la deuxième année : Échelles de l'EOWPVT et de l'ÉVIP-R (non normalisées)

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux).

Nous présentons les estimations de l'impact des indices de fidélité/qualité globale du programme sur les scores de l'EOWPVT et de l'ÉVIP-R au tableau 6.8. Les résultats de la série complète d'analyses avec l'inclusion séquentielle des covariables sont rapportés (modèles 1 à 5). De plus, comme pour les analyses précédentes avec ces deux résultantes, nous avons employé un estimateur de l'effet traitement reposant sur le contrôle statistique des scores de Communication des enfants au niveau de base. Les résultats démontrent qu'après un ajustement statistique pour les covariables du modèle, on observe un impact positif de la qualité globale du programme tant pour l'ÉVIP-R que pour l'EOWPVT. Dans les deux cas, l'effet est un changement de l'ordre de deux mots par unité de qualité du programme. En revanche, les résultats ont indiqué que les estimations de la fidélité globale du programme n'étaient pas constamment reliées aux mesures de vocabulaire.

Comme prévu, les analyses effectuées en fonction de la qualité se sont avérées être plus sensibles à l'impact de la qualité du programme que les analyses plus simples fondées sur l'assignation aux groupes expérimentaux. Le résultat est important, car il donne de la crédibilité à notre postulat selon lequel l'effet du groupe expérimental observé relativement au groupe témoin en garderie (mais qui n'a pas persisté suite à l'inclusion de la série complète de covariables) est: a) réel et donc, b) susceptible de devenir statistiquement significatif avec une plus grande taille d'échantillon. Nous ne présentons pas ici les estimations ajustées de la fidélité/qualité globale du programme, parce que l'effet du groupe expérimental n'était pas statistiquement significatif. Nous ne savons pas encore si l'insertion, dans les analyses, des données de la seconde cohorte d'enfants inscrits au projet Capacité d'apprentissage confirmera notre attente selon laquelle l'indication que nous observons ici d'un effet est réelle.

**Tableau 6.8 : Scores de Vocabulaire de la deuxième année en fonction des indices globaux de fidélité et de qualité du programme – Scores de l'ÉVIP-R et de l'EOWPVT (non normalisés)**

	<i>Inclusion séquentielle des covariables – Erreurs groupées par garderie</i>									
	(1)		(2)		(3)		(4)		(5)	
Effets programme	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste	Effet	ET robuste
<b>Scores de l'ÉVIP-R</b>										
Fidélité globale	4,56	3,75	5,89	4,50	4,40	4,91	4,05	5,35	3,31	4,81
Qualité globale	1,97 <sup>m</sup>	1,04	2,20 <sup>m</sup>	1,10	2,44 <sup>†</sup>	,697	2,13*	,789	1,79*	,724
<b>Scores de l'EOWPVT</b>										
Fidélité globale	6,12*	2,42	4,98	3,22	4,56	3,78	3,06	4,89	2,29	4,84
Qualité globale	1,67 <sup>†</sup>	,402	1,90 <sup>†</sup>	,516	1,99 <sup>†</sup>	,473	2,04 <sup>†</sup>	,542	1,95 <sup>†</sup>	,521
Nombre de participants	154		151		150		149		147	

**Note :** Les indices globaux de qualité et de fidélité du programme sont dérivés d'analyses qualitatives. Les scores de fidélité globale du programme vont de 0 à 1, tandis que les scores de qualité globale du programme vont de 0 à 7. Les erreurs-types ont été estimées grâce au programme SPSS pour le calcul de l'estimateur robuste à l'hétérogénéité de White (White, 1980). De plus, elles ont été groupées par garderie afin d'ajuster pour l'effet commun du milieu de garderie. Les résultats sont basés sur l'inscription à la garderie au moment du test, mais ils sont comparables aux analyses reposant sur l'appartenance au groupe à la fin de la première année. Les détails de spécification sont fournis à la section 6.2.2.

<sup>m</sup> dénote une signification à  $p < 0,10$ , \*  $p < 0,05$ , †  $p < 0,01$  (tous les tests sont bilatéraux, 20 degrés de liberté).



### 6.2.3 Caractéristiques linguistiques de l'échantillon

Plusieurs effets rapportés dans les analyses principales des résultantes de la deuxième année étaient marginalement significatifs et/ou leur ampleur était moins importante que prévue (voir les effets du groupe expérimental pour l'ÉVIP-R et l'EOWPVT). Il est logique à ce stade de vérifier si une liste ciblée de covariables peut modérer les effets du groupe expérimental. Comme dans les analyses des résultantes de la première année, nous vérifierons ici s'il y a indication que les caractéristiques linguistiques de l'échantillon au niveau de base peuvent aider à identifier la population-cible idéale pour le programme testé.

Le programme visait à favoriser le développement des enfants francophones vivant en milieu minoritaire. L'intervention repose sur la prémisse que le degré d'exposition de nombreux enfants à la langue et à la culture francophones est insuffisant pour assurer leur développement en tant que francophones. Comme dans la section 5.1.3, nous prévoyions que les enfants, pour qui ce besoin était le plus grand, bénéficieraient davantage du programme. Cette possibilité a été explorée dans les analyses qui suivent.

Les données ont été analysées de la même manière que dans la section 5.1.3. Toutefois, dans le présent cas, les résultats ont varié plus largement selon la variable modératrice considérée. Pour compenser pour cette complexité additionnelle, la présentation des résultats a été simplifiée. Nous rapportons seulement les estimations pour la condition de « faible exposition au français » ou pour la condition de « forte exposition au français », en fonction de celle ayant montré un gain quant à l'effet observé du programme. Les estimations conditionnelles qui ont été éliminées intentionnellement des tableaux sont décrites de façon sommaire dans le corps du texte.

*Échelle de Communication (standardisée)*. Contrairement à la section 5.1.3, nous rapportons seulement les estimations conditionnelles pour la condition de faible exposition au français (voir tableau 6.9). Comme pour les résultats rapportés dans le tableau 5.22, aucun effet programme positif et significatif n'a été observé pour la condition de « forte exposition » (non rapportée ici). Pour certains indicateurs, les résultats suggèrent que les différences entre les groupes tendent à être plus grandes à la fin de la deuxième année, pour ceux dont l'exposition au français est faible. Toutefois, le patron de résultats n'est pas constant à travers les différents indicateurs de profil linguistique, ce qui donne à penser a) que l'exposition au français en tant que construit unitaire est un modérateur moins pertinent du développement en deuxième année que du développement lors de la première année et, donc b) qu'il se serait utile, dans l'interprétation des résultats, de tenir compte des attributs idiosyncrasiques des indicateurs qui démontrent l'effet. Par exemple, on a soutenu précédemment que le continuum linguistique parlé par l'enfant et la langue des activités de littératie seraient les modérateurs les plus constants de l'effet du programme. Les caractéristiques de chaque groupe de comparaison peuvent aussi être pertinentes pour déterminer le modérateur le plus pertinent.

Relativement au groupe témoin en garderie, cet effet modérateur du profil linguistique était statistiquement significatif au cinquième post-test lorsque le continuum linguistique parlé par l'enfant, la langue des activités de littératie et le type de foyer (p. ex., le PLOP défini purement sur la base de la langue parlée à l'enfant) étaient utilisés. Cet effet n'était pas significatif du point de vue statistique lorsqu'on l'estimait pour l'échantillon dans son ensemble (voir les tableaux 6.1 et 6.2). Il semblerait alors que l'effet de la deuxième année pour l'échelle de Communication fait irruption quatre mois plus tôt lorsque l'exposition au français est faible (selon les modérateurs

sélectionnés). Au sixième post-test, le même effet s'observe pour la moyenne de l'échantillon dans son ensemble (voir les tableaux 6.1 et 6.2).

En comparaison avec le groupe témoin hors garderie, l'ampleur de l'effet du programme a varié significativement en fonction de la langue des activités de littératie et du continuum linguistique parlé par la mère à l'enfant, pour les trois périodes post-test rapportées dans le tableau 6.9. La convergence de ces deux résultats est justifiable étant donné que la mère est habituellement le principal pourvoyeur de soins à l'enfant dans le groupe témoin hors garderie. Les données probantes à l'appui d'un effet programme étaient les plus solides pour le quatrième et le cinquième post-test alors que le développement du groupe programme en garderie et du groupe témoin hors garderie était estimé comme statistiquement équivalent à compter du sixième post-test.

**Tableau 6.9 : Effets estimés du programme pour les scores de Communication standardisés (français) conditionnels à la faible exposition au français – Échantillon total**

	<i>Modérateurs de l'effet du programme</i>											
	Continuum linguistique (enfant)		Langue des activités de littératie		Type de foyer (PLOP)		Type de foyer (relatif à l'enfant)		Continuum linguistique (mère-enfant)		Continuum linguistique (père-enfant)	
	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET
<b>G1<sup>a</sup> v. G2</b>												
Niveau de base	,283*	,112	,330*	,100	,422*	,153	,438*	,153	,352*	,113	,137	,137
4 <sup>e</sup> post-test DD	-,216	,195	-,256	,199	-,204	,225	-,270	,187	-,172	,164	-,226	,141
5 <sup>e</sup> post-test DD	-,379 <sup>Δ</sup>	,224	-,407* <sup>Δ</sup>	,195	-,200	,227	-,454* <sup>Δ</sup>	,198	-,294	,241	-,258	,155
6 <sup>e</sup> post-test DD	-,313 <sup>m</sup>	,177	-,510*	,174	-,209	,192	-,192	,213	-,264	,197	-,140	,169
<b>G1 v. G3</b>												
Niveau de base	-,292*	,095	-,194 <sup>m</sup>	,111	,009	,103	,024	,116	-,192	,116	-,269	,106
4 <sup>e</sup> post-test DD	-,216	,283	-,698* <sup>Δ</sup>	,165	-,373 <sup>Δ</sup>	,220	-,426 <sup>Δ</sup>	,247	-,386* <sup>Δ</sup>	,172	-,307	,341
5 <sup>e</sup> post-test DD	,037	,189	-,286 <sup>Δ</sup>	,171	,072	,195	-,403* <sup>Δ</sup>	,180	-,069 <sup>Δ</sup>	,195	-,041	,304
6 <sup>e</sup> post-test DD	,018	,347	-,338 <sup>Δ</sup>	,313	,076	,292	-,177	,319	-,158 <sup>Δ</sup>	,354	-,002	,302
Interaction triple (F de Wald)	F (5, 15) = 3,63*		F (5, 15) = 6,45*		F (5, 15) = 6,12*		F (5, 15) = 18,96 <sup>†</sup>		F (5, 15) = 3,73*		F (5, 15) = 4,73*	
Nombre de participants	221		221		221		221		221		215	

**Note :** Les résultats pour le troisième post-test ne sont pas rapportés, mais ils figuraient dans les analyses. Les variables de foyers sont dichotomiques. Par conséquent, les estimations représentent le contraste entre les familles endogames francophones et les autres types de familles. Pour les deux autres variables, le tableau rapporte les estimations obtenues selon un modèle de régression de l'effet du programme à des niveaux d'utilisation de la langue française choisis arbitrairement. L'estimation de la borne supérieure a été calculée pour le score maximal sur l'échelle, tandis que l'estimation de la borne inférieure a été estimée selon la valeur se situant à un écart-type sous la moyenne. **Le groupe programme en garderie (G1) est mis en référence, donc les effets négatifs indiquent un avantage pour ce groupe.** L'évaluation au niveau de base et le groupe programme en garderie sont les catégories de référence pour tous les effets DD et DDD.

<sup>a</sup> G1, G2, et G3 indiquent le groupe programme en garderie, le groupe témoin en garderie et le groupe témoin hors garderie, respectivement.

<sup>m</sup>  $p < 0,10$ , \*  $p < 0,05$ , \*\*  $p < 0,01$ , <sup>†</sup>  $p < 0,001$ ; <sup>Δ</sup> indique les effets qui varient significativement en fonction des divers degrés d'exposition au français  $p < 0,10$ .

***Échelle de Conscience de soi (standardisée).*** Les analyses de l'échelle de Conscience de soi ont offert peu de preuves que l'ampleur des effets du programme variait significativement et systématiquement en fonction du profil linguistique, et certains des résultats obtenus dans ces analyses étaient contradictoires. Au lieu de présenter des données portant à confusion, nous avons écarté du présent rapport les résultats issus de ces analyses. Notons par ailleurs qu'il n'était attendu que cette résultante particulière dépende du profil linguistique de la même manière que les mesures de vocabulaire et la Communication.

***Échelle d'Habilités cognitives (standardisée).*** Dans les analyses de la première année (section 5.1.3), les effets du programme en relation avec les Habiletés cognitives étaient plus grands pour les enfants à forte exposition au français. Les données de la deuxième année ont présenté un portrait similaire. Les estimations de l'effet du programme pour les enfants à forte exposition sont rapportées dans le tableau 6.10. Il est à noter que, sauf pour une seule exception, aucune des estimations pour la condition de « faible exposition » n'était statistiquement significative (non rapportées).<sup>103</sup>

Les analyses présentées ici indiquent que ce sont les participants à forte exposition qui étaient principalement responsables des effets du programme rapportés dans le tableau 6.2. Pour trois des six modérateurs potentiels, l'ampleur de l'effet du programme a varié significativement en fonction du profil linguistique de la famille, selon le test global de Wald. Trois des six modérateurs ont atteint la signification statistique au sixième post-test pour la comparaison des groupes en garderie (comme l'indique le symbole  $\Delta$ ). En fait, le groupe témoin en garderie engendre des effets programme positifs et significatifs aux cinquième et sixième post-tests, pour tous les modérateurs linguistiques.

Les résultats présentés au tableau 6.10 indiquent également que le programme exerce un effet considérable, comparativement au groupe témoin hors garderie, lorsqu'il est conditionnel aux indicateurs de profil linguistique. Ce dernier effet n'a pas été observé avec la langue des activités de littératie et il était propre au cinquième post-test. En d'autres mots, l'effet programme le plus robuste est toujours obtenu avec les garderies témoins, même dans les présentes analyses.

---

<sup>103</sup> L'exception était les effets DD, relatifs au groupe témoin hors garderie, pour les cinquième et sixième post-tests conditionnels au continuum linguistique parlé par le père à l'enfant. Les effets estimés étaient statistiquement significatifs et plus grands pour le groupe à forte exposition. La source de cet effet n'est pas claire.

**Tableau 6.10 : Effets estimés du programme pour les scores d'Habilités cognitives standardisés (français) conditionnels à la forte exposition au français**

	<i>Modérateurs de l'effet du programme</i>											
	Continuum linguistique (enfant)		Langue des activités de littératie		Type de foyer (PLOP)		Type de foyer (relatif à l'enfant)		Continuum linguistique (mère-enfant)		Continuum linguistique (père-enfant)	
	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET
<b>G1 v. G2</b>												
Niveau de base	,364*	,165	,390*	,153	,216	,155	,253 <sup>m</sup>	,124	,363*	,142	,392*	,150
4e post-test DD	-,274	,259	-,304	,251	-,256	,212	-,246	,211	-,201	,243	-,234 <sup>Δ</sup>	,236
5e post-test DD	-,443**	,128	-,354*	,157	-,477**	,129	-,358**	,126	-,474**	,124	-,394**	,120
6e post-test DD	-,513*	,200	-,500**	,142	-,584** <sup>Δ</sup>	,161	-,465**	,161	-,501** <sup>Δ</sup>	,144	-,501** <sup>Δ</sup>	,171
<b>G1 v. G3</b>												
Niveau de base	,389 <sup>m</sup>	,207	,331*	,145	,330 <sup>m</sup>	,192	,352 <sup>m</sup>	,186	,336*	,146	,337	,206
4e post-test DD	-,163	,256	-,037	,249	-,266 <sup>Δ</sup>	,204	-,240 <sup>Δ</sup>	,221	-,020	,217	-,104	,232
5e post-test DD	-,441 <sup>†</sup>	,100	-,173	,100	-,463 <sup>†</sup>	,105	-,278**	,089	-,361 <sup>†</sup>	,074	-,295*	,105
6e post-test DD	-,345	,278	-,248	,226	-,524* <sup>Δ</sup>	,199	-,341	,208	-,249	,220	-,297	,244
Interaction triple (F de Wald)	F (5, 15) = 5,53**		F (5, 15) = ,719		F (5, 15) = 2,50 <sup>m</sup>		F (5, 15) = 7,54 <sup>†</sup>		F (5, 15) = 1,24		F (5, 15) = 1,98	
Nombre de participants	211		211		211		211		211		210	

**Note :** Les variables de foyer sont dichotomiques. Par conséquent, les estimations représentent le contraste entre les familles endogames francophones et les autres types de familles. Pour les deux autres variables, le tableau rapporte les estimations obtenues selon un modèle de régression des effets du programme à des niveaux d'utilisation du français choisis arbitrairement. L'estimation de la borne supérieure a été calculée pour le score maximal sur l'échelle, tandis que l'estimation de la borne inférieure a été évaluée en fonction de la valeur se situant à un écart-type sous la moyenne. **Le groupe programme en garderie (G1) est mis en référence, donc les effets négatifs indiquent un avantage pour ce groupe.** L'évaluation au niveau de base et le groupe programme en garderie sont les catégories de référence pour tous les effets DD et DDD.

G1, G2, et G3 indiquent le groupe programme en garderie, le groupe témoin en garderie et le groupe témoin hors garderie, respectivement.

<sup>m</sup> $p < 0,10$ , \* $p < 0,05$ , \*\* $p < 0,01$ , <sup>†</sup> $p < 0,001$ ; <sup>Δ</sup> indique des effets qui varient significativement en fonction des divers degrés d'exposition au français  $p < 0,10$

***Échelle de Vocabulaire expressif (standardisée)***. Les estimations conditionnelles de l'effet du programme sont rapportées pour la condition de « faible exposition au français » dans le tableau 6.11. Le patron des résultats pour la condition de « forte exposition au français » était très semblable à celui qui a été obtenu lorsque l'effet moyen était estimé pour cette résultante pour tout l'échantillon; il serait donc redondant de rapporter ici ces estimations (résultats offerts sur demande). Dans ces analyses, les impacts positifs du programme ont été observés pour le cinquième et le sixième post-test (voir tableau 6.2).

Deux principaux résultats présentés au tableau 6.11 distinguent ces résultats de ceux qui ont été présentés dans le tableau 6.2. Lorsqu'on compare les groupes en garderie, l'effet positif du programme à la fin de l'année émerge quatre mois plus tôt, soit dès le quatrième post-test, pour les enfants à faible exposition au français. Ici, le continuum linguistique parlé par l'enfant ainsi que le continuum linguistique parlé par le père à l'enfant étaient des modérateurs statistiquement significatifs de l'effet (comme l'indique le symbole  $\Delta$ ). La deuxième distinction intéressante réside dans le fait qu'un effet relatif au groupe témoin hors garderie se dégage brièvement pour la période du cinquième post-test. Comme dans les analyses précédentes (c.-à-d. Habiletés cognitives), ce dernier effet ne persiste pas jusqu'au sixième post-test, comme celui qui met en cause les deux groupes en garderie.

Les résultats basés sur le type de foyer PLOP (« foyer » basé principalement sur la première langue officielle parlée) dans le tableau 6.11 constituent clairement des données aberrantes. Nous attribuons ce résultat particulier au fait qu'il n'est pas aussi explicitement lié à des comportements réels des personnes dans l'environnement de l'enfant que les autres indicateurs. Il convient de se rappeler que les autres variables linguistiques employées ici comme modérateurs se basent sur des variables explicitement liées aux interactions avec l'enfant.

**Tableau 6.11 : Effets estimés du programme pour les scores standardisés de Vocabulaire expressif (français) conditionnels à la faible exposition au français**

	<i>Modérateurs de l'effet du programme</i>											
	Continuum linguistique (enfant)		Langue des activités de littératie		Type de foyer (PLOP)		Type de foyer (relatif à l'enfant)		Continuum linguistique (mère-enfant)		Continuum linguistique (père-enfant)	
	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET
<b>G1<sup>a</sup> v. G2</b>												
Niveau de base	,222	,148	,307*	,150	,295	,192	,417	,176	,347	,165	,302 <sup>m</sup>	,172
4 <sup>e</sup> post-test DD	-,502 <sup>**Δ</sup>	,151	-,468*	,169	-,111	,128	-,366 <sup>**</sup>	,122	-,219 <sup>m</sup>	,118	-,432 <sup>**Δ</sup>	,134
5 <sup>e</sup> post-test DD	-,556 <sup>†</sup>	,131	-,516*	,183	-,227 <sup>Δ</sup>	,168	-,496 <sup>**</sup>	,147	-,384*	,155	-,558 <sup>**</sup>	,171
6 <sup>e</sup> post-test DD	-,396 <sup>m</sup>	,200	-,446*	,203	-,126 <sup>Δ</sup>	,209	-,236	,233	-,233	,233	-,367 <sup>m</sup>	,204
<b>G1 v. G3</b>												
Niveau de base	-,280	,380	-,202	,464	,129	,276	,245	,388	,172	,284	-,120	,444
4 <sup>e</sup> post-test DD	-,224	,384	-,515 <sup>Δ</sup>	,340	,049	,184	-,233	,394	-,277	,370	-,184	,532
5 <sup>e</sup> post-test DD	-,441	,463	-,716 <sup>m</sup>	,346	-,516	,361	-,927 <sup>*Δ</sup>	,501	-,830 <sup>*Δ</sup>	,397	-,677	,628
6 <sup>e</sup> post-test DD	-,344	,586	-,530	,483	-,360	,488	-,792	,597	-,827	,495	-,580	,691
Interaction triple (F de Wald)	F (5, 15) = ,615		F (5, 15) = 10,67 <sup>†</sup>		F (5, 15) = 2,68*		F (5, 15) = 7,38 <sup>†</sup>		F (5, 15) = 3,23*		F (5, 15) = 3,65*	
Nombre de participants	211		211		211		211		211		201	

**Note :** Les variables de foyer sont dichotomiques. Par conséquent, les estimations représentent le contraste entre les familles endogames francophones et les autres types de familles. Pour les deux autres variables, le tableau rapporte les estimations obtenues selon un modèle de régression des effets du programme à des niveaux d'utilisation du français choisis arbitrairement. L'estimation de la borne supérieure a été calculée pour le score maximal sur l'échelle, tandis que l'estimation de la borne inférieure a été évaluée en fonction de la valeur se situant à un écart-type sous la moyenne. **Le groupe programme en garderie (G1) est mis en référence, donc les effets négatifs indiquent un avantage pour ce groupe.** L'évaluation au niveau de base et le groupe programme en garderie sont les catégories de référence pour tous les effets DD et DDD.

<sup>a</sup> G1, G2 et G3 indiquent le groupe programme en garderie, le groupe témoin en garderie et le groupe témoin hors garderie, respectivement.

<sup>m</sup>  $p < 0,10$ , \* $p < 0,05$ , \*\* $p < 0,01$ , † $p < 0,001$ ; Δ indique les effets qui varient significativement en fonction des divers degrés d'exposition au français  $p < 0,10$ .

**EOWPVT & ÉVIP-R.** Les analyses par profil linguistique ont été effectuées, tel que précédemment, pour l'EOWPVT et pour l'ÉVIP-R. Les analyses ont été basées sur l'appartenance au groupe expérimental au moment du test. Les résultats pour l'ÉVIP-R ne sont pas rapportés étant donné qu'ils ont fourni peu d'évidences selon lesquelles le profil linguistique constituait un modérateur pour cette variable (d'autres études d'intervention n'ont pas réussi à observer des effets programme avec cette variable, voir ci-dessous; pour un contre-exemple, voir Maltais, 2007). La seule exception observée était auprès de la variable langue des activités de littératie. Pour les enfants provenant de familles ayant obtenu un score faible pour cette variable, un effet important du programme s'observe relativement au groupe témoin hors garderie, effet = -6,30\*, ET = 2,28. Autrement, les effets estimés pour cette résultante étaient statistiquement non distinguables de ceux qui ont été rapportés dans les analyses précédentes où aucun effet n'a été discerné (voir les tableaux 6.3 et 6.4).

Lorsqu'on le compare à l'ÉVIP-R, l'EOWPVT s'est avéré être plus sensible à l'intervention lorsque les effets du programme étaient estimés pour les enfants à faible exposition au français. Ce patron de résultats est congruent avec les résultats obtenus de recherches antérieures qui ont également conclu que l'EOWPVT était une mesure plus sensible (voir Whitehurst et al., 1988; Arnold, Loginan, Whitehurst, et Epstein, 1994; Hargrave & Sénéchal, 2000). De plus, la sous-échelle de Vocabulaire expressif de l'ÉPE-AD souffrait du problème de l'étendue restreinte de réponse, alors que ceci n'était pas un facteur pour l'EOWPVT. Les effets estimés basés sur l'EOWPVT pour les enfants à « faible exposition au français » sont rapportés dans le tableau 6.12. Comme on s'y attendait, le continuum linguistique parlé par l'enfant au niveau de base est le modérateur le plus constant de l'effet du programme, suivi par la langue des activités de littératie. Les enfants qui utilisaient moins le français pour s'exprimer lors de la période de base semblaient avoir profité le plus de l'intervention pour ce qui est du développement de leur vocabulaire expressif. Cette affirmation était vraie peu importe si c'était le groupe programme en garderie ou le groupe témoin hors garderie qui servait de point de comparaison. Les effets observés sont de l'ordre de quatre à six mots, ce qui représente un impact relativement modeste (à peu près équivalent à celui observé dans une étude évaluant les impacts du programme « Head Start » sur le vocabulaire, Wasik, Bond, & Hindman, 2006).



**Tableau 6.12 : Effets estimés du programme sur l'EOWPVT pour les enfants à «faible exposition».**

	<i>Modérateurs de l'effet du programme</i>											
	Continuum linguistique (enfant)		Langue des activités de littératie		Type de foyer (PLOP)		Type de foyer (relatif à l'enfant)		Continuum linguistique (mère-enfant)		Continuum linguistique (père-enfant)	
	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET	Effet	ET
<b>G1<sup>a</sup> v. G2</b>	-4,16* <sup>Δ</sup>	1,97	-3,43	2,22	-,948	2,52	-2,25	2,86	-1,81	2,66	-3,91	2,37
<b>G1 v. G3</b>	-6,23 <sup>mΔ</sup>	3,04	-9,24 <sup>†Δ</sup>	2,05	-1,81	4,07	-5,27	4,65	-5,27 <sup>Δ</sup>	3,38	-7,47 <sup>mΔ</sup>	4,26
Interaction groupe x langue (F de Wald)	F (2, 21) = 3,99*		F (2, 21) = 26,99 <sup>†</sup>		F (2, 21) = 1,87		F (2, 21) = ,82		F (2, 21) = 4,33*		F (2, 21) = 1,89	
Nombre de participants	222		222		222		222		222		215	

**Note :** Les variables de foyer sont dichotomiques. Par conséquent, les estimations représentent le contraste entre les familles endogames francophones et les autres types de familles. Pour les deux autres variables, le tableau rapporte les estimations obtenues selon un modèle de régression des effets du programme à des niveaux d'utilisation de la langue française choisis arbitrairement. L'estimation de la borne inférieure est rapportée en fonction de la valeur se situant à un écart-type sous la moyenne de l'échantillon. **Les garderies programme (G1) sont mis en référence, donc les effets négatifs indiquent un avantage pour le groupe des garderies programme.** L'évaluation au niveau de base et le groupe des garderies programme sont les catégories de référence pour tous les effets DD et DDD. Les groupes à l'étude étaient ceux qui étaient là au moment où les données pour la variable résultante ont été recueillies.

<sup>a</sup> G1, G2 et G3 indiquent le groupe programme en garderie, le groupe témoin en garderie et le groupe témoin hors garderie respectivement.

<sup>m</sup>  $p < 0,10$ , \*  $p < 0,05$ , \*\*  $p < 0,01$ , †  $p < 0,001$ ; <sup>Δ</sup> indique les effets qui varient considérablement en fonction de divers degrés d'exposition au français  $p < 0,10$ .

## 6.3 SOMMAIRE

Tout au long de ce rapport, nous avons employé une stratégie reposant sur le principe d'opérations convergentes, où le concept «intensité du traitement» ou «exposition au traitement» a été défini de plusieurs manières afin de vérifier dans quelle mesure ces analyses produisaient un ensemble cohérent de résultats. Dans les analyses des résultantes de la première année, les estimations de l'effet programme basées sur l'appartenance au groupe expérimental, le dosage et la qualité du programme ont mené à la même conclusion: le programme testé a eu un impact petit mais significatif sur le développement des enfants inscrits aux garderies programme comparativement à ceux inscrits dans les garderies témoins, impact qui était détectable au premier post-test seulement, mais qui est compatible avec un effet « de rattrapage » pour toutes les évaluations. L'effet relatif au groupe témoin hors garderie n'était pas statistiquement significatif dans les analyses par groupe expérimental; il s'agit d'un résultat nul qui n'a pu être vérifié avec les deux autres indices d'exposition au traitement étant donné qu'ils étaient propres aux garderies. Le degré d'accord entre les estimations de l'intensité du traitement a varié en fonction de la résultante considérée, bien qu'on ait obtenu une validation croisée pour de nombreux effets. *Les analyses des résultantes par profil linguistique pour la première année* ont indiqué qu'au moment où la plupart des enfants étaient admis à l'école (52 % de l'échantillon au moment du troisième post-test), le programme avait eu un impact positif. Les enfants qui ont commencé l'intervention avec une faible exposition au français ont présenté des gains quant leurs compétences relatives à la communication (Communication, ÉPE-AD) et au vocabulaire expressif (Vocabulaire expressif, ÉPE-AD) par rapport à leurs pairs des deux groupes de comparaison, (garderie, hors garderie). En ce qui a trait aux enfants qui ont commencé l'intervention avec une forte exposition au français, ils ont présenté des gains importants quant à leur développement cognitif, au moment où la plupart des enfants étaient admis à la prématernelle (à l'âge de quatre ans).

Nous avons appliqué la même stratégie d'opérations convergentes pour les analyses des *résultantes de la deuxième année*. Deux séries d'analyses par groupe expérimental ont été effectuées. La première incluait les trois groupes expérimentaux et permettait aux participants de changer de groupe (pour les échelles standardisées de l'ÉPE-AD, voir le tableau 6.2), alors que la seconde reposait seulement sur les groupes en garderie – elle a été menée dans le contexte des analyses par dosage – et ne permettait pas aux participants de changer de groupe (pour les échelles standardisées de l'ÉPE-AD, voir le tableau 6.5). Les résultats étaient pour ainsi dire identiques : des effets programme statistiquement significatifs (basés sur l'estimateur DD) ont été observés pour les cinquième et sixième post-tests. L'effet sur le Vocabulaire expressif était beaucoup plus marqué et plus persistant dans la seconde analyse (quatrième, cinquième et sixième post-tests). En somme, nous observons durant la deuxième année que le groupe programme en garderie a surmonté leur désavantage initial en *démontrant des gains développementaux plus rapides que le groupe témoin en garderie*. Les enfants du groupe programme en garderie semblaient être «disposés à apprendre » lorsqu'ils sont entrés à la prématernelle et, au moment de la fin de l'année scolaire, ils avaient fait des gains correspondants au niveau des résultantes de l'ÉPE-AD. Les analyses par groupe expérimental n'ont pu révéler un patron de résultats stable à la faveur d'un impact positif du programme relativement au groupe témoin hors garderie.

Nous avons pu observer les effets du groupe expérimental, associés au groupe témoin en en garderie, sur les quatre échelles de l'ÉPE-AD pour la deuxième année : Communication, Conscience de soi, Habiletés cognitives, et Vocabulaire expressif. Nous nous attendions à ce que nos analyses deviennent plus sensibles lorsque la version révisée de l'ÉPE-AD a commencé à être employée. En effet, l'émergence d'effets programme statistiquement significatifs pour la plupart des mesures a coïncidé avec le changement effectué vers l'ÉPE-AD révisée, qui a eu lieu à la sixième évaluation. Ces effets n'ont pas été attestés par les effets du dosage en fonction des groupes expérimentaux. Dans tous les cas, différents degrés d'exposition à la garderie ont eu un effet similaire sans égard au groupe expérimental.

*Les analyses des résultantes de l'ÉPE-AD de la deuxième année en fonction du profil linguistique* ont produit un patron de résultats qui était généralement semblable à celui qui a été rapporté au chapitre 5. Les effets programme positifs sur les compétences relatives à la communication et au vocabulaire expressif étaient plus fortement mis en évidence lorsque l'exposition au français au niveau de base était faible. Par contre, les données probantes à l'appui d'un impact positif sur les habiletés cognitives étaient plus fortes lorsque l'exposition au français au niveau de base était forte. Contrairement aux analyses de la première année, les effets dépendant de la langue relatifs au groupe témoin hors garderie ont bel et bien émergé, même s'ils étaient moins constants à travers les analyses que ceux mettant en cause les groupes en garderie.

Les effets du groupe expérimental sur les scores de Communication et d'Habiletés cognitives ont été attestés par des effets correspondants pour la qualité globale des services de garde. La fidélité globale a également attesté l'effet observé quant aux Habiletés cognitives. Pour ce qui est de la version des échelles utilisées au cours de cette période, les échelles de Communication et d'Habiletés cognitives contiennent de nombreux items qui servent à évaluer la « conscience phonologique », un prédicteur de la réussite académique. Le résultat soutient donc la thèse voulant que les leviers manipulés par le programme influencent favorablement les enfants qui participent à celui-ci. En outre, les analyses de la deuxième année comportaient deux résultantes qui n'étaient pas basées sur l'ÉPE-AD, soit l'EOWPVT et l'ÉVIP-R. Les résultats faisant référence à ces résultantes étaient équivoques lors de la comparaison des groupes expérimentaux, laissant entrevoir un impact positif du programme.

Cependant, nous avons détecté pour les deux mesures de vocabulaire une association positive avec la qualité globale du programme en garderie, association qui était robuste à l'inclusion de covariables et qui, encore une fois, appuie la thèse voulant que les leviers manipulés par le programme dans une tentative de favoriser le développement de l'enfant sont valides. Nous développons cette question au chapitre 7 (Discussion et conclusion). Comme prévu, les effets du programme sur ces mesures ont été plus faciles à déceler pour les enfants à faible degré d'exposition au français (voir les analyses par profil linguistique). Les enfants qui ont commencé le projet avec des degrés relativement faibles d'exposition au français ont démontré un effet programme sur la mesure de vocabulaire expressif (EOWPVT) et ce, tant comparativement au groupe témoin en garderie qu'au groupe témoin hors garderie. De façon similaire à d'autres interventions en garderie (voir Hargrave & Sénéchal, 2000, et plusieurs exemples cités dans le présent document), les bénéfiques ont favorisé, dans cette étude, le vocabulaire expressif plutôt que le vocabulaire réceptif.

## 7. Discussion et conclusion

Le but de la présente recherche était d'évaluer l'impact d'une intervention à deux volets auprès de jeunes enfants de minorité francophone et de leurs parents. Le but du programme testé était de promouvoir le développement des compétences langagières et la préparation à la scolarisation des enfants francophones vivant dans un contexte de langue minoritaire. Certains théoriciens ont proposé que le moteur de l'acquisition de ces compétences soit l'interaction d'un enfant avec son environnement (Bronfenbrenner & Morris, 1998). En s'inspirant de la théorie écologique de Bronfenbrenner (1979) sur le développement de l'enfant, nous avons conçu l'univers de l'enfant comme une série de sphères d'influence qui se chevauchent partiellement et qui comprennent le foyer, la garderie (ou l'école) et la communauté. Comme nous l'avons vu dans l'introduction, la langue majoritaire peut influencer les personnes qui emploient la langue minoritaire dans chacune de ces sphères, y compris les deux milieux primaires, soit le foyer et la garderie. Par conséquent, une approche à deux volets a été employée pour obtenir un effet maximal sur les résultantes de l'enfant (Reese et al., 2010) : a) des Ateliers familles cherchant à influencer positivement l'environnement au foyer par l'entremise des parents, et b) un programme francophone en garderie de grande qualité cherchant à influencer les enfants plus directement.

L'*effet combiné* de ces interventions sur le développement de l'enfant a été évalué par l'entremise d'un devis quasi-expérimental comparant les familles qui ont reçu l'intervention à deux groupes témoins: enfants en garderie (groupe témoin en garderie) et enfants dans un milieu de garde informel (groupe témoin hors garderie). La première comparaison a servi à identifier l'effet de l'intervention relativement aux services de garde habituellement offerts aux enfants francophones. La seconde comparaison a permis de dégager l'effet du programme testé comparativement aux milieux de garde informels. Nous avons observé plusieurs impacts positifs de l'intervention dans les analyses rapportées aux chapitres 5 et 6, qui font l'objet d'une discussion plus détaillée dans les sections suivantes.

Chaque résultat pris individuellement est informatif, mais peut manquer de validité. Ensemble, les conclusions convergent de manière à se renforcer mutuellement, permettant d'atteindre un haut niveau de confiance quant aux résultats grâce à notre utilisation d'une approche de méthodes mixtes à l'évaluation du programme. Cette approche a entraîné le recours à un éventail d'outils, tant qualitatifs que quantitatifs, provenant de plusieurs sources d'information et qui ont tous été sélectionnés en fonction des objectifs de recherche. La richesse de l'information ainsi recueillie facilite la triangulation des résultats de recherche, une stratégie qui aide les chercheurs à arriver à des conclusions solides. Par ailleurs, la complémentarité des données recueillies fournit un portrait plus achevé et plus nuancé du phénomène à l'étude. Parmi d'autres sources, les évaluations des enfants, les sondages auprès des parents, les observations dans les classes de garderie et au cours des Ateliers familles de même que certaines données administratives (p.ex., le registre des présences et des absences des enfants à la garderie) ont tous servi à alimenter les analyses de régression menées. Les principales conclusions du présent rapport sont présentées sommairement pour les résultantes des enfants et des parents dans les tableaux 7.1 et 7.2, respectivement.

**Tableau 7.1 : Sommaire des principales conclusions du Rapport des résultats de la première cohorte du projet Capacité d'apprentissage – Résultantes des enfants**

	Effets de la première année (Niveau de base v. 1 <sup>er</sup> , 2 <sup>e</sup> et 3 <sup>e</sup> post-tests)	Effets de la deuxième année (Niveau de base v. 3 <sup>e</sup> , 4 <sup>e</sup> , 5 <sup>e</sup> et 6 <sup>e</sup> post-tests)
<b>Intégrité du programme et résultantes des enfants</b>		
Fidélité/Qualité/Dosage	<p><b>Fidélité et qualité du programme (différenciation du programme) :</b></p> <ul style="list-style-type: none"> <li>On a démontré que la fidélité et la qualité modulent de façon <b>significative</b> les impacts précoces du programme sur les résultantes des enfants du groupe G1 par rapport aux enfants du groupe G2 <sup>a</sup>.</li> </ul> <p><b>L'importance de la fidélité et de la qualité du programme pour les impacts :</b></p> <ul style="list-style-type: none"> <li>On a démontré que la fidélité et la qualité modulent de façon <b>significative</b> l'impact du programme sur les enfants du groupe G1 par rapport aux enfants du groupe G2 <sup>a</sup>.</li> </ul>	<p><b>L'importance de la fidélité et de la qualité du programme pour les impacts :</b></p> <ul style="list-style-type: none"> <li>On a démontré que la fidélité et la qualité modulent de façon <b>significative</b> l'impact du programme sur les résultantes du groupe G1 par rapport aux enfants du groupe G2 <sup>a</sup>.</li> </ul>
<b>Résultantes des enfants</b>		
Préparation à la scolarisation	<p><b>Les enfants du groupe programme (G1) présentent :</b></p> <ul style="list-style-type: none"> <li>Comme prévu, des gains <b>non significatifs</b> quant aux Habilités physiques par rapport au G2 et au G3;</li> </ul> <p><b>Lorsque les enfants du groupe programme (G1) entrent à la prématernelle (3<sup>e</sup> post-test) <sup>b, c</sup>:</b></p> <ul style="list-style-type: none"> <li>Des gains <b>significatifs</b> quant aux compétences de Communication en français pour les enfants exposés moins activement à la langue, par rapport au G2</li> <li>Des gains <b>significatifs</b> quant aux Habilités cognitives (selon le test en français) pour les enfants exposés plus activement à la langue, par rapport au G2 et au G3 <sup>a</sup></li> </ul>	<p><b>Lorsque les enfants du groupe programme (G1) entrent à la maternelle (6<sup>e</sup> post-test) <sup>b</sup>:</b></p> <ul style="list-style-type: none"> <li>Des gains globaux <b>significatifs</b> en Communication, en Conscience de soi et en Habilités cognitives (réémergence avec des mesures mieux étalonnées) par rapport au G2.</li> <li>Des gains <b>significatifs</b> quant aux compétences de Communication en français pour les enfants exposés moins activement à la langue, par rapport au G2 et au G3.</li> <li>Des gains <b>significatifs</b> quant aux Habilités cognitives (selon le test en français) pour les enfants exposés plus activement à la langue, par rapport au G3.</li> </ul>
Vocabulaire	<p><b>Les enfants du groupe programme (G1) présentent :</b></p> <ul style="list-style-type: none"> <li>Des gains globaux <b>significatifs</b> quant au Vocabulaire expressif (en français), par rapport au G2</li> </ul> <p><b>Lorsque les enfants du groupe programme (G1) entrent à la</b></p>	<p><b>Lorsque les enfants du groupe programme (G1) entrent en maternelle (6<sup>e</sup> post-test) <sup>b</sup>:</b></p> <ul style="list-style-type: none"> <li>Des gains globaux <b>significatifs</b> quant au Vocabulaire expressif (en français), par rapport au G2 (ÉPE-AD)</li> </ul>

Effets de la première année (Niveau de base v. 1 <sup>er</sup> , 2 <sup>e</sup> et 3 <sup>e</sup> post-tests)	Effets de la deuxième année (Niveau de base v. 3 <sup>e</sup> , 4 <sup>e</sup> , 5 <sup>e</sup> et 6 <sup>e</sup> post-tests)
<p><b>prématernelle (3<sup>e</sup> post-test) <sup>b, c</sup>:</b></p> <ul style="list-style-type: none"> <li>Des gains <b>significatifs</b> quant au Vocabulaire expressif (en français) pour les enfants exposés moins activement à la langue, par rapport au G2 (sous-échelle de l'ÉPE-AD).</li> </ul>	<ul style="list-style-type: none"> <li>Des gains <b>significatifs</b> quant au Vocabulaire expressif (en français) pour les enfants exposés moins activement à la langue, par rapport au G2 et au G3 (sous-échelle de l'ÉPE-AD; EOWPVT)</li> </ul>

**Note :** Les résultantes en matière de préparation à la scolarisation comprenaient: a) Communication, b) Conscience de soi, c) Hâbiletés cognitives, et d) Hâbileté physique. Les mesures de Vocabulaire comprenaient a) mesures de Vocabulaire expressif et réceptif fondées sur l'ÉPE-AD (générées par la SRDC) et b) Vocabulaire expressif (EOWPVT) et réceptif (ÉVIP-R) normalisé. Certains effets étaient significatifs pour certaines périodes, mais pas pour d'autres. **Nous soulignons les résultats importants qui présentent un intérêt, tandis que les résultats non significatifs sont implicites.**

<sup>a</sup> G1 se rapporte au groupe programme en garderie; G2 au groupe témoin en garderie et G3 au groupe témoin hors garderie.

<sup>b</sup>Les enfants avaient environ trois ans lorsqu'ils étaient en garderie; quatre ans au moment où ils entraient en prématernelle et cinq au moment d'entrer en maternelle.

<sup>c</sup> Un peu plus que la moitié de l'échantillon inscrit à la prématernelle.

**Tableau 7.2 : Sommaire des principales conclusions du Rapport des résultats de la première cohorte du projet Capacité d'apprentissage – Résultats des parents**

	Effets de la première année (Niveau de base v. post-tests)	Effets de la deuxième année (Niveau de base v. post-tests)
<b>Intégrité du programme et résultantes des parents</b>		
Fidélité/Qualité/Dosage	<ul style="list-style-type: none"> <li>Les effets positifs pour les participants aux ateliers sont <b>significativement</b> liés aux éléments de la mise en œuvre du programme (fréquentation, qualité).</li> </ul>	
<b>Résultantes des parents</b>		
Activités de littératie	<ul style="list-style-type: none"> <li><b>Aucun effet significatif</b> sur la fréquence ou la langue d'usage, soit des activités impliquant l'enfant ou de la modélisation des comportements d'alphabétisation : résultat attribué à des problèmes de mesure.</li> </ul>	<ul style="list-style-type: none"> <li><b>Aucune différence significative</b> sur la fréquence ou la langue d'usage entre les groupes expérimentaux: problèmes de mesure.</li> </ul>
Connaissances, attitudes et opinions	<ul style="list-style-type: none"> <li>Pour les participants aux ateliers, des gains <b>significatifs</b> quant au <i>Sentiment d'auto-efficacité</i> dans leur rôle de parents et aux <i>Connaissance</i> du développement de l'enfant et des pratiques exemplaires, comparativement au pré-test et à un groupe auto-sélectionné de familles non participantes.</li> <li><b>Aucun effet significatif</b> sur les opinions : résultat attribué à des problèmes de mesure.</li> </ul>	<ul style="list-style-type: none"> <li>Sans objet.</li> </ul>

*Note* : Nous soulignons les résultats importants qui présentent un intérêt; la plupart des résultats non significatifs sont implicites.

## 7.1. PRÉPARATION À FRÉQUENTER L'ÉCOLE EN FRANÇAIS?

La principale question à laquelle le présent rapport a tenté de répondre est celle de savoir si l'intervention a eu une incidence positive sur le développement des marqueurs de la préparation à la scolarisation pour les enfants participants. La réponse à cette question semble être positive lorsque l'on tient compte de l'ensemble des sources de données probantes. Cette affirmation demeure vraie que l'impact du programme soit estimé pour l'ensemble de l'échantillon ou spécifiquement pour les enfants qui présentent, au niveau de base, un certain degré d'exposition à la langue française.

### 7.1.1 Effets moyens du programme

Dans l'ensemble, les comparaisons ont révélé des effets positifs basés sur les mesures de préparation à la scolarisation pour le groupe programme en garderie comparativement au groupe témoin en garderie, sur les résultantes de la première année (chapitre 5) et de la deuxième année (chapitre 6). Le même patron de résultats n'a toutefois pas été mis en évidence relativement au groupe témoin hors garderie. Les résultats étaient robustes à l'inclusion de covariables et ils ont, dans plusieurs cas, été confirmés en employant l'une ou plusieurs des définitions alternatives de « l'intensité du traitement », comme la qualité/fidélité du programme et le dosage. Nous avons constaté que les Ateliers familles exercent une influence positive sur certaines résultantes des parents, ce qui donne à penser que les parents sont indirectement responsables d'une partie de l'impact observé du programme sur le développement des enfants. Cette conclusion serait compatible avec la philosophie dont s'est inspiré le programme et avec les recherches antérieures indiquant que l'effet combiné des interventions ciblant les enfants et les parents est supérieur à celui des programmes qui ne portent que sur un seul aspect (Brooks-Gunn, Berlin & Fuligni, 2000; Reese et coll., 2010). Nous présentons maintenant les résultats de façon plus détaillée.

#### ***Groupe programme en garderie versus groupe témoin en garderie***

L'étude portait sur la capacité des enfants francophones à relever leurs futurs défis éducationnels en français. En effet, la préparation à la scolarisation des participants du programme (en français) a été influencée de façon positive par le nouveau programme préscolaire, comme le démontrent les effets statistiquement significatifs de trois des quatre dimensions de l'ÉPE-AD. Les effets observés du programme étaient spécifiques, car nous n'avons pas observé d'effet là où aucun effet n'était prévu (c.-à-d. pour la dimension d'Habilité physique de l'ÉPE-AD), et généraux, car les résultantes ciblées ont été positivement affectées. Les résultats convergent pour former un patron cohérent en ce qui a trait aux résultantes de la préparation à la scolarisation, qui dépendent toutes fortement de la maîtrise de la langue française. La liste de résultantes ayant subi un effet positif comprend: Communication, Conscience de soi, et Habilités cognitives. Nous avons également observé un effet programme sur la sous-échelle de Vocabulaire expressif composée d'items provenant de l'ÉPE-AD. Les tailles d'effet associées à ces effets observés sont de l'ordre de 0,30 à 0,50 écart-type, valeurs se situant à l'intérieur de l'étendue des tailles d'effet observées dans des études comparables menées aux États-Unis, où les effets rapportés allaient de 0,26 à 0,79 écart-type (pour une revue, voir Howes et coll., 2008).



Dans la présente étude, le groupe programme en garderie a commencé le projet avec un désavantage *non significatif* par rapport au groupe témoin en garderie en ce qui a trait à trois des quatre domaines de l'ÉPE-AD. Le groupe programme en garderie a surmonté ce désavantage initial de façon à démontrer un avantage développemental initial, où des gains développementaux supérieurs à ceux du groupe témoin en garderie ont été observés quatre mois après la mise en œuvre du programme pour trois des quatre domaines de l'ÉPE-AD et pour la sous-échelle de Vocabulaire expressif (premier post-test). Ces gains initiaux sont réapparus, comme prévu, 16 et 20 mois après la mise en œuvre du programme (cinquième et sixième post-tests).

Tel que discuté à la section 6.3, nous avons prévu la réémergence de l'effet programme basé sur l'utilisation, pour ces évaluations, d'une version révisée de l'ÉPE-AD qui permettrait de mieux discriminer les enfants présentant des compétences et des habiletés variées en Communication, Cognition et Conscience de soi (p. ex, l'ÉPE-AD modifiée par la SRDC). L'inscription à l'école (prématernelle, âge de quatre ans) était statistiquement équivalente entre les groupes expérimentaux au début de la deuxième année. Néanmoins, celle-ci a été employée comme covariable afin de contrôler pour son effet qui potentiellement est dû aux taux significativement inférieurs d'inscription à l'école chez les enfants du groupe témoin hors garderie, taux qui ont été observés plus tard au cours de l'année en raison du changement de groupes. En bref, la réapparition d'un effet du traitement au cours de la deuxième année ne peut être attribuée à une association entre l'inscription à l'école et le groupe expérimental. Il est toutefois possible et même probable que les enfants les mieux préparés à la scolarisation profitent mieux de l'environnement scolaire. Dans la mesure où le programme a contribué à une augmentation de la préparation à la scolarisation, cela se traduirait par un effet positif du programme au cours de la deuxième année comparativement aux groupes témoins, toutes autres choses étant égales. En effet, c'est ce que nous avons observé (voir ci-dessous).

Les résultats appuient la conclusion que les enfants ayant reçu l'intervention sont davantage préparés à l'école que leurs pairs du groupe témoin en garderie, lorsqu'ils atteignent l'âge de cinq ans (maternelle). Les enfants qui faisaient partie du groupe programme ont réalisé des gains significativement plus importants quant aux résultantes associées à la préparation à la scolarisation que s'il n'y avait eu aucune intervention. De tels gains sont importants, car on associe une meilleure préparation à la scolarisation à une transition plus facile en classe (Cunha, Heckman, Lochner & Masterov, 2006), où les enfants seront plus susceptibles de profiter pleinement de l'environnement social et pédagogique qui s'offre à eux (Janus & Offord, 2000). Bien sûr, nous nous référons ici qu'aux mesures de préparation à la scolarisation qui ont été administrées en français.

Dans des rapports futurs, nous nous demanderons si ces gains en matière de préparation à la scolarisation s'appliquent également à des précurseurs plus immédiats de réussite scolaire comme la conscience phonologique et les connaissances alphabétiques (Lonigan et coll., 2008; Sénéchal, 2005). Dans la version modifiée de l'ÉPE-AD (sixième et septième évaluations), ces deux concepts sont représentés au niveau de l'échelle d'Habiletés cognitives, et deux des quatre items de l'échelle de Communication évaluent la conscience phonologique. Des impacts positifs du programme testé ont été démontrés pour ces deux échelles. Ainsi, nous anticipons pouvoir observer des effets positifs du programme sur des instruments validés, spécifiquement conçus pour mesurer ces construits. Ces résultats paraîtront dans des rapports ultérieurs.

Des mesures de vocabulaire français ont été prises pour répondre aux objectifs de la recherche concernant la langue française. Le but était de déterminer si le programme testé avait un effet positif sur le développement du vocabulaire français dans un contexte de minorité. Pour la sous-échelle de Vocabulaire expressif construite selon les items de l'ÉPE-AD, nous avons observé des impacts positifs du programme aux cinquième, sixième et septième évaluations. Les analyses ont été menées en reconnaissant d'abord les changements de groupes survenus lors de la deuxième année du projet et ensuite, en fixant la composition des groupes comme elle l'était à la fin de la première année. De façon générale, les deux analyses ont produit des résultats équivalents, ce qui suggère que les changements de groupe survenus au cours de la deuxième année du projet n'ont pas eu d'effets importants sur les résultats. Nous avons observé, exceptionnellement, un effet pour la cinquième évaluation qui n'a émergé que des analyses qui ne tenaient pas compte des changements de groupe survenus à la deuxième année. Dans ce cas particulier, la migration des enfants qui ont quitté le groupe programme en garderie immédiatement avant la cinquième évaluation a eu pour effet d'obscurcir le véritable impact du programme testé. Tout compte fait, les estimations associées à chacune des administrations de l'ÉPE-AD au cours de la deuxième année du programme reflétaient, de façon cohérente, un impact positif sur le développement du Vocabulaire expressif. Aucun effet n'a été observé sur la sous-échelle de Vocabulaire réceptif de l'ÉPE-AD, mais ceci est attribué aux mauvaises propriétés psychométriques de cette mesure.

Les analyses reposant sur les mesures normalisées de Vocabulaire réceptif et expressif, l'ÉVIP-R (septième évaluation) et l'EOWPVT (sixième évaluation), respectivement, suggèrent un effet programme modeste d'environ deux à trois mots qui n'a pas persisté quand toutes les covariables ont été incluses dans le modèle. Ce résultat nul a été obtenu peu importe la définition des groupes expérimentaux (appartenance aux groupes à la fin de la première année v. lors de l'administration des tests). L'absence d'un effet sur ces deux mesures de vocabulaire fait contraste à l'effet positif du programme que nous avons observé relativement aux sous-échelles de Vocabulaire expressif de l'ÉPE-AD. Nous attribuons cette incongruité aux différences méthodologiques de ces deux analyses. Les échelles de Vocabulaire normalisées ont été analysées au moyen d'un estimateur basé sur le contrôle statistique et les échelles de l'ÉPE-AD ont été analysées au moyen de l'estimateur DD. L'estimateur DD aurait permis un meilleur contrôle des caractéristiques au niveau de base si les échelles de vocabulaire normalisées en avaient permis le recours (c.-à-d. le biais de sélection, Cronbach & Furby, 1970; Winship & Morgan, 1999). Des résultats plus encourageants ont été obtenus pour les enfants qui avaient été moins exposés à la langue française au début du projet (voir les sections qui suivent).

### ***Garderie programme versus groupe hors garderie***

Même si le groupe témoin hors garderie a obtenu des scores significativement plus élevés par rapport aux mesures de langue et culture françaises (chapitre 4), le groupe programme en garderie a maintenu une trajectoire développementale statistiquement équivalente sur les résultantes de l'ÉPE-AD durant les deux premières années du projet (avant et après les ajustements pour les caractéristiques linguistiques). Toutefois, les analyses comparant les deux groupes n'ont indiqué à aucun moment un impact du programme statistiquement significatif. Cela est vrai même si les effets de la communauté ont fait l'objet d'un contrôle expérimental et qu'une variété d'autres caractéristiques sur la famille et l'enfant a été soumise à un contrôle statistique. Les principales analyses des résultantes développementales ne démontrent pas de

façon probante que les enfants ayant reçu l'intervention se distinguent à cet égard de leurs pairs du groupe hors garderie. Les effets positifs du programme se sont plutôt manifestés dans les analyses par profil linguistique.

### **7.1.2. Effets du programme en fonction du profil linguistique**

La littérature sur le bilinguisme fait une distinction claire entre les cas où l'acquisition d'une langue seconde profite au développement général d'un enfant et les cas où le développement semble en être affecté (p. ex., Landry, Allard, Deveau, 2009). Ces différents types de bilinguisme sont appelés respectivement « additif » et « soustractif ». La distinction cruciale entre les deux types de bilinguisme fait référence à la question suivante : est-ce que la *langue maternelle* est suffisamment développée pour soutenir l'acquisition d'une deuxième langue sans entraîner de retard dans le développement approprié à l'âge des habiletés cognitives ou des compétences dans la langue maternelle (UNESCO, 2010).

Cette condition minimale ne semble pas être remplie pour beaucoup d'enfants francophones vivant en milieu minoritaire, une population chez qui on a parfois observé des retards par rapport à leurs pairs pour des résultantes comme la performance en lecture (Bussière, Cartwright et al., 2001; Chartier et al., 2008; Conseil canadien sur l'apprentissage, 2008; Education Quality and Accountability Office, 2007). Les causes sous-jacentes à cet écart de performance sont raisonnablement bien comprises. La littérature sur le bilinguisme les explique en faisant appel à un mécanisme assez clair : l'exposition à la langue. Pour que le bilinguisme soit additif, un seuil minimal d'exposition ou d'utilisation de la langue maternelle doit être dépassé (pour une revue, voir Pearson 2007). Pour diverses raisons complexes (p. ex., motivationnelles, plus grande exposition à la langue de la majorité dans des milieux multiples, Landry et al., 2009), ce seuil minimal est plus élevé lorsque la langue maternelle est une langue minoritaire (Pearson et al., 1997; Vihman, Lum, Tierry, Nakai, & Keren-Portnoy, 2006). En conséquence, les enfants francophones en milieu minoritaire qui sont élevés comme bilingues ont besoin d'un soutien spécial pour améliorer leurs chances d'atteindre un bilinguisme additif.

C'est dans cet esprit que nous avons étudié la possibilité que les enfants qui ont été exposés à des langues autres que le français (habituellement l'anglais) profiteraient le plus du programme pilote préscolaire. Nous référons ici explicitement aux contributions conjointes des volets d'alphabétisation en famille ou en garderie en ce qui a trait à l'exercice d'une influence sur l'environnement linguistique des enfants. Nous avons tenu compte du fait que ces enfants courent le risque de développer un bilinguisme de type soustractif, étant donné qu'ils vivent en milieu linguistique minoritaire, lorsque nous avons prévu que les effets les plus grands du programme seraient précisément observés auprès de cette population parce qu'il a été conçu pour répondre à leurs besoins. Cette prédiction a été confirmée dans une série d'analyses où les effets du programme ont été estimés pour les enfants présentant divers degrés d'exposition de base à la langue française.

#### ***Garderies programmes versus garderies témoins***

À la fin de la première année (c.-à-d. au troisième post-test), la plupart des enfants du projet (52 %) étaient inscrits à l'école soit à temps partiel ou à temps plein (âge de quatre ans, prématernelle). Il est utile de s'interroger si les enfants du groupe programme en garderie étaient plus préparés à la scolarisation à ce moment-là qu'ils ne l'auraient été en l'absence du traitement.

Les principales analyses des effets du programme n'ont pas fourni de données probantes irréfutables à cet effet. Toutefois, les analyses subséquentes ont révélé que ces résultats mitigés étaient en partie attribuables à la variabilité des réponses au traitement. Les enfants ayant reçu l'intervention semblent effectivement être mieux préparés à la scolarisation, mais les dimensions touchées par le programme dépendent des caractéristiques de l'enfant.

Nous rapportons des résultats indiquant que les enfants *moins exposés* au français au début du projet en ont bénéficié le plus pour ce qui est de l'acquisition des compétences linguistiques de base telles qu'elles ont été mesurées par les échelles de Vocabulaire expressif et de Communication. Ceci s'est avéré être vrai lorsque la plupart des enfants de l'échantillon sont entrés à la prématernelle (âge de quatre ans) et cet avantage a été de nouveau observé lorsque la quasi-totalité des enfants de l'échantillon sont entrés en maternelle (à l'âge de cinq ans), soit à la fin de la deuxième année du projet (selon les échelles ÉPE-AD et EOWPVT). *Le Rapport de mise en œuvre du projet* (Bérubé et al., 2014) et les résultats signalés dans le chapitre 5 du présent rapport démontrent clairement que la qualité des activités de *lecture* constitue une caractéristique importante pour distinguer le programme testé de ce qui est normalement offert dans un programme de garderie de langue française. Les effets observés du programme pourraient s'expliquer par cette caractéristique du programme testé. La spécificité apparente de l'effet quant au langage de type de expressif est cohérente avec d'autres évaluations d'interventions préscolaires (c.-à-d. Hargrave & Sénéchal, 2000) qui insistent sur l'importance d'activités linguistiques interactives (p. ex., la lecture dialoguée), tandis que la spécificité apparente de l'effet du programme quant aux compétences linguistiques est cohérente avec les études qui ont été effectuées auprès de francophones et qui ont tenu compte de cette variable comme variable modératrice (c.-à-d. Maltais, 2007).

En revanche, les enfants admis dans l'étude avec une *forte exposition* au français et, conséquemment, avec une facilité relative à la langue plus élevée étaient en position de développer leurs compétences associées à l'échelle d'Habilités cognitives. Les données présentées ici suggèrent que les compétences linguistiques de base facilitent l'obtention de gains secondaires en habiletés cognitives (p. ex., littératie). Cette explication trouve un appui auprès des théories développementales sur l'acquisition des compétences (Bloom, 1976; Vygotsky, 1978) et avec l'importance que revêt la maîtrise de la langue pour le succès scolaire (Cummins, 1979; Doherty, 1997; Hindman, Skibbe, Miller, and Zimmerman, 2010). Elle est aussi cohérente avec les constatations de Maltais (2007) qui a signalé que l'effet d'un programme de prématernelle pleine journée sur le développement linguistique dépendait du profil linguistique de l'enfant, mais que les avantages cognitifs plus larges (c.-à-d. gains en habiletés de lecture) s'observaient même pour les enfants dont l'exposition au français était forte, enfants pour qui les effets étaient davantage importants.

En somme, le haut degré de qualité du programme testé a mené à une transmission plus efficace de compétences et de connaissances. Les enfants ont simplement tiré du programme des avantages déterminés selon leur préparation à apprendre certains types particuliers de compétences dans un environnement francophone. Un patron généralement similaire de résultats a été obtenu dans les analyses de la première année et de la deuxième année, ce qui tend à indiquer que les effets précoces ont persisté même lorsqu'ils échappaient à la détection dans certaines analyses statistiques. Pris dans leur ensemble, ces résultats ont des implications quant à la manière dont le programme est mis en œuvre dans d'autres contextes, parce que les avantages

attendus du programme semblent varier qualitativement et quantitativement selon l'exposition linguistique de la clientèle.

### ***Garderies programme versus groupe témoin hors garderie***

Au départ, nous avons fait l'hypothèse que le programme aurait des effets positifs comparativement aux deux groupes témoin. Cette prévision était fondée sur le fait que les deux volets de l'intervention (garderie programme, Ateliers d'alphabétisation familiale) avaient été conçus pour susciter un changement dans l'environnement des enfants à la garderie et à la maison, respectivement. Les effets attendus selon nos hypothèses n'ont pas été observés dans les principales analyses relatives au groupe témoin hors garderie. L'absence d'un effet dans ces analyses est peut-être attribuable au fait que le groupe témoin hors garderie a profité de degrés relativement élevés d'exposition au français au cours de la période de base (voir tableau 4.4). En fait, les effets attendus ont été relevés lorsqu'une analyse plus subtile a été effectuée, qui a permis aux effets du programme de varier en fonction de l'exposition au français.

Dès la fin de la première année, les enfants « à risque » présentant une faible exposition au français lorsqu'ils ont commencé à recevoir l'intervention s'étaient développés plus rapidement au niveau de la dimension de Communication que les enfants hors garderie avec un niveau d'exposition similaire. Pratiquement le même patron de résultats a été obtenu peu importe la variable qui a été utilisée comme indicateur de l'exposition de l'enfant au français. Toutefois, la langue parlée par la mère à l'enfant a mené aux effets « conditionnels » les plus constants dans le temps. L'exposition linguistique de base des enfants dans un milieu hors garderie (p. ex., à la maison) est nécessairement un déterminant plus important du développement que pour les enfants qui passent une période de temps substantielle en garderie. Ceci explique ainsi le rôle plus important que joue cette caractéristique en tant que modérateur des effets du programme lors des comparaisons relatives à ce groupe. Dans la deuxième année du projet, le groupe programme a démontré un vocabulaire expressif supérieur selon l'EOWPVT. Encore une fois, la spécificité apparente de l'effet sur le langage de type expressif est cohérente avec d'autres évaluations d'interventions préscolaires (Hargrave & Sénéchal, 2000).

En somme, les résultats ont fourni des évidences appuyant les répercussions positives du programme sur le développement des enfants francophones, que ce soit les enfants de garderies régulières ou du groupe témoin hors garderie qui aient été pris comme base de comparaison. Pour les enfants les plus exposés au risque de bilinguisme de type soustractif, le programme semble avoir son effet le plus marqué sur les compétences linguistiques et en communication. Les enfants qui étaient les moins à risque ont profité de meilleures résultantes cognitives provenant de l'exposition à un programme de garderie de meilleure qualité.

## **7.2 RÉSULTANTES DES PARENTS**

L'attention individuelle qu'une éducatrice en service de garde peut accorder à un enfant donné est limitée (Wasik, 2008), ce qui explique pourquoi certains ont fait valoir que les parents sont une ressource importante dans toute intervention de soins à l'enfant (Reese et coll., 2010). Les ateliers pour les parents ont été conçus pour mettre cette ressource au profit du développement de la langue et de la culture francophones, et nous avons pris un certain nombre de résultantes en considération pour en évaluer le succès.

Des effets robustes ont été observés pour les Connaissances et l'Auto-efficacité déclarées par les parents. L'intensité de ces effets a varié en fonction des mesures de qualité des ateliers (c.-à-d. le style d'animation de l'intervenante et la longueur de la séance) et de la participation des parents (pour une conclusion similaire, voir Jordan, Snow & Porche, 2000). L'ensemble convergent de résultats obtenus pour le programme Atelier qui ont été produits dans des analyses distinctes selon la participation (participant versus non-participant) et selon la qualité du programme Atelier augmente la crédibilité de l'effet programme, établissant fermement le lien entre les ateliers pour parents et le changement d'attitude chez les parents qui ont reçu l'intervention. Établir ce lien est la première étape pour démontrer que les ateliers ont influencé indirectement le développement des enfants par l'entremise des parents et de l'environnement qu'ils ont créé à la maison.

Une fois le lien atelier-parent établi, on pourrait élargir la question de recherche pour déterminer si la participation des parents ou le score de changement des parents est significativement lié aux résultantes développementales des enfants. Démontrer un tel lien compléterait la chaîne causale qui se termine par les effets positifs sur les résultantes des enfants. Pour l'instant, nous concluons que le lien entre la participation aux ateliers et les attitudes parentales a été établi. Nous présumons que le lien existe entre les résultantes développementales des enfants et les attitudes parentales en nous basant sur des évaluations antérieures de l'efficacité d'interventions où la combinaison d'un programme pour enfants et d'ateliers d'alphabétisation (Brooks-Gunn et coll., 2000; Reese et coll., 2010) a produit des effets plus marqués que si l'un ou l'autre des volets avaient été pris de façon isolée. Un manque de puissance statistique nous a empêchés de tester formellement le caractère plausible de la structure causale décrite pour l'échantillon du projet Capacité d'apprentissage. Dans des rapports subséquents, un test de médiation formel sera effectué (p. ex. Baron & Kenny, 1986) afin de décomposer les effets programme observés selon leur source, soit, dans ce cas, les programmes en garderie et les ateliers d'alphabétisation. Ces analyses pourront être menées lorsque les données de la deuxième cohorte d'enfants pourront y être incluses.<sup>104</sup>

La plupart des autres résultantes (c.-à-d. la fréquence des activités de littératie, comportements de modélisation, croyances quant au développement de l'enfant) ont été déformées par des problèmes d'expérimentation et de mesure, ce qui explique pourquoi aucun effet programme n'a été observé avec ces mesures. La conséquence est qu'un véritable changement des comportements en matière d'alphabétisation a pu avoir lieu en plus des avantages associés sur les résultantes touchant les enfants (Bus, van IJzendoorn & Pelligrini, 1995; Scarborough & Dobrich, 1994). Les résultats équivoques ne tranchent aucunement, ni d'une façon ni de l'autre. Un facteur contribuant au résultat nul est peut-être le fait que les sujets présentés durant les ateliers étaient trop généraux. Les ateliers d'alphabétisation tendent à produire des effets constants lorsqu'ils ciblent le développement de compétences spécifiques et concrètes par l'entremise d'un programme qui consacre plus de temps à des sujets précis que celui offert par le projet Capacité d'apprentissage (Sénéchal, 2006; pour une liste d'exemples, voir Reese et al., 2010).

Nous avons observé un effet positif, comparativement au groupe témoin hors garderie, sur la variable «langue des activités de littératie» au troisième post-test seulement. Cet effet était

---

<sup>104</sup> Présentement, l'échantillon de participants aux ateliers des quatre communautés qui ont servi aux analyses d'impact est de moins de 50.

intéressant, en ceci qu'il n'a pas été observé lors des sondages mi-atelier, mais seulement lors du sondage suivant. Le résultat est compatible avec le fait que la question de la langue des activités de littératie a été abordée vers la fin de la série d'ateliers (c.-à-d. à l'atelier 6). Il serait plus facile d'interpréter le résultat comme un effet programme si l'on avait observé un avantage correspondant lors de la comparaison avec le groupe témoin en garderie.

### **7.3 EFFETS DIFFÉRENCIÉS DU DOSAGE**

La littérature concernant les services de garde commencent à aborder la question cherchant à déterminer si l'effet de l'exposition aux services de garde (p. ex. les heures passées chaque semaine en garderie) varie selon la qualité du programme offert (Vandell, 2004). En effet, on observe des signes que les garderies de grande qualité atténuent les effets comportementaux négatifs associés à de longues heures passées en garderie (National Institute of Child Health et Human Development Early Child Care Research Network, 2003). Il est possible que la qualité ait un effet modérateur similaire sur les associations négatives parfois observées entre le dosage et les résultantes cognitive ou linguistique (Vandell, 2004). Toutes les parties prenantes (c.-à-d. parents, enfants, éducatrices, évaluatrices) ont été informées du programme, des effets prévus et des personnes visées par l'intervention. Cette conscience peut avoir contribué, de diverses façons, aux résultats observés. En revanche, toutes les parties prenantes étaient plus ou moins aveugles à la variable de dosage et à ses effets prévus. Les effets observés relativement à la variable de dosage sont par conséquent exempts des formes les plus évidentes de biais dus aux attentes.

Dans la mesure où le programme testé a mené à un programme en garderie de meilleure qualité, il est raisonnable de se demander si l'effet de l'exposition à la garderie varie en fonction de l'appartenance des enfants au groupe programme en garderie ou au groupe témoin en garderie. En effet, les analyses présentées à la section 5.1.2 ont révélé que la différence observée quant au domaine de la Communication entre les deux groupes en garderie lors du premier post-test avait augmenté de façon directement proportionnelle aux heures passées en garderie. Pour cette résultante du moins, nous observons des effets du dosage qui sont différenciés selon le groupe expérimental, ce qui donne plus de crédibilité aux impacts du programme rapportés. Dans toutes les autres analyses (c.-à-d. autres résultantes de la première année et analyses de toutes les résultantes de la deuxième année), le degré d'exposition à la garderie ne permettait pas de prédire les résultantes associées à la préparation à la scolarisation, *au-delà de ce qui pouvait déjà être expliqué par l'appartenance au groupe expérimental*.

### **7.4. FIDÉLITÉ ET QUALITÉ DES GARDERIES**

Toutes les garderies varient quant à la qualité du programme qu'elles offrent. Le programme testé a été conçu pour répondre aux besoins spécifiques des familles et des enfants francophones dans le contexte d'un programme de haute qualité. Nous avons prédit que les garderies assignées au nouveau programme offrirait un programme de meilleure qualité que celles qui ont été assignées au groupe témoin. Comme nous l'avons affirmé au chapitre 3, cette division des garderies en deux groupes est une simplification d'une dimension continue qui peut être mesurée directement : la qualité du programme.

Plus la qualité de l'enseignement augmente, plus les enfants exposés au programme sont susceptibles d'assimiler les concepts ciblés (Howes et coll., 2008), soit, dans le cas présent, la langue et la culture françaises. En tant que simplification, l'appartenance au groupe expérimental ne capture pas l'entière variabilité dans la qualité qui existe entre les garderies, bien qu'elle puisse en expliquer une grande partie (voir le tableau 5.7). Par exemple, certaines garderies témoins offraient un service d'une qualité relativement élevée, tandis que certaines garderies programme mettaient en œuvre le programme moins efficacement. Les analyses descriptives ont indiqué que tant les garderies programme que les garderies témoins ont fourni une qualité de service qui se situait dans l'étendue supérieure (*Rapport de mise en œuvre du projet*, 2014). On pourrait arguer sur cette base que tout changement de qualité causé par l'intervention est superflu. Un tel argument irait à l'encontre des recherches indiquant que l'effet de la qualité sur les résultantes développementales de l'enfant n'est pas linéaire. C'est dans l'étendue de qualité moyenne à supérieure que les différences quant à la qualité des programmes de garderie ont l'incidence la plus forte (Burchinal, Vandergrift, Pianta, & Mashburn, 2009).

Les statistiques descriptives et inférentielles ont confirmé que les garderies du groupe programme se distinguaient de façon positive quant aux dimensions de fidélité et de qualité du programme. Tant la fidélité structurelle que la fidélité de contenu étaient plus élevées dans le groupe programme en garderie que dans les autres groupes, mais la différence la plus marquée a été observée, dans cette analyse à faible puissance, pour la fidélité structurelle. Parmi les indices de qualité considérés ici (qualité structurelle, qualité éducative, sensibilité des éducatrices et lecture), les plus grands avantages des garderies programme par rapport aux garderies témoins ont été observés pour la qualité structurelle et la qualité de lecture. En somme, tous les indices de qualité et de fidélité soutiennent la validité de notre hypothèse selon laquelle l'appartenance au groupe expérimental (c.-à-d. garderie programme versus garderie témoin) constitue une substitution adéquate de la qualité du programme en garderie.

Malgré cette cohérence, nous avons néanmoins procédé à des analyses de suivi où on a employé les mesures de fidélité et de qualité directement comme prédicteurs des résultantes développementales des enfants. L'objectif de ces analyses était de contribuer à l'ensemble des données probantes à l'appui de la validité de construit de l'étude. Les résultats de ces analyses tendaient effectivement à confirmer les résultats des analyses de la première année par groupe expérimental (chapitre 5). Les effets les plus constants ont été observés avec la *fidélité structurelle*, ce qui semble contredire les résultats d'une étude rapportant que la fidélité structurelle permet d'expliquer peu de variance additionnelle lorsqu'un grand assortiment de covariables est inclus dans le modèle (Howes et coll., 2008). La même liste de covariables n'était pas disponible pour les analyses présentées ici, ce qui empêche toute comparaison définitive. Nous suggérons qu'en l'absence de mesures de contrôle adéquates, les indices de fidélité structurelle capturent à la fois la fidélité structurelle et d'autres concepts tels le niveau de « ralliement » des parties prenantes au sein d'une garderie donnée. Cette explication permettrait de justifier l'effet imprévu associé à l'Habilité physique lors du troisième post-test, qui était le seul effet d'impact impliquant cette dimension. Notons également que la qualité de lecture était fortement corrélée avec la fidélité structurelle et la fidélité de contenu (voir le tableau 5.5). En résumé, ces indices ne constituent pas une mesure pure des construits visés, mais appuient néanmoins la validité de l'étude.

La *fidélité de contenu* était aussi associée aux résultantes développementales des enfants. La fidélité de contenu était particulièrement associée à la Communication, à la Conscience de soi et



au Vocabulaire expressif. Aucune association n'a été observée avec les Habilités cognitives, ce qui est compatible avec le fait que le contenu du programme visait spécifiquement les résultantes concernant la langue française plutôt que le développement cognitif en tant que tel.

De façon plus générale, les *effets de la qualité du programme* ont été observés pour les échelles de Communication, de Conscience de soi et de Vocabulaire expressif. Ces associations aux compétences relatives au langage de type expressif sont compatibles avec les travaux antérieurs menés auprès d'enfants anglophones (Mashburn et coll., 2008). Il est intéressant de noter que la qualité structurelle et la qualité éducative étaient significativement liées au Vocabulaire expressif, tandis que la qualité éducative et la sensibilité des éducatrices étaient liées à la Communication. Cette légère dissociation entre les résultats soutient l'idée que la sensibilité des éducatrices est principalement associée aux aspects sociaux de la communication (Peisner-Feinberg et coll., 2001), alors que les divers éléments structurels d'une garderie visent des objectifs plus concrets tels le développement du vocabulaire. La sensibilité des éducatrices est importante, car leur relation avec les enfants constitue le principal mécanisme qui permet aux ressources du programme d'atteindre ces derniers (Howes et coll., 2008). Par exemple, les effets négatifs associés à la taille plus importante des classes peuvent s'expliquer en partie par la diminution concomitante de la sensibilité des éducatrices observée dans ce contexte (Phillipsen, Burchinal, Howes & Cryer, 1997). L'impact du programme testé sur cette dimension suggère que les résultantes développementales peuvent être améliorées si nous concentrons nos efforts sur cette dimension de la qualité en classe.

Dans l'analyse des résultantes de la deuxième année, la fidélité globale et la qualité globale du programme (moyennes basées sur certains des indices plus précis discutés ci-dessus) ont permis de prédire le développement sur l'échelle d'Habilités cognitives, au milieu de la deuxième année. Étant donné que cette échelle contient des items qui évaluent la conscience phonologique et la connaissance alphabétique, nous pouvons conclure que la fidélité et la qualité du programme peuvent prédire des précurseurs importants de l'acquisition de la littératie et de la réussite académique subséquente (Belsky et coll., 2007). Nous attribuons l'association observée entre la qualité globale et la Communication à une cause semblable, soit les deux items de conscience phonologique inclus dans cette échelle. Il est à noter que les deux autres items sont partagés avec la sous-échelle de Vocabulaire expressif, qui n'est pas liée à la qualité globale. En dernier lieu, nous avons observé une association entre la qualité globale et les mesures normalisées de Vocabulaire réceptif et expressif : ÉVIP-R et EOWPVT (voir Mashburn et coll., 2008). Nous n'avons pas observé une association similaire pour la fidélité globale, ce qui suggère que le succès du programme nécessite davantage que de simplement cocher une liste d'exigences de base à respecter. Les normes de qualité associées à la mise en œuvre de ces exigences sont aussi importantes. Pris dans leur ensemble, les résultats appuient notre théorie selon laquelle le programme testé manipule les bons leviers pour influencer les résultantes développementales.

## **7.5 LIMITES ET TRAVAUX FUTURS**

La théorie écologique de Bronfenbrenner (1979) fournit un cadre riche pour explorer diverses contingences mises en cause dans le succès d'un programme de garderie comme celui qui est évalué ici. En testant si l'ampleur des répercussions du programme variait selon l'exposition linguistique, nous avons seulement commencé à explorer égratigner la surface. Les mesures de

Vitalité ethnolinguistique, prises au niveau des communautés, peuvent aussi contribuer à modérer les effets du programme. La combinaison, dans un lieu donné, d'une faible exposition au français et d'une faible vitalité de la communauté francophone pourrait agir de façon à augmenter les avantages du programme. Un examen approprié de tels effets complexes nécessiterait un échantillon plus grand que celui disponible pour le présent rapport. Ces questions pourront être examinées plus raisonnablement lorsque les données de la deuxième cohorte seront disponibles. L'ajout de participants pourrait aussi nous permettre de distinguer les contributions indépendantes du volet garderie et Ateliers familles du programme à l'aide d'autres stratégies d'analyse des données (p. ex., tests de médiation, analyses des coefficients de direction).

Une deuxième limite de la présente étude concerne la manière dont la préparation à la scolarisation a été mesurée. Du fait de leur nature même, les outils d'évaluation de la préparation à la scolarisation, comme l'Instrument de mesure du développement de la petite enfance (IMDPE) et l'ÉPE-AD, ne sont pas bien étalonnés pour mesurer les répercussions du programme. Ils conviennent mieux pour dépister une population d'enfants lorsque l'objectif est d'identifier ceux qui sont à un risque de présenter un retard développemental et qui ont besoin d'une « stimulation » pour acquérir un ensemble de compétences fonctionnelles approprié à leur groupe d'âge. Tout ce qui est nécessaire pour atteindre cet objectif est une mesure générale du développement, ce qui correspond aux forces de l'IMDPE et de l'ÉPE-AD. Toutefois, cette force devient une faiblesse lorsque l'objectif change : évaluer l'impact d'un programme. Dans ce cas, les questions de recherche se centrent sur des effets qui peuvent être plus subtils (c.-à-d. des effets programme potentiellement petits plutôt que de grands retards ou écarts de performance), et les compétences spécifiques touchées par le programme testé sont d'un intérêt certain. En bref, un portrait plus précis du développement est nécessaire, soit un portrait qui nous éclaire non seulement au sujet du développement général, mais aussi sur les compétences (mesurées avec précision) qui prédisent le mieux les résultantes ultérieures (p. ex., l'acquisition de mots et les connaissances phonologiques nous éclairent sur les compétences en lecture ultérieures qui sont à leur tour un bon prédicteur du succès scolaire) (Hirsh-Pasek, Kochanoff, Newcombe & de Villiers, 2005).

Une troisième limite de l'étude concerne la validité externe. La question à savoir si les effets rapportés du programme se reproduiront au sein d'autres populations d'enfants francophones d'âge préscolaire est limitée par la nature même de l'intervention. Le programme testé a été mis en œuvre dans des garderies francophones, situées principalement dans des communautés urbaines et caractérisées par des niveaux intermédiaires à élevés de Vitalité ethnolinguistique (sauf pour Durham). De plus, la plupart des familles participant au projet étaient canadiennes et elles comprenaient peu de nouveaux immigrants francophones (voir le tableau 4.5 du chapitre 4). D'autres travaux sont nécessaires pour dégager les impacts d'un tel programme sur une population minoritaire francophone vivant en région rurale, sur les francophones vivant dans des milieux très minoritaires, de même que sur une population de nouveaux immigrants de minorité francophone dont plusieurs déclarent parler plus d'une première langue.

Enfin, bien que des analyses exhaustives ont été menées pour vérifier la robustesse des résultats aux diverses spécifications des modèles, la petite taille d'échantillon ( $n < 100$  par groupe expérimental) et le petit nombre de garderies ( $n < 30$ ) tempèrent quelque peu notre confiance dans les résultats. Cela est particulièrement vrai étant donné que dans le projet Capacité d'apprentissage, comme pour toute recherche quasi-expérimentale, la possibilité qu'un facteur confondant non contrôlé soit la cause des résultats ne peut pas être écartée définitivement. La

recommandation globale du présent rapport est qu'une nouvelle analyse des données, comprenant les participants de la première et de la seconde cohorte, soit effectuée pour confirmer la stabilité des estimations.

## 7.6 CONCLUSION

L'analyse d'impact présentée dans ce document a révélé *des effets programme positifs tant sur les résultantes concernant les enfants que les parents*. En ce qui concerne les résultantes touchant les enfants, les effets positifs se sont manifestés comme des gains développementaux supérieurs relativement aux groupes témoins en garderie et hors garderie, aussi bien lors du premier post-test qu'au milieu de la deuxième année du projet pour certains enfants. La nature de ces gains dépendait de l'exposition de l'enfant au français au début du projet. Les enfants ayant l'exposition la plus faible ont profité le plus du programme en terme de développement linguistique (p. ex., vocabulaire expressif), tandis que les enfants présentant l'exposition la plus forte ont connu des gains pour ce qui est de leur développement cognitif. Le volet du programme s'adressant aux parents a réussi à influencer certaines dimensions de leurs attitudes. Les parents ont rapporté des gains quant à leurs Connaissances et à leur sentiment d'Auto-efficacité après avoir assisté aux ateliers, ce qui signifie qu'une partie des impacts observés sur les enfants peut être attribuée au volet Ateliers familles de l'intervention.

Nous pouvons donc conclure que le programme testé a eu un impact modeste sur la préparation à la scolarisation des enfants de minorité francophone. Dans des rapports ultérieurs, les analyses bénéficieront d'une plus grande stabilité et puissance statistiques en raison de l'ajout d'une deuxième cohorte d'enfants à l'échantillon. Ces rapports futurs se pencheront sur la question cherchant à déterminer si les gains apparents sur les dimensions de la préparation à la scolarisation et de la maîtrise de la langue française entraîneront des avantages à moyen et à long terme. La première étape pour répondre à une telle question consistera à déterminer si le programme a eu un impact positif sur des précurseurs plus immédiats de la réussite scolaire tels la conscience phonologique, la connaissance alphabétique et le vocabulaire (Lonigan, 2008). Cette question fera l'objet de la deuxième phase du projet Capacité d'apprentissage.

## Références

- 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*. Ottawa, ON: Direction générale de la recherche appliquée, Politique stratégique, Développement des ressources humaines Canada.
- Arnold, D. H., Lonigan, C. J., Whitehurst, G. J., & Epstein, J. N. (1994). Accelerating language development through picture-book reading: Replication and extension to a videotape training format. *Journal of Educational Psychology*, 86(2), 235-243.
- 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.
- 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.
- Baylor College of Medicine. (2001). Guide for using the classroom observation tool.
- Becker, G. S. (1993). Nobel lecture: The economic way of looking at behaviour. *The Journal of Political Economy*, 101(3), 385-409.
- Behrman, J. R., Cheng, Y., & Todd, P. E. (2004). Evaluating preschool programs when length of exposure to the program varies: A nonparametric approach. *The Review of Economics and Statistics*, 86(1), 108 - 132.
- Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., & Owen, M. T., & The NICHD Early Child Care Research Network. (2007). Are there long-term effects of early child care? *Child Development*, 78(2), 681-701.
- Benoît, S. (s.d.). *Questionnaire utilisés par CIRCEM dans l'étude de l'impact de l'impact de l'alphabétisation familiale sur les familles vivant en milieu minoritaire*. 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*. Center for Research on Child Wellbeing, Princeton, NJ: Princeton University.
- 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.
- Bertrand, J. (2007). Programmes d'enseignement efficaces au préscolaire. Commentaires sur Kagan et Kauerz et sur Schweinhart. In R. E. Tremblay, R. G. Barr, & R. de V. Peters (Eds.), *Encyclopédie sur le développement des jeunes enfants* [online] (pp. 1-7). Montréal, QC: Centre d'excellence pour le développement des jeunes enfants. Available online at: <http://www.child-encyclopedia.com/pages/PDF/BertrandFRxp.pdf>.

- Bérubé, A., Legault, L., Janisse, L., Carson, R., Saucier, M.-A. et Lefebvre, L. (2014). *Capacité d'apprentissage dans les communautés francophones en situation minoritaire : Rapport de mise en œuvre du projet*. Ottawa (Ontario) : Société de recherche sociale appliquée.
- Bloom, B. S. (1976). *Human characteristics and school learning*. New York, NY: McGraw-Hill.
- Bornstein, L., & Bornstein, M. H. (Eds.). (2007). Pratiques parentales et développement social de l'enfant. Dans R. E. Tremblay, R de V. Peters, M. Boivin et R. G. Barr (éd). *Encyclopédie sur le développement des jeunes enfants* [sur Internet]. Montréal, QC : Centre d'excellence pour le développement des jeunes enfants; 2006 : 1-2. Disponible en ligne au <http://www.enfant-encyclopedie.com/Pages/PDF/BornsteinFRxp.pdf>.
- 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* (2<sup>nd</sup> ed., pp. 132-224). New York, NY: HarperCollins.
- 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. Dans W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of huma 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? Dans J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2<sup>nd</sup> ed., pp. 549-588). New York, NY: Cambridge University Press.
- Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *The Future of Children*, 7(2), 55-71.
- Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A., (2009). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*. 25, 166-176.
- 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.
- Bus, A. G., van IJzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research*, 65(1), 1-21.
- Bussière, P., Cartwright, F., Crocker, R., Ma, X., Oderkirk, J., & Zhang, X. (2001). *À la hauteur: la performance des jeunes du Canada en lecture, en mathématiques et en sciences*. Ottawa, ON: Développement des ressources humaines Canada, Statistique Canada, Conseil des ministres de l'Éducation (Canada).
- Centre for Family Literacy. (2002). *Statements of good practice for family literacy programs in Alberta*. Edmonton, AB: Author.

- 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 boy. *Journal of School Psychology, 42*(3), 201–220.
- Chartier, M. J., Dumaine, J., Daudet-Mitchell, L., Gosselin, J., & Vielfaure, K. (2008). Vivre en français à la petite enfance et apprendre à l'école française en milieu minoritaire, y a-t-il un lien? *Enfants en santé Manitoba, Division scolaire franco-manitobaine*. Document soumis pour publication.
- 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* (2<sup>nd</sup> ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioural sciences* (3<sup>rd</sup> ed.). Hillsdale, NJ: Psychology Press.
- Commission nationale des parents francophones. (2005). *Le meilleur est...avenir! Dans N. Lafrenière-Davis (Ed.), La petite enfance dans les communautés francophones en milieu minoritaire au Canada : une analyse transformative de la situation*. Ottawa, ON. Consulté le 2 février, 2009. Disponible en ligne au [http://cnpf.ca/documents/SCAN\\_FRANCAIS\\_2005.pdf](http://cnpf.ca/documents/SCAN_FRANCAIS_2005.pdf).
- 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*. Ottawa, ON. Direction générale de la recherche appliquée, Politique stratégique, Développement des ressources humaines Canada. Disponible en ligne au <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. Dans M. H. Bornstein (Ed.), *Handbook of parenting. Volume 4: Applied and practical parenting* (pp. 437-458). Mahwah, NJ: Lawrence Erlbaum Associates.
- Conseil canadien sur l'apprentissage. (2007). *Rapport sur l'apprentissage au Canada : rapport sur l'état de l'apprentissage chez les jeunes enfants*. Ottawa, ON.
- Conseil canadien sur l'apprentissage. (2008). *Literacy among Canadian students in minority language contexts*. Unpublished report.
- Corbeil, J.-P., & Blaser, C. (2007). *Le portrait linguistique en évolution : Recensement de 2006*. Ottawa, ON: Statistique Canada. Disponible en ligne au <http://www12.statcan.ca/census-recensement/2006/as-sa/97-555/pdf/97-555-XIE2006001-fra.pdf>.
- Corbeil, J.-P., Grenier, C., & Lafrenière, S. (2007). *Les minorités prennent la parole : résultats des l'Enquête sur la vitalité des minorités de langue officielle 2006*. Consulté le 11

- décembre 2007. Disponible en ligne au <http://www.statcan.gc.ca/pub/91-548-x/91-548-x2007001-fra.pdf>.
- Corter, C., & Pelletier, J. (2005). Parents and community involvement in schools: Policy panacea or pandemic? Dans N. Bascia, A. Cumming, A. Datnow, K. Leithwood, & D. Livingstone (Eds.), *International Handbook of Educational Policy* (pp. 295 - 327). Great Britain: Springer.
- Cronbach, L. J., & Furby, L. (1970). How we should measure "change" – or should we? *Psychological Bulletin*, 74(1), 68-80.
- Cummins, J. (1979). *Cognitive academic language proficiency, linguistic interdependence, the optimum age question and some other matters*. Working Papers in Bilingualism, No. 19, 12–129.
- Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. Dans E. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (1st ed., Vol. 1, pp. 307-451). Amsterdam: North-Holland: Elsevier.
- D'Amours, Y. (2010). *La scolarité des francophones et des anglophones, à travers les groupes d'âge, au Québec et en Ontario: données sociodémographique en bref*. Québec, QC. Disponible en ligne au [http://www.stat.gouv.qc.ca/publications/conditions/pdf2010/sociodemobref\\_fev10.pdf](http://www.stat.gouv.qc.ca/publications/conditions/pdf2010/sociodemobref_fev10.pdf).
- 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. Dans 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*. Montréal, QC : Insitut de la statistique Québec. Disponible en ligne au [http://www.jesuisjeserai.stat.gouv.qc.ca/pdf/publications/feuille/fascicule\\_ecole\\_bon\\_pied.pdf](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*. Présenté au 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 : Direction générale de la recherche appliquée, Politique stratégique, Développement des ressources humaines Canada. Disponible en ligne au <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.
- 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* (pp. 147): 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. Disponible en ligne au [http://www.eqao.com/pdf\\_e/09/Cpr PJ9e\\_0909\\_web.pdf](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. Consulté le 23 mars 2010. Disponible en ligne au [http://www.who.int/child\\_adolescent\\_health/documents/pdfs/lancet\\_child\\_dev\\_series\\_paper3.pdf](http://www.who.int/child_adolescent_health/documents/pdfs/lancet_child_dev_series_paper3.pdf).
- Erikson, E. H. (1994). *Identity and the life cycle*. New York, NY: W. W. Norton & Company.
- Fédération canadienne pour l'alphabétisation en français. (2007). *Apprendre ça commence à la maison : guide de pratiques exemplaires en alphabétisation familiale en contexte francophone minoritaire*. Ottawa, ON : Auteur.
- Forgues, É., & Landry, R. (2006). *Définitions de la francophonie en situation minoritaire : analyse de différentes définitions statistiques et leurs conséquences*. Ottawa, ON: Institut Canadien de recherche sur les minorités linguistiques.
- 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. (s.d.). Questionnaire utilisés dans l'Évaluation du programme alpha District 3, Edmundston, Nouveau-Brunswick.
- Gilbert, A. (2003). *Une vision nationale : la petite enfance : porte d'entrée à l'école de langue française : Rapport final de la recherche*. Ottawa, ON: Centre interdisciplinaire de recherche sur la citoyenneté et les minorités, Université d'Ottawa en partenariat avec la Fédération canadienne des enseignantes et des enseignants.
- Glymour, M. M., Weuve, J., Berkman, L. F., Kawachi, I., & Robins, J. M. (2005). When is baseline adjustment useful in analyses of change? An example with education and cognitive change. *American Journal of Epidemiology*, 162(3), 267-278.



- Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, 26(3), 499-510.
- Guimond, L. (2003). *Revue des travaux sur la vitalité des communautés francophones du Canada*. Ottawa: Université d'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., & Cryier, D. (1998). Early Childhood Environment Rating Scale: Revised Edition (ECERS-R). New York, NY: Teachers College Press.
- Haveman, R., & Wolfe, B. (1995). The determinants of children's attainments: A review of methods and résultats. *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), 3-5.
- 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.
- Hirsh-Pased, K., Kochanoff, A., Newcombe, N. S., de Villiers, J. (2005). Using scientific knowledge to inform preschool assessment: Making the case for “empirical validity”. *Social Policy Report*, 19(1), 1-20.
- Horizon Research. (2000). Inside the classroom: Observation and analytic protocol. Consulté janvier 2008. Disponible en ligne au <http://www.horizon-research.com/instruments/clas/cop.pdf>.
- 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., et al. (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* (1<sup>st</sup> ed.). New York, NY: Routledge.
- Human Resources and Social Development Canada (HRSDC), & Statistics Canada. (2005). *Building on our competencies: Canadian results of the International Adult Literacy and Skills Survey*. Ottawa, ON: Statistics Canada.
- Institute de la statistique Québec. (2003). *En 2002...j'avais 5 ans! Questionnaire informatisé rempli par l'intervieweuse (QIRI) – E6*. Québec, QC.

- 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.
- Janus, M., & Offord, D. (2000). Readiness to learn at school. *ISUMA: Canadian Journal of Policy Research*, 1(2), 71-75.
- 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 IRPP*, 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.
- Kézdi, G. (2004). Robust estimation des erreurs-types in fixed-effects panel models. *Hungarian Statistical Review, Special number 9*, 95-116.
- 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.
- 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 pré-test*. Papier présenté au le Annual meeting of the American Sociological Association. Disponible en ligne au [http://www.allacademic.com/meta/p\\_mla\\_apa\\_research\\_citation/1/0/8/1/9/p108194\\_index.html](http://www.allacademic.com/meta/p_mla_apa_research_citation/1/0/8/1/9/p108194_index.html).
- Lamb, T. A., & Tschillard, R. (2005). Evaluating learning in professional development workshops: Using the retrospective pré-test. *Journal of Research in Professional Learning*. Disponible en ligne au <http://www.nsdsc.org/library/publications/research/lamb.pdf>.
- 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 ou le nombre le justifie... V*. Research report for the Commission nationale des parents francophones. Canadian Institute for Research on Linguistic Minorities. Moncton, NB.
- 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. Disponible en ligne au <http://www.telugu.quebec.ca/diverscite/entree.htm>.

- Landry, R., Allard, R., & Deveau, K. (2007a). A macroscopic intergroup approach to the study of ethnolinguistic development. *International Journal of the Sociology of Language*, 185, 225-253.
- Landry, R., Allard, R., & Deveau, K. (2007b). 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.
- Legault, L., Mák, G., Verstraete, M. et Bérubé, A. (2014). *Capacité d'apprentissage dans les communautés francophones en situation minoritaire : Rapport de référence*. Ottawa (Ontario) : Société de recherche sociale appliquée.
- Lemelin, J.-P., & Boivin, M. (2007). *Mieux réussir dès la première année : l'importance de la préparation à l'école*. Québec, QC : Institut de la statistique Québec. Disponible en ligne au <http://www.jesuisjeserai.stat.gouv.qc.ca/pdf/publications/feuille/Fasc2Vol4.pdf>.
- LeTouzé, S. (2006). *Pour mon enfant d'abord : étude de l'impact de l'alphabétisation familiale sur les familles vivaant en milieu minoritaire. Étape 2 : Cohortes 2 et 3 : 2005-2006*. Ottawa, ON: Centre interdisciplinaire de recherche sur la citoyenneté et les minorités.
- 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. Papier présenté au le Effective Early Learning Programs: Research, Policy and Practice, Banff, ON.
- Lonigan, C. J., Shanahan, T., Westberg, L., & The National Early Literacy Panel. (2008). Identification of children's skills and abilities linked to later outcomes in reading, writing, and spelling. Dans *Developing early literacy: Report of the National Early Literacy Panel*. (pp. 55–101). Jessup, MD: National Institute for Literacy. Disponible en ligne au <http://www.nifl.gov/publications/pdf/NELPReport09.pdf>.
- 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*, 33(2), 6–18.
- Martel, A. (2001). *Droits, écoles et communautés en milieu minoritaire : 1986 - 2002. Analyse pour un aménagement du français par l'éducation*. Ottawa, ON : Commissariat aux langues officielles.
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., ... Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79(3), 732-749.
- Masny, D. (2006). Le développement de l'écrit en milieu de langue minoritaire : l'apport de la communication orale et des habiletés métalinguistiques. *Éducation et francophonie*, 34(2), 126-149.

- Meisels, S. J., & Atkins-Burnett, S. (2006). Evaluating early childhood assessments: A differential analysis. Dans K. McCartney & D. Phillips (Eds.), *The Blackwell handbook of early childhood development* (pp. 533-549). Malden, MA: Blackwell Publishing.
- Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. San Francisco, CA: Jossey-Bass.
- Monheit, A. C., & Steinburg Schone, B. (2004). How has small group market reform affected employee health insurance coverage? *Journal of Public Economics*, 88(1-2), 237-254.
- Moore, T. (2005). *Towards improved support for families of young children: The role of communities*. Papier présenté au 9th Australian Institute of Family Studies Conference. Disponible en ligne au <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 pre-test 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 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.
- National Institute of Child health and Human Development Early Child Care Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten. *Child Development*, 74(4), 976-1005.
- Neuman, S.D., & Celano, D. (2001). Access to print in low-income and middle-income communities: An ecological study of four neighborhoods. *Reading Research Quarterly*, 36(1), 8-26.
- Office de la qualité et de la responsabilité en éducation. (2009). *Rendement des élèves de l'Ontario : rapport provincial de l'OQRE sur les résultats des tests en lecture, écriture et mathématiques, cycle primaire (de la 1re à la 3e année) et cycle moyen (de la 4e à la 6e année) et du test de mathématiques, 9e année, 2008-2009. Élèves de langue française*. Toronto, ON.
- Pearson, B. (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). Input factors in lexical learning of bilingual infants (ages 10 to 30 months). *Applied Psycholinguistics*, 18(1), 41-58.
- Peisner-Feinberg, E. S., Burchinal, M.R., Clifford, R.M, Culkin, M.L., Howes, C., Kagan, S.L., et al. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, 72(5), 1534-1553.
- 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.
- Phillipsen, L. C., Burchinal, M. R., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly, 12*(3), 281-303.
- 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.
- Ressources humaines et Développement des compétences Canada (2006). *Projet pilote de garde d'enfants : principes directeurs du projet pilote, 2006*. Disponible en ligne au [http://www.RHDCC.gc.ca/fra/sm/comm/ds/garde\\_enfants/2006/principes\\_directeurs.shtml](http://www.RHDCC.gc.ca/fra/sm/comm/ds/garde_enfants/2006/principes_directeurs.shtml).
- Rockwell, S. K., & Kohn, H. (1989). Post-the-pre evaluation. *Journal of Extension, 27*(2). Disponible en ligne au <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? Dans R. E. Tremblay, R. de V. Peters, M. Boivin et R. G. Barr (éd.), *Encyclopédie sur le développement des jeunes enfants* [sur Internet] (pp. 1-14). Montréal, QC : Centre d'excellence pour le développement des jeunes enfants. Disponible en ligne au <http://www.enfant-encyclopedie.com/Pages/PDF/Sanders-MorawskaFRxp.pdf>.
- Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review, 14*(3), 245-302.
- Sénéchal, M. (2005). Alphabétisation, langue et développement affectif. Dans R. E. Tremblay, R. G. Barr, R. de V. Peters (éd.), *Encyclopédie sur le développement des jeunes enfants* [sur Internet] (pp. 1-6). Montréal, QC : Centre d'excellence pour le développement des jeunes enfants. Disponible en ligne au <http://www.enfant-encyclopedie.com/Pages/PDF/SenechalFRxp.pdf>.
- Sénéchal, M. (2006). Testing the home literacy model: Parent involvement in kindergarten is differentially related to grade 4 reading comprehension, fluency, spelling, and reading for pleasure. *Scientific Studies of Reading, 10* (1), 59-87.
- 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.
- Shrigley, R. L. (1990). Attitude and behavior are correlates. *Journal of Research in Science Teaching, 27*(2), 97-113.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage Publications.

- Statistique Canada. (2004). *Revue trimestrielle de l'éducation*. 9(4). Disponible en ligne au <http://www.statcan.gc.ca/pub/81-003-x/81-003-x2002004-fra.pdf>.
- Statistique Canada. (2006). *Enquête longitudinale nationale sur les enfants et les jeunes : Matériel d'enquête pour la collecte de données 2004-2005. Cycle 6. Livre 1 : Parent, enfant et jeune*.
- Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5<sup>th</sup> 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. Dans R. E. Tremblay, R. G. Barr, R. de V. Peters, M. Boivin (éd.), *Encyclopédie sur le développement des jeunes enfants* [sur Internet] (pp. 1-11). Montréal: Centre d'excellence pour le développement des jeunes enfants. Disponible en ligne au <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 pré-tests. *American Journal of Evaluation*, 30(1), 31-43.
- Thomas, E. M. (2006). *La disposition à apprendre à l'école pour les jeunes de cinq ans au Canada*. Consulté le 17 février 2008. Disponible en ligne au <http://www.statcan.gc.ca/pub/89-599-m/89-599-m2006004-fra.pdf>.
- 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.
- UNESCO. (2010). *Enhancing learning of children from diverse language backgrounds: Mother tongue-based bilingual or multilingual education in early childhood and early primary school years*. Available online at: <http://www.ecdip.org/docs/pdf/UNESCO%20Mother-tongue%20based%20EY%202010.pdf>.
- Vandell, D. L. (2004). Early child care: The known and the unknown. *Merrill Palmer Quarterly Journal of Developmental Psychology*, 50(3), 387-414.
- Venetsanou, F., & Kambas, A. (2010). Environmental factors affecting preschoolers' motor development. *Early Childhood Education Journal*, 37, 319-327.
- Vihman, M. M., Lum, J., Tierry, 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. (2008). When fewer is more: Small groups in early childhood classrooms. *Early Childhood Education Journal*, 35(6), 515-521.



- Wasik, B.A., Bond, M.A., & Hindman, A. (2006). The effect 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.
- Whitehurst, G. J., Falco, F. L., Lonigan, C., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca M. C., & Caulfield, M. (1988). Accelerating language development through picture book reading. *Developmental Psychology*, 24(4), 552-558.
- 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.
- Williams, R. L. (2000). A note on robust variance estimation for cluster-correlated data. *Biometrics*, 56(2), 645-646.
- Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, 25, 659-707.
- Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W., & Wolf, M. (2007). The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills, word identification skills, and reading comprehension by children with reading disabilities. *Journal of Speech, Language, and Hearing Research*, 50, 1093-1109.

## Annexe A: Calendrier et outils d'évaluation

Le tableau présenté ci-dessous fournit le calendrier des évaluations des enfants. De plus, on y retrouve, pour chacune des vagues de collecte de données, le construit à l'étude (*en italique*) et l'outil utilisé pour évaluer son développement auprès des enfants.

**Table A1: Calendriers des évaluations des enfants**

octobre 2007 (Niveau de base)	février 2008 (1 <sup>er</sup> post-test)	juin 2008 (2 <sup>e</sup> post-test)	octobre 2008 (3 <sup>e</sup> post-test)	février 2009 (4 <sup>e</sup> post-test)	juin 2009 (5 <sup>e</sup> post-test)	octobre 2009 (6 <sup>e</sup> post-test)
<i>Préparation à la scolarisation</i>						
ÉPE-AD <sup>105</sup> : Conscience de soi (A)						
ÉPE-AD : Habiletés cognitives (B)						
ÉPE-AD : Langage et Communication (C)						
ÉPE-AD : Habiletés physique et motrice (D)						
<i>Vocabulaire expressif</i>						
Sous-échelle de l'ÉPE-AD						
					EOWPVT-R	
<i>Vocabulaire réceptif</i>						
Sous-échelle de l'ÉPE-AD						
						ÉVIP-R

<sup>105</sup> Évaluation de la petite enfance — Appréciation directe (Willms, 2007)



## **Annexe B : Procédure pour l'administration de l'ÉPE–AD (Mesure pré-intervention)**

Les évaluatrices qui ont administré l'ÉPE–AD (le EYE-DA en anglais) aux enfants ont été recrutées à compter de la fin de l'été 2007. Elles ont reçu de la SRSA une formation théorique et pratique d'une durée d'environ six heures aux mois d'août, septembre et, pour la communauté d'Orléans, octobre. Cette formation visait, en plus de la présentation du protocole d'administration du test, la présentation du projet Capacité d'apprentissage ainsi que des procédures relatives à la confidentialité. Les évaluatrices se sont engagées par contrat à respecter le protocole d'administration et de confidentialité. Les étapes complètes du protocole pour l'administration du test sont :

1. Les évaluatrices appellent les parents pour prendre les rendez-vous des évaluations à la maison ou pour les prévenir du moment de l'évaluation en garderie. Ces appels ont aussi pour but de confirmer l'âge en mois de l'enfant ainsi que les réponses aux questions concernant les langues parlées avec les parents et les amis pour le domaine E (*Conscience et engagement envers la culture francophone*).
2. Les évaluatrices et la coordonnatrice communautaire prennent contact avec les garderies participantes pour définir l'horaire et aménager un endroit dans la classe favorable à une bonne évaluation.
3. L'évaluatrice applique la méthode « médicale », c'est-à-dire qu'elle attend de vérifier que l'enfant soit le bon avant de remplir les renseignements d'identification sur le questionnaire papier.
4. L'évaluatrice s'adresse à l'enfant dans sa langue maternelle tout d'abord, puis applique le protocole pour déterminer la langue du test.
5. L'évaluatrice suit les indications de notation fournies par le concepteur de l'outil, ce qui veut dire qu'elle arrondit les points à l'entier inférieur dans le but de réaliser une évaluation prudente.
6. L'évaluatrice encourage l'enfant, mais n'offre pas d'indices sur les réponses à donner, sauf si le protocole l'indique.
7. Si l'enfant est fatigué en cours de test, l'évaluatrice s'arrête et peut recommencer plus tard au début du domaine où elle s'était arrêtée.
8. À la fin de l'évaluation, l'évaluatrice remet à l'enfant un autocollant pour le remercier de sa participation.
9. Dans les cas où l'enfant ne voudrait vraiment pas participer, l'évaluatrice doit tenter d'évaluer l'enfant au moins une autre fois (un autre jour).

Le point 4 est définitivement un point clé pour l'évaluation puisque le respect ou le non-respect de cette règle peut influencer les résultats de manière considérable. Le protocole d'administration pour déterminer la langue du test, suggéré par M. Willms (présenté le 8 décembre 2006 et révisé en juillet 2007), est présenté au tableau B.1.

**Tableau B1 Arbre décisionnel pour déterminer la langue d'administration du test**

<i>Étapes</i>	<i>Protocole</i>
1. Administration du domaine E	Administrier les six questions du domaine E directement à <u>l'enfant, et au parent pour les questions E4 à E6</u> , au début du test.
2. Arbre de décision pour déterminer la langue du test (domaine E)	Si le score est supérieur à 6 pour le domaine E, administrer le reste de l'ÉPE-AD en français. Si le score est inférieur ou égal à 6 pour le domaine E, évaluer le domaine C en français et en anglais.
3. Arbre de décision pour déterminer la langue du test (domaine C)	Si le score en français pour ce domaine est supérieur à 14, le reste de l'évaluation se déroule en français. Si le score en français pour ce domaine est inférieur ou égal à 14 et le score en anglais est inférieur ou égal à 14, le reste du test se déroule également en français. Si le score pour le domaine C est inférieur ou égal à 14 en français, mais le score est supérieur à 14 en anglais, le reste de l'évaluation se fait en anglais.

Également, il y a deux points de départ pour l'administration du test, dépendant de l'âge de l'enfant. Si l'enfant est âgé de moins de quatre ans, l'évaluatrice débute avec le premier item pour le domaine. Autrement, elle débute plus loin dans le test, avec l'option de retourner au point de départ si l'enfant présente des difficultés. Le Rapport de référence n'a pas fait mention de cette règle puisque tous les enfants étaient âgés de moins de quatre ans lors de l'évaluation de base.

## **Annexe C : Une comparaison des différentes versions de l'ÉPE-AD**

Cette section concerne le contenu de trois versions de l'ÉPE-AD (le EYE-DA en anglais) utilisées au cours des deux premières années de mise en œuvre du projet Capacité d'apprentissage. De nouvelles versions de l'ÉPE-AD ont été créées en réponse à deux enjeux notés à l'automne 2008 et suite aux décisions prises lors d'une rencontre avec le Comité consultatif de RHDCC qui a eu lieu le 17 décembre 2008. Le premier enjeu fait référence au risque élevé d'un effet plafond soulevé par les projections de scores pour la 5e évaluation (février 2009) à partir des résultats aux 3e et 4e évaluations. Les services du concepteur de l'outil ont donc été retenus par RHDCC pour créer de nouveaux items plus difficiles permettant ainsi de suivre la trajectoire développementale des enfants quant aux domaines A, B, et C. Le deuxième enjeu concerne la nécessité d'ajouter une mesure d'habiletés linguistiques afin de faire une discrimination plus fines des enfants. Bien que l'instrument ÉPE-AD mesure bien la préparation à scolarisation des enfants, il n'était pas suffisamment « sensible » pour capter les dimensions développementales chez les enfants et il ne mesure qu'en surface leurs habiletés de communication.

Pour les fins de la discussion qui suit, nous nommons « ÉPE-AD initiale » l'outil utilisé en février 2009 projet Capacité d'apprentissage et nous nommons « ÉPE-AD étendue » l'outil administré au cours de l'hiver 2009 (comprenant les questions plus difficiles développées par Willms et excluant les questions plus faciles de l'ÉPE-AD initiale). Enfin, nous nommons « ÉPE-AD modifiée » la version de l'outil qui a été retravaillée par la SRSA et qui a été utilisée lors des évaluations menées de juin à octobre 2009.

### ***Domaines mesurés par l'ÉPE-AD initiale***

L'ÉPE-AD conçue par Doug Willms comprend quatre domaines et un cinquième conçu spécifiquement pour le projet Capacité d'apprentissage :

- Domaine A = Conscience de soi;
- Domaine B = Habiletés cognitives;
- Domaine C = Langue et communication;
- Domaine D = Habiletés physique et motrice; et
- Domaine E = Conscience et engagement envers la culture francophone.

### ***Domaines mesurés par l'ÉPE-AD modifiée***

L'ÉPE-AD modifiée comprend sept dimensions, notamment:

- Vocabulaire expressif;
- Conscience de soi;
- Conscience phonologique;

- Numératie;
- Mémoire pour les informations personnelles;
- Connaissances alphabétiques; et
- Lecture orale.

Ces échelles permettent de poursuivre la trajectoire développementale des enfants pour les domaines A, B et C. La version de l'ÉPE-AD modifiée comprend la totalité des items du domaine A (moins la question A18). Pour le domaine B, la totalité des items a été conservée, soit dans les sous-échelles de Conscience phonologique, de Numératie, ou de Connaissances alphabétiques. Enfin, le domaine C est en partie conservé dans la sous-échelle de Vocabulaire expressif, avec laquelle il présente une corrélation de 0,87, confirmant ainsi qu'il sera possible de suivre la trajectoire développementale de ce domaine.

Les tableaux suivants offrent une comparaison des items retrouvés dans les trois versions de l'ÉPE-AD. La version modifiée de l'ÉPE-AD a été conçue en français seulement; les items présentés dans les tableaux sont des verbatims issus de cette évaluation. Le lecteur est prié de noter que les items non retenus dans la version de l'EPE-AD modifiée ne se retrouvent pas dans le tableau. Ces items incluent ceux éliminés par Willms dans la version de l'EPE-AD étendue.

**Tableau C1 : Sous-échelle de Vocabulaire expressif**

Items	ÉPE-AD initiale	ÉPE-AD étendue	ÉPE-AD modifiée
<b>Vocabulaire expressif</b>			
Comment s'appelle cet objet? <i>unseau/une chaudière, des boutons, des chandelles, une horloge/un réveille-matin</i>	<b>A11</b>	<b>x</b>	<b>V1</b>
Pointez chacune des images, une à la fois. <b>Peux-tu me nommer quatre couleurs?</b> <b>Peux-tu me nommer quatre fruits?</b> <b>Peux-tu me nommer quatre animaux?</b> <b>Peux-tu me nommer quatre vêtements?</b>	<b>A12</b>	<b>x</b>	<b>V2</b>
Comment se nomme cette partie du corps? <i>le menton, le coude, le poignet, l'épaule</i>	<b>A14</b>	<b>x</b>	<b>V3</b>
Qu'est-ce que cette personne fait comme métier? <i>le policier, l'enseignant(e), le fermier/jardinier, le médecin</i>	<b>A15</b>	<b>x</b>	<b>V4</b>
Qu'est-ce qu'on utilise pour écrire? pour s'asseoir? pour se promener? pour balayer?	<b>C8</b>	<b>x</b>	<b>V5</b>
Nomme le plus d'animaux possible.	<b>C9 (nomme huit animaux)</b>	<b>x</b>	<b>V6</b>

Items	ÉPE–AD initiale	ÉPE–AD étendue	ÉPE–AD modifiée
<b>Vocabulaire expressif</b>			
Quel temps fait-il dans chacune des images? <i>Il vente, il pleut, il neige, il fait soleil</i>		<b>A21</b>	<b>V7</b>
Comment s'appelle ceci? <i>thermomètre, tasse à mesurer, règle, balance</i>		<b>A22</b>	<b>V8</b>

*Note* : L'échelle de Vocabulaire expressif (six items) est corrélée à 0,87 aux items du domaine C de l'échelle ÉPE–AD étendue (C7 à C14). Nous avons conservé uniquement les items qui demandent à l'enfant de nommer un objet, de façon à être cohérent avec la méthodologie des tests de vocabulaire conventionnels, tel l'EOWPVT.

**Tableau C2 : Sous-échelle de Conscience de soi**

Items	ÉPE-AD initiale	ÉPE-AD étendue	ÉPE-AD modifiée
<b>Conscience de soi</b>			
Que devrais-tu faire quand tu : <b>es fatigué, as faim, as froid, as soif?</b>	<b>A7</b>	<b>x</b>	<b>CS1</b>
La souris est petite; l'éléphant est ____ ( <b>gros ou grand</b> ). Cet oiseau est dehors : cet oiseau est ____ ( <b>en dedans ou à l'intérieur</b> ). Ce pot à biscuits est plein; ce pot à biscuits est ____ ( <b>vide</b> ). Cette échelle est courte; cette échelle est ____ ( <b>longue</b> ).	<b>A8</b>	<b>x</b>	<b>CS2</b>
Placer des objets pour démontrer sa compréhension de : <b>premier, dernier, devant, derrière</b> .	<b>A10</b>	<b>x</b>	<b>CS3</b>
Nommer le moment de la journée (le matin, l'après-midi, la soirée, la nuit) associé à deux de ces situations communes : <b>des étoiles dans le ciel, prendre le petit déjeuner, retourner à la maison après l'école</b>	<b>A13</b>	<b>x</b>	<b>CS4</b>
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?		<b>A17</b>	<b>CS5</b>
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?		<b>A20</b>	<b>CS6</b>

**Tableau C3 : Sous-échelle de Conscience phonologique**

Items	ÉPE–AD initiale	ÉPE–AD étendue	ÉPE–AD modifiée
<b>Conscience phonologique</b>			
<p><i>Écoute-moi prononcer le mot jouet : jouet commence par le son /j/.</i></p> <p><b>Par quel son commence le mot piano?</b>  <b>Par quel son commence le mot tomate?</b>  <b>Par quel son commence le mot boîte?</b>  <b>Par quel son commence le mot rouge?</b></p>		<b>C16</b>	<b>CP1</b>
<p>Dis-moi si les mots riment, s'ils finissent par le même son.</p> <p><b>pomme/homme; boîte/lune; chien/chat;</b>  <b>pain/main</b></p>	<b>B11</b>	<b>x</b>	<b>CP2</b>
<p><b>Trouve deux mots qui commencent par /b/ comme dans « ballon ».</b></p>	<b>B13</b>	<b>x</b>	<b>CP3</b>
<p>Dis-moi si les mots suivants commencent par le même son :</p> <p><b>mère/lait, balle/beau, plat/clou, peau/pas</b></p>	<b>B14</b>	<b>x</b>	<b>CP4</b>
<p><i>Écoute pendant que je prononce le mot canif. Le mot canif se termine par le son /f/.</i></p> <p><b>Quel est le son qui termine le mot cheval?</b>  <b>Quel est le son qui termine le mot album?</b>  <b>Quel est le son qui termine le mot autobus?</b>  <b>Quel est le son qui termine le mot neuf (9)?</b></p>		<b>C19</b>	<b>CP5</b>
<p><i>Voici un bateau (une pomme, une table, une fenêtre et du lait). Le mot bateau commence par un son /b/. Écoute /b/, bateau.</i></p> <p><b>Quelle image commence par le son t?</b>  <b>Quelle image commence par le son f?</b>  <b>Quelle image commence par le son l?</b>  <b>Quelle image commence par le son p?</b></p>		<b>B18</b>	<b>CP6</b>

*Note :* Pour l'item C16, nous demandons seulement de dire par quel son commence le mot. Nous ne demandons pas d'identifier la lettre. Le but étant de mesurer la conscience phonologique, nous tentons de garder cette mesure pure. Les mots « piano » et « tomate » ont été ajoutés pour arriver à un score maximum de 4, tout comme avec les autres versions du test.

**Tableau C4 : Sous-échelle de Numératie**

Items	ÉPE-AD initiale	ÉPE-AD étendue	ÉPE-AD modifiée
<b>Numératie</b>			
Dire le nombre de parties : <b>Un chat a combien de queues?</b> <b>Un chien a combien de pattes?</b> <b>Un oiseau a combien d'ailes?</b> <b>Ta main a combien de doigts?</b>	<b>A9</b>	<b>x</b>	<b>N1</b>
Compte toutes les étoiles et dis-moi combien il y en a.	<b>B9</b>	<b>x</b>	<b>N2</b>
Regarde chaque chiffre et dis-moi lequel est <u>le plus</u> grand. <b>6 ou 8, 12 ou 10, 9 ou 7, 11 ou 12.</b>	<b>B10</b>	<b>x</b>	<b>N3</b>
Mets ensemble des jetons pour faire un groupe de... <b>5, 7, 8, 9.</b>	<b>B12</b>	<b>x</b>	<b>N4</b>
<b>Montre-moi le chiffre...</b> 13, 20, 45, 112.		<b>B19</b>	<b>N5</b>

*Note :* Pour l'item B19, l'ordre des chiffres a été changé sur l'image afin que les chiffres ne soient pas présentés dans le même ordre que demandés.

**Tableau C5 : Mémoire pour les informations personnelles**

Items	ÉPE-AD initiale	ÉPE-AD étendue	ÉPE-AD modifiée
<b>Mémoire pour les informations personnelles</b>			
Peux-tu me dire quel âge tu as et quels sont le jour et le mois de ton anniversaire (de ta fête)?	<b>A16</b>	<b>x</b>	<b>MP1</b>
Savoir où l'on habite : <b>Quel est le nom de la rue où tu habites?</b> <b>Quel est le nom de la ville où nous habitons?</b> <b>Quel est le nom de notre province?</b> <b>Quel est le pays où nous habitons?</b>		<b>A19</b>	<b>MP2</b>



**Tableau C6 : Sous-échelle de connaissances alphabétiques**

Items	ÉPE-AD initiale	ÉPE-AD étendue	ÉPE-AD modifiée
<b>Connaissances alphabétiques</b>			
<b>Dire le son de la lettre majuscule</b>	<b>B16</b>	<b>B17</b>	<b>Livret</b>
<b>Dire la lettre majuscule</b>	<b>B15</b>		
<b>Dire la lettre minuscule</b>		<b>B20</b>	<b>Livret</b>

NOTE : Les lettres sont présentées en ordre de difficulté croissante pour les francophones canadiens. Nous avons présenté un total de 26 lettres, incluant celles avec des accents « é, è, ê ». Chaque réponse correcte obtient un score de 1. L'étendue des scores est donc, pour B17 et B20, de 0 à 26.

**Tableau C7 : Lecture orale**

Items	ÉPE-AD initiale	ÉPE-AD étendue	ÉPE-AD modifiée
<b>Lecture orale</b>			
<b>Lire huit mots perçus de façon globale à fréquence élevée.</b> Dis-moi quel mot est écrit ici.	<b>B17</b>	<b>B16</b>	<b>L1</b>
		<b>Chien</b> <b>Oui</b> <b>Non</b> <b>Balle</b> <b>Chat</b> <b>Le</b> <b>La</b> <b>Nez</b>	<b>Le</b> <b>La</b> <b>Oui</b> <b>Non</b> <b>Balle</b> <b>Maman</b> <b>Nez</b> <b>Papa</b>

**Note :** Nous avons changé l'ordre des mots de façon à ce que leur fréquence d'utilisation en français soit décroissante. Nous avons également changé deux mots pour les remplacer par d'autres mots plus communs en français. Les mots anglais « dog » et « cat » font souvent partie des livres d'alphabet anglophones alors que leurs équivalents en français présentent une orthographe plus complexe (c.-à-d. les lettres « ch » (pour chien et chat) constituent un symbole de plusieurs lettres pour représenter un seul son). Nous avons substitué ces mots par les mots « maman » et « papa », qui sont plus long, mais plus fréquents dans les livres francophones pour enfants.

# **SOCIÉTÉ DE RECHERCHE SOCIALE APPLIQUÉE**

## **BUREAU D'OTTAWA**

55, rue Murray, bureau 400  
Ottawa (Ontario) K1N 5M3  
Tél. : 613.237.4311  
Télééc. : 613.237.5045

## **BUREAU DE TORONTO**

481, avenue University, bureau 705  
Toronto (Ontario) M5G 2E9  
Tél. : 416.593.0445  
Télééc. : 647.725.6293

## **BUREAU DE VANCOUVER**

128, rue Pender Ouest, bureau 301  
Vancouver (Colombie-Britannique) V6B 1R8  
Tél. : 604.601.4070  
Télééc. : 604.601.4080

[www.srdc.org](http://www.srdc.org)