

Annexes au rapport « Quels programmes de prévention du harcèlement pour la Fédération Wallonie-Bruxelles ? »

Morgane Senden, Benoit Galand

UCLouvain

Novembre 2018

Avec le soutien du Ministère de la Fédération Wallonie-Bruxelles

Annexes

Table des matières des annexes

Annexe 1 : Evaluation de <i>KiVa</i>	1
Implantation de <i>KiVa</i> en Finlande	1
Qualité de l'évaluation de <i>KiVa</i> lors de la phase d'évaluation RCT de 2007-2008	2
Effets de <i>KiVa</i> lors de la phase d'évaluation RCT de 2007-2008	7
Qualité de l'évaluation de <i>KiVa</i> lors de la phase d'évaluation RCT de 2008-2009	11
Effets de <i>KiVa</i> lors de la phase d'évaluation RCT de 2008-2009	16
Qualité de l'évaluation de <i>KiVa</i> lors de la phase de dissémination nationale de 2009-2010	20
Effets de <i>KiVa</i> lors de la phase de dissémination nationale de 2009-2010	21
Implantation de <i>KiVa</i> aux Pays-Bas	23
Qualité de l'évaluation de <i>KiVa</i> aux Pays-Bas	23
Effets de <i>KiVa</i> aux Pays-Bas	27
Implantation de <i>KiVa</i> au Royaume-Uni	32
Qualité de l'évaluation de <i>KiVa</i> au Royaume-Uni	32
Effets de <i>KiVa</i> au Royaume-Uni	33
Implantation de <i>KiVa</i> en Italie	35
Qualité de l'évaluation de <i>KiVa</i> en Italie	35
Effets de <i>KiVa</i> en Italie	37
Résumé des effets de <i>KiVa</i>	39
Annexe 2 : Evaluation de <i>Friendly Schools</i>	42
Implantation de <i>Friendly Schools</i>	42
Qualité de l'évaluation de <i>Friendly Schools</i> (2000-2002)	42
Effets de <i>Friendly Schools</i>	45
Implantation de <i>Friendly Schools Friendly Families</i>	48
Qualité de l'évaluation de <i>Friendly Schools Friendly Families</i> (2002-2004)	48
Effets de <i>Friendly Schools Friendly Families</i> selon l'évaluation de 2002-2004	52
Implantation de <i>Cyber Friendly Schools</i>	57
Qualité de l'évaluation de <i>Cyber Friendly Schools</i> (2010-2012)	57
Effets de <i>Cyber Friendly Schools</i> selon l'évaluation de 2010-2012	59
Résumé des effets de <i>Friendly Schools</i>	61
Annexe 3 : Evaluation de <i>Zero</i>	62
Implantation de <i>Zero</i> en Norvège	62

Annexes

Qualité de l'évaluation pilote de <i>Zero</i>	63
Effets de <i>Zero</i> selon l'évaluation pilote	64
Qualité de l'évaluation de 2003-2004	65
Effets de <i>Zero</i> selon l'évaluation de 2003-2004	67
Implantation de <i>Zero</i> en Irlande	69
Qualité de l'évaluation de <i>Zero</i> en Irlande	69
Effets de <i>Zero</i> selon l'évaluation en Irlande	71
Résumé des effets de <i>Zero</i>	72
Annexe 4 : Evaluation de <i>ViSC</i>	73
Implantation de <i>ViSC</i> en Autriche	73
Qualité de l'évaluation de 2000-2001	74
Effets de <i>ViSC</i> selon l'évaluation de 2000-2001	75
Qualité de l'évaluation de 2009-2010	76
Effets de <i>ViSC</i> selon l'évaluation de 2009-2010	80
Implantation de <i>ViSC</i> en Allemagne	83
Qualité de l'évaluation en Allemagne	83
Effets de <i>ViSC</i> selon l'évaluation en Allemagne	86
Implantation de <i>ViSC</i> à Chypre	88
Qualité de l'évaluation de <i>ViSC</i> à Chypre	88
Effets de <i>ViSC</i> selon l'évaluation à Chypre	91
Implantation de <i>ViSC</i> en Roumanie	93
Qualité de l'évaluation de <i>ViSC</i> en Roumanie	93
Effets de <i>ViSC</i> selon l'évaluation en Roumanie	95
Implantation de <i>ViSC</i> en Turquie	96
Qualité de l'évaluation de <i>ViSC</i> en Turquie	97
Effets de <i>ViSC</i> selon l'évaluation en Turquie	99
Résumé des effets de <i>ViSC</i>	101
Annexe 5 : Evaluation de <i>Second Step</i>	103
Implantation de <i>Steps to Respect</i>	103
Qualité de l'évaluation de 2000-2001	103
Effets de <i>Steps to Respect</i> selon l'évaluation de 2000-2001	108
Qualité de l'évaluation de 2008-2009	112
Effets de <i>Steps to Respect</i> selon l'évaluation de 2008-2009	117

Annexes

Implantation de <i>Second Step</i>	121
Qualité de l'évaluation de 2010-2011 dans les écoles secondaires	121
Effets de <i>Second Step</i> selon l'évaluation de 2010-2011 dans les écoles secondaires	125
Autres études portant sur <i>Second Step</i>	127
Résumé des effets de <i>Second Step</i>	130
Annexe 6 : Evaluation de <i>Zippy's Friends</i>	132
Implantation de <i>Zippy's Friends</i> en 1998 au Danemark	132
Implantation de <i>Zippy's Friends</i> en 2000 au Danemark et en Lituanie	132
Implantation de <i>Zippy's Friends</i> en 2004 au Royaume-Uni (Southampton)	134
Implantation de <i>Zippy's Friends</i> en 2004 en Chine (Hong Kong)	135
Implantation de <i>Zippy's Friends</i> en 2006 au Canada (Québec)	136
Implantation de <i>Zippy's Friends</i> en 2007 en Norvège	137
Implantation de <i>Zippy's Friends</i> en 2008 en Irlande	138
Implantation de <i>Zippy's Friends</i> en 2009 au Royaume-Uni (Northamptonshire)	140
Implantation de <i>Zippy's Friends</i> en 2011 en Belgique	141
Implantation de <i>Zippy's Friends</i> en 2012 aux Pays-Bas	142
Implantation de <i>Zippy's Friends</i> en 2014 au Royaume-Uni (pour les enfants à besoins spécifiques)	143
Implantation de <i>Zippy's Friends</i> en 2015 en République Tchèque	143
Résumé des effets de <i>Zippy's Friends</i>	145

Annexe 1 : Evaluation de *KiVa*

KiVa a tout d'abord été évalué en Finlande. Les effets du programme ont été testés via deux études contrôlées puis, suite aux résultats encourageants de ces évaluations, le programme a été diffusé à très grande échelle dans le pays. Là encore, les effets du programme ont été évalués. Le programme s'est ensuite exporté dans plusieurs pays et a été évalué aux Pays-Bas, au Royaume-Uni et en Italie. Dans la suite, l'annexe, les évaluations du programme seront présentées les unes à la suite des autres, pays par pays.

Implantation de *KiVa* en Finlande

Depuis que les enquêtes PISA de l'OCDE existent, la Finlande est devenue célèbre pour les excellentes performances scolaires de ses élèves. Le système éducatif finlandais est reconnu pour être performant et égalitaire, avec des enseignant·e·s très bien formé·e·s et une grande autonomie laissée aux écoles (Salmivalli et al., 2013). Toutefois, un rapport de l'OMS de 2014 a montré que la Finlande était un des pays où les élèves rapportaient le moins aimer l'école (Sandal, Dür, & Freeman, 2004, cités par Salmivalli et al., 2013). Suite à ces résultats, le ministère finlandais de l'Éducation a souhaité la création et mise en place d'un programme national contre le harcèlement. Il existait déjà, depuis 1999, une législation indiquant que chaque élève a le droit de se sentir en sécurité à l'école et donc que le pouvoir organisateur de l'école a la responsabilité de faire en sorte que les élèves ne subissent pas de harcèlement. Plus précisément, depuis 2003, une loi prévoit que le pouvoir organisateur d'une école doit prévoir des plans d'action pour lutter contre le harcèlement, mettre en œuvre ces plans d'action et superviser leur implantation effective. Malgré ces législations, les résultats d'études réalisées chaque année par le National Institut for Health and Welfare sur la prévalence du harcèlement dans les écoles secondaires ont montré que le phénomène n'avait aucunement diminué (Salmivalli et al., 2011). C'est pourquoi, en 2006, le ministère de l'éducation finlandais a passé un contrat (de 3 ans au départ, renouvelé par la suite) avec l'University of Turku afin qu'un programme anti-harcèlement soit développé et évalué (Salmivalli et al., 2013). Depuis le départ, le programme *KiVa* a donc pour objectif de pouvoir être implanté dans toutes les écoles finlandaises. Le programme a été implanté et évalué pour la première fois lors de l'année scolaire 2007-2008 chez les élèves de 4ème, 5ème et 6ème année puis lors de l'année 2008-2009 chez les élèves de la 1ère à la 3ème année et de la 7ème à la 9ème année. Lors de l'année scolaire 2009-2010, le programme a été implanté à très grande échelle, dans 1450 écoles. Les effets de *KiVa* ont continué à être évalués lors de cette deuxième phase d'implantation, quoique de façon moins rigoureuse. D'autres écoles ont encore rejoint le programme par la suite, et, à partir de l'année 2011-2012, environ 2500 écoles implantaient *KiVa*, ce qui représente 90 % des écoles finlandaises (Salmivalli et al., 2013). Le programme est actuellement toujours en place dans ces écoles.

Qualité de l'évaluation de *KiVa* lors de la phase d'évaluation RCT de 2007-2008 (années 4-6)

Article principal : Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011b). A Large-Scale Evaluation of the KiVa Antibullying Program: Grades 4-6. *Child Development*, 82(1), 311-330. doi:10.1111/j.1467-8624.2010.01557.x

1. Quel était le plan de recherche de l'évaluation

Il s'agit d'une étude longitudinale avec un design RCT (*Randomized Control Trial*, les écoles sont aléatoirement réparties entre une condition contrôle et une condition expérimentale). Les écoles faisant partie de la condition contrôle ont été placées sur liste d'attente pour participer au programme l'année suivante si elles le désiraient toujours. En attendant, elles devaient continuer à gérer le harcèlement comme à leur habitude. Le programme a été implanté à partir d'août 2007. Les données pré-test ont été récoltées avant l'implantation du programme (en mai 2007) et les données post-test ont été récoltées à deux reprises : 4 et 9 mois après le début de l'implantation (en décembre 2007-janvier 2008 et en mai 2008).

Les données ont été analysées via des modèles linéaires multi-niveaux afin de tenir compte de la présence de données nichées. Les modèles possèdent quatre niveaux (le temps, les élèves, les classes et les écoles). Avant de tester les effets de l'intervention, les chercheur·euse·s ont simplifié les modèles afin d'obtenir une certaine parcimonie et une bonne convergence. Pour tenir compte de la présence de données manquantes, une procédure d'imputations multiples a été utilisée.

2. Quelles étaient les caractéristiques de l'échantillon ?

Pour recruter les écoles participantes, des lettres décrivant le programme *KiVa* ont été envoyées en 2006 à toutes les 3418 écoles finlandaises organisant de l'enseignement primaire et/ou secondaire inférieur. 275 d'entre elles se sont portées volontaires pour participer à l'étude. Après avoir exclu les écoles organisant uniquement de l'enseignement spécialisé, les écoles volontaires ont été classées par région et langue (finlandais ou suédois) et 78 d'entre elles ont été aléatoirement assignées à la condition expérimentale ou à la condition contrôle. Les écoles participantes se situent à travers tout le pays et sont représentatives des écoles finlandaises en termes de taille des classes et de proportion d'élèves immigrants. Par contre, les écoles de langue suédoise sont légèrement surreprésentées. Dans ces 78 écoles, il y avait 429 classes et 8248 élèves. Une école entière a abandonné l'étude avant la première récolte des données. En mai 2007, les élèves avaient entre 9 et 11 ans et étaient en 3^{ème}, 4^{ème} ou 5^{ème} année (avec environ 33 % des élèves par niveau). Il y avait environ autant de filles que de garçons et 97,5 % d'élèves d'origine finlandaise. La façon de gérer les données manquantes et l'échantillon final sur lequel les analyses ont été réalisées varie d'un article à l'autre. Dans les articles de Kärnä et al. (2011b), Salmivalli et al. (2011), Kärnä et al. (2010), Saarento, Boulton, & Salmivalli (2015), Juvonen, Schacter, Sainio, & Salmivalli (2016) et Williford et al. (2012), le nombre d'élèves inclus-es dans les analyses varie entre 5651 et 8166 élèves (il s'agit des élèves qui ont eu le consentement parental, n'ont pas quitté l'école après la vague 1 et ont répondu aux questionnaires portant sur les variables d'intérêt).

Les analyses réalisées par Garandeau, Lee, & Salmivalli (2018), portaient sur un échantillon final de 4466 élèves répartis dans 211 classes (115 dans le groupe contrôle et 96 dans les écoles contrôles) venant de 61

Annexe 1 Evaluation de KiVa

écoles. Au temps 1, 623 élèves (15 %) rapportaient être victimes de harcèlement. Au temps 3, ils étaient 420 (10,5 %). Parmi ces élèves, 170 (4,3 %) sont restés victimes du temps 1 au temps 3. Leur âge moyen était de 10,6 ans et 53 % étaient des garçons. Dans l'article de Garandeau, Lee, & Salmivalli (2014), les analyses portent sur 7957 élèves entre 9 et 14 ans, parmi lesquels 911 élèves harceleurs ont été identifiés au temps 1 (89 % étaient des garçons). Les articles de Veenstra, Lindenberg, Huitsing, Sainio, & Salmivalli (2014) et de van der Ploeg, Kretschmer, Salmivalli, & Veenstra (2017) portent uniquement sur les élèves de la condition contrôle.

3. Par qui et où l'évaluation a-t-elle été menée ?

Les auteur·e·s impliqué·e·s dans ces études sont Elisa Poskiparta et Christina Salmivalli (codirectrices du projet KiVa, University of Turku), Antti Kärnä, Anne Kaljonen, Claire Garandeau, Silja Saarento et Miia Sainio (University of Turku), Aaron Boulton, Todd Little, Brian Noland et Anne Williford (University of Kansas), Gijs Huitsing, Tina Kretschmer, Siegwart Lindeberg, René Veenstra et Rozmarijn van der Ploeg (University of Groningen), Jaana Juvonen et Hannah Schacter (University of Southern California) et Marinus Voeten (Radboud University). Ce sont donc essentiellement des personnes impliquées dans le développement de KiVa qui ont évalué son efficacité.

4. Quelles et combien de mesures ont été prises ?

La récolte des données s'est faite en mai 2007 (temps 1), décembre 2007-janvier 2008 (temps 2) et mai 2008 (temps 3). Au temps 2, les élèves n'ont dû remplir qu'une partie des questionnaires. Ils et elles ont répondu aux questionnaires en ligne pendant les heures de cours en étant encadré·e·s par leurs enseignant·e·s qui avaient reçu des instructions détaillées sur la passation des questionnaires et avaient eu l'occasion de poser des questions aux chercheur·euse·s si nécessaire. Les élèves ont reçu des mots de passe afin de pouvoir s'enregistrer dans les questionnaires. Cela permet de relier les données aux différents points de mesure tout en assurant leur anonymat.

Le questionnaire commençait par des questions démographiques (genre, âge, statut d'immigrant, classe et école) suivies de la définition du harcèlement du *Olweus Bully/Victim Questionnaire* (« *It is bullying when one or more children deliberately and repeatedly make another child feel bad. The bully usually has power over the victim and the victim of bullying is usually unable to defend himself or herself against the bully. A child is being bullied when one or more children say mean or unkind things about him or her, make fun of him or her, or call him or her mean and unkind names, completely ignore him or her, leave him or her out of their group of friends, or leave him or her outside on purpose, hit, kick, push or order him or her around or, for example, lock him or her in a room try to make other children dislike him or her by spreading lies about him or her, or by sending mean notes or doing other unkind things than the ones mentioned above. Also, it is bullying when a child is teased repeatedly in a mean and unkind way. Friendly and playful teasing is not bullying. It is also not bullying when children willingly argue or fight.* »). De plus, chaque question portant sur le harcèlement était accompagnée d'une version réduite de la définition. Mis à part pour le *Olweus Bully/Victim Questionnaire*, l'ordre des questions, items et échelles était aléatoire afin d'éviter des effets d'ordre.

Annexe 1 Evaluation de KiVa

Les mesures qui ont été prises sont les suivantes :

- * **Harcèlement auto-rapporté** : Le harcèlement auto-rapporté a été mesuré via l’item global de la version révisée du *Olweus Bully/Victim Questionnaire* (Olweus, 1996; cité par Kärnä et al., 2011b) : « *How often have you bullied others at school in the last couple of months ?* ». Les élèves ont répondu via une échelle de Likert à 5 points (0=*Jamais*, 1=*Seulement une ou deux fois*, 2=*Deux ou trois fois par mois*, 3=*Environ une fois par semaine*, 4=*Plusieurs fois par semaine*). Pour calculer les OR (odds ratio), les élèves ont été catégorisé·e·s comme harceleur·euse s’ils ou elles avaient répondu 2 ou plus. Les élèves ont également répondu à l’item concernant le cyberharcèlement du *OBVQ* (« *I bullied others by cell phone or through the Internet: I sent mean or hurtful messages, calls, or pictures.* »).
- * **Victimisation auto-rapportée** : La victimisation auto-rapportée a été mesurée via l’item global du *OBVQ* : « *How often have you been bullied at school in the last couple of months ?* ». Les élèves y ont répondu via la même échelle à 5 points que pour le harcèlement et la dichotomisation en victime/non victime s’est faite de la même manière. L’item global était suivi par 9 items décrivant des formes spécifiques de harcèlement, auxquels les élèves devaient toujours répondre sur la même échelle à 5 points. Ces items concernaient le harcèlement verbal (« *I was called mean names, was made fun of or teased in a hurtful way.* »), l’exclusion (« *Other students ignored me completely or excluded me from things or from their group of friends.* »), le harcèlement physique (« *I was hit, kicked, or shoved.* »), la manipulation (« *Other students tried to make others dislike me by spreading lies about me.* »), le harcèlement matériel (« *Somebody took money or other things from me or damaged my things.* »), les menaces (« *I was threatened or forced to do things I would not have wanted to do.* »), le racisme (« *I was bullied by calling me names, making remarks or gestures about my ethnicity or skin color. Such a name can be, for example, nigger.* »), le harcèlement sexuel (« *I was bullied by sexual name calling (such a name can be, for instance, slut or gay), sexual gestures or actions.* ») et le cyberharcèlement (« *I was bullied by cell phone or through the Internet: I received mean or hurtful messages, calls, or pictures. (Please also report any other ways of bullying by means of a cell phone or a computer.)* »).
- * **Rôles dans les situations de harcèlement rapportés par les pairs** : Les rôles dans les situations de harcèlement ont été évalués via le *Participant Role Questionnaire* (Salmivalli et al., 1996; cité par Kärnä et al., 2011b). Il était demandé aux élèves de penser à une situation de harcèlement et de nommer les élèves qui se comporteraient de la façon décrite par les différents items de l’échelle. Une liste des élèves de la classe était présente sur l’écran et les élèves pouvaient nommer autant d’élèves qu’ils ou elles voulaient, y compris aucun·e. Les 12 items décrivaient des comportements de harcèlement (« *Starts bullying* », « *Makes the others join in the bullying* », « *Always finds new ways of harassing the victim* »), de victimisation (« *He/She is being pushed around and hit* », « *He /She is called names and mocked* » et « *Nasty rumors are spread about him /her* »), d’assistance (« *Joins in the bullying, when someone else has started it* », « *Assists the bully* », « *Helps the bully, maybe by catching the victim* »), de renforcement (« *Comes around to watch the situation* », « *Laughs* », « *Incites the bully by shouting or saying: Show him/her!* ») et de défense (« *Comforts the victim or encourages him/her to tell the teacher about the bullying* », « *Tells the others to stop bullying* »,

Annexe 1 Evaluation de KiVa

«*Tries to make the others stop bullying*»). Le nombre de nominations par les pairs était ensuite, pour chaque élève, divisé par le nombre total d'élèves de leur classe ayant répondu. Cela donne un score compris entre 0 et 1 pour chaque item. Les scores des items ont ensuite été moyennés de façon à obtenir un score global pour chaque rôle. Les alphas (mesure de la corrélation entre les différents items du test) se situent entre .84 et .91.

- * **Acceptation par les pairs** : Pour mesurer l'acceptation par les pairs, chaque élève devait choisir parmi une liste les trois élèves de sa classe qu'il ou elle aimait le plus. Le nombre de nominations reçues est divisé par le nombre d'élèves de la classe ayant répondu. Il en résulte donc un score entre 0 et 1 pour chaque élève.
- * **Anxiété** : L'anxiété a été mesurée à l'aide de deux échelles combinées : la *Fear of Negative Evaluation* (Garcia-López et al., 2001; cité par Williford et al., 2012) composée de 5 items évaluant à quel point l'évaluation par les pairs cause du stress et de l'inquiétude et la *Social Avoidance and Distress* (Garcia-López et al., 2001; cité par Williford et al., 2012) composée de 4 items évaluant à quel point l'élève évite les interactions sociales et se sent mal à l'aise dans les situations de groupe. Les 9 items ont été moyennés pour former un score global (l'alpha est égal à .88).
- * **Anxiété sociale** : L'anxiété sociale est mesurée via une version à 9 items de la *Social Anxiety Scale for Adolescent* (La Greca & Lopez, 1998; cité par Kärnä et al., 2010). Les élèves ont répondu sur une échelle à 5 points allant de *Pas du tout* à *Tout le temps* à des items tels que «*I'm afraid to invite other to do things with me because they might say no.*». Les 9 items ont été moyennés en un score global (l'alpha est égal à .88).
- * **Attitudes anti-harcèlement** : Les attitudes anti-harcèlement ont été évaluées via une version réduite à 10 items de la *Provicim scale* (Rigby & Slee, 1991; cité par Kärnä et al., 2011b). Les élèves ont répondu à des items tels que «*It's okay to call some kids nasty names*» sur une échelle à 5 points allant de *Désaccord complet* à *Accord complet*. Une fois les items négatifs inversés, les dix items ont été moyennés en un score total (l'alpha est égal à .79).
- * **Bien-être à l'école** : Le bien-être à l'école a été évalué avec des items initialement développés par le *Finnish National Board of Education* (Metsämuuronen & Svedlin, 2004 ; cité par Kärnä et al., 2011b) concernant l'appréciation générale de l'école (ex. «*My school days are generally nice*»), le concept scolaire de soi-même (ex. «*Learning brings me joy*»), le climat de la classe (ex. «*There is a good climate in our class*») et le climat de l'école (ex. «*I feel safe at school*»). Les élèves y ont répondu sur une échelle de 5 points (allant de *Désaccord complet* à *Accord complet*). Les 14 items ont été moyennés pour former un score global (l'alpha est égal à .88). Les 5 items mesurant le climat de l'école ont également été moyennés pour former un score de perception du climat scolaire, tout comme les 3 items mesurant l'appréciation générale de l'école pour former un score d'attitude envers l'école.
- * **Dépression** : Le niveau de dépression des élèves a été mesuré avec une échelle à 7 items dérivée du *Beck Depression Inventory* (Beck et al., 1996; cité par Williford et al., 2012). Les items qui ne semblaient pas adéquats pour des enfants (ceux en rapport avec les idées suicidaires, l'intérêt sexuel ou les plaintes somatiques) ont été retirés. Les items ont été moyennés pour former un score global (l'alpha est égal à .89).

Annexe 1 Evaluation de KiVa

- * **Efficacité personnelle à défendre la victime** : L'efficacité personnelle à défendre la victime a été évaluée avec des items dérivés de la partie défense du *Participant Role Questionnaire*. Les élèves ont dû évaluer sur une échelle à 4 points allant de *Très difficile pour moi* à *Très facile pour moi* si adopter les comportements présentés leur serait difficile. Un score global a été calculé (l'alpha est égal à .69).
- * **Empathie envers la victime** : L'empathie envers la victime a été mesurée via une échelle d'empathie de 7 items (tels que « *When a bullied child is sad I feel sad as well.* ») développée par Pöyhönen et al. (2010; cité par Kärnä et al., 2011b). Les élèves ont répondu sur une échelle à 5 points allant de *Jamais* à *Toujours*. Les scores de chaque item ont été moyennés pour faire un score total (l'alpha est égal à .84). Un score d'empathie cognitive (4 items) et un score d'empathie affective (3 items) ont également été calculés.
- * **Estime de soi** : L'estime de soi a été mesurée via une version traduite et modifiée (de façon à mesurer l'estime de soi parmi les pairs) de la *Rosenberg Self-Esteem Scale* (Rosenberg, 1965; cité par Juvonen et al., 2016). Les élèves ont répondu aux dix items (ex. « *I feel I am not good enough for anything when among my peers* ») sur une échelle à cinq points allant de *Pas du tout vrai* à *Tout à fait vrai*.
- * **Perceptions de l'enseignant-e en lien avec le harcèlement** : La perception des attitudes, de l'efficacité et des efforts de l'enseignant-e a été mesurée via trois items. Les élèves devaient répondre sur une échelle de 5 points à la question de savoir ce que leur enseignant-e pensait du harcèlement (Saarento et al., 2013; cité par Saarento et al., 2015), à quel point il ou elle était capable de diminuer le harcèlement et à quel point il ou elle faisait des efforts afin de diminuer le harcèlement. Les possibilités de réponse allaient de *L'enseignant-e pense que le harcèlement est une bonne chose* à *L'enseignant-e pense que le harcèlement est tout à fait mal* ou de *Il n'y a rien que l'enseignant-e peut faire/fait* à *l'enseignant-e peut faire/fait beaucoup*.
- * **Perception des pairs** : Les croyances des élèves à propos de leurs pairs ont été mesurées via le *Generalized Perception-of-Peers Questionnaire* (Salmivalli et al., 2005; cité par Williford et al., 2012) une échelle comportant 13 items (ex. « *They can be relied on* ») auxquels les élèves ont dû répondre avec une échelle à 5 points allant de *Pas du tout vrai* à *Tout à fait vrai*. Les items ont été moyennés pour former un score total (l'alpha est égal à .89).
- * **Popularité perçue** : Afin de mesurer la popularité perçue des élèves, il leur a été demandé de répondre à la question « *Who are the most popular in your class ?* » en sélectionnant autant d'élèves qu'ils ou elles le désiraient parmi la liste des élèves de la classe. Le nombre de nominations par les pairs a été, pour chaque élève, divisé par le nombre d'élèves de la classe ayant répondu. Cela donne un score de popularité perçue entre 0 et 1 pour chaque élève. Les valeurs obtenues ont ensuite été standardisées pour pouvoir classer les élèves en fonction de leur popularité perçue (popularité haute pour ceux ayant un score supérieur à 0,5 écart-type, moyenne pour ceux ayant un score situé entre -0,5 et 0,5 écart-type et faible pour les élèves ayant un score inférieur à -0,5 écart-type).
- * **Rejet par les pairs** : Pour mesurer le rejet par les pairs, chaque élève devait choisir les trois élèves de sa classe qu'il ou elle aimait le moins. Le nombre de nominations reçues a été divisé par le

Annexe 1 Evaluation de *KiVa*

nombre d'élèves de la classe ayant répondu. Il en résulte donc un score entre 0 et 1 pour chaque élève.

Effets de *KiVa* lors de la phase d'évaluation RCT de 2007-2008 (années 4-6)

1. Des effets ont-ils été observés ?

Données soutenant la théorie sous-jacente au programme :

L'étude de Kärnä et al. (2010) investigate le lien entre les facteurs de risque de victimisation, les comportements des témoins et la victimisation. Elle porte uniquement sur les données prétest de l'étude. Une fois l'âge et le genre contrôlés, le rejet par les pairs et l'anxiété sociale sont des facteurs prédictifs de la victimisation : les élèves qui sont davantage rejeté·e·s par leurs pairs et qui présentent des niveaux élevés d'anxiété sociale ont une probabilité plus importante que les autres d'être victimes de harcèlement. Mais cette relation est influencée par le contexte dans lequel se trouve l'élève. Le coefficient interclasse pour la victimisation est de .13, ce qui signifie que 13 % de la variance de la victimisation est due aux caractéristiques des élèves de la classe. Le rejet et l'anxiété sociale prédisent davantage la victimisation dans les classes où de nombreux élèves adoptent des comportements de renforcement face aux situations de harcèlement. Et l'anxiété sociale prédit davantage la victimisation dans les classes où peu d'élèves défendent les victimes de harcèlement. L'attitude des témoins dans les situations de harcèlement a donc bien un impact sur la victimisation. Les enseignant·e·s ont également un impact. Veenstra et al. (2014) ont montré que la relation entre les caractéristiques des élèves (attitudes anti-harcèlement, comportements de défense et efficacité personnelle à défendre la victime) et leur tendance à harceler dépendait de leur perception des capacités et efforts de leur enseignant·e (cette relation est moins forte dans les classes où l'enseignant·e est perçu comme étant efficace et ne devant pas faire beaucoup d'efforts pour réduire le harcèlement).

van der Ploeg et al. (2017) ont investigué la relation entre l'empathie cognitive et affective, l'efficacité personnelle et la tendance à défendre les victimes de harcèlement. Selon leurs résultats, les élèves ayant de hauts scores d'empathie affective et/ou d'efficacité personnelle sont plus susceptibles de défendre les élèves victimes de harcèlement. De plus, la tendance à défendre les élèves victimes de harcèlement est associée à une augmentation de la popularité perçue parmi les pairs. Les résultats sont les mêmes que les élèves soient eux ou elles-mêmes victimes de harcèlement ou non. Ces résultats soutiennent l'idée derrière le programme *KiVa* selon laquelle augmenter l'empathie (affective) et l'efficacité personnelle des élèves permettrait d'augmenter les comportements de défense vis-à-vis des pairs victimes de harcèlement.

Saarento et al. (2015) se sont intéressé·e·s aux médiateurs de l'effet du programme *KiVa* sur la réduction du harcèlement et de la victimisation. Leurs résultats suggèrent que *KiVa* parvient à contrecarrer le harcèlement et la victimisation via une augmentation des comportements des témoins dans les situations de harcèlement, des attitudes anti-harcèlement des élèves et de leur perception des attitudes anti-harcèlement de l'enseignant·e.

Annexe 1 Evaluation de KiVa

Effets du programme :

Après 4 mois d'implantation du programme, des effets sont observés sur 4 des 11 variables dépendantes (voir tableau 1). Dans les écoles ayant eu l'intervention, par rapport aux écoles contrôles, il y a plus de défense des élèves victimes et moins de victimisation rapportée par les pairs et les élèves présentent davantage d'attitudes anti-harcèlement ainsi qu'une plus grande empathie vis-à-vis des victimes. Après 7 mois d'implantation du programme, des effets sont observés sur 7 des 11 variables dépendantes. Dans les écoles ayant eu l'intervention par rapport aux écoles contrôles, il y a moins de victimisation (auto-rapportée et rapportée par les pairs), moins de harcèlement auto-rapporté, moins d'assistance et de renforcement, une efficacité personnelle pour défendre la victime renforcée et un bien-être à l'école plus important. Par contre, il n'y a plus d'effet sur le harcèlement hétéro-rapporté, la défense de la victime, les attitudes anti-harcèlement et l'empathie vis-à-vis des victimes. L'évolution de la victimisation et du harcèlement auto-rapportés sont repris dans le tableau 2.

Williford et al. (2012) s'intéressent à d'autres variables dépendantes : la perception des pairs, la dépression et l'anxiété. La perception positive des pairs diminue davantage dans le groupe contrôle que dans le groupe expérimental, tandis que l'anxiété diminue davantage dans le groupe expérimental (il n'y a par contre pas d'effet du programme sur la dépression). Des analyses supplémentaires sur les relations entre ces variables suggèrent que, dans les deux groupes, une réduction de la victimisation cause une augmentation de la perception positive des pairs et une diminution du niveau d'anxiété

Tableau 1. Taille des effets (d de Cohen) après 4 et 7 mois d'implantation de KiVa

	Décembre 2007	Mai 2008
Victimisation auto-rapportée	0.03 (n.s.)	0,17
Harcèlement auto-rapporté	0.05 (n.s.)	0,10
Victimisation rapportée par les pairs	0,18	0,33
Harcèlement rapporté par les pairs	0.10 (n.s.)	0.14 (n.s.)
Assistance	0.12 (n.s.)	0,14
Renforcement	0.12 (n.s.)	0,17
Défense	0,11	0.08 (n.s.)
Attitudes anti-harcèlement	0,09	0.06 (n.s.)
Empathie envers la victime	0,10	0.06 (n.s.)
Efficacité personnelle	0.01 (n.s.)	0,08
Bien-être à l'école	0.05 (n.s.)	0,10
Perception des pairs	-	0,08
Dépression	-	-0,04
Anxiété	-	-0,11

Inspiré de Kärnä et al. (2011b, p. 323) et Williford et al. (2012b, p.302). Les effets significatifs sont indiqués en gras

Annexe 1 Evaluation de *KiVa*

Tableau 2. Évolution de la victimisation et du harcèlement auto-rapporté

	Mai 2007	Décembre 2007	Mai 2008	Différence en faveur des écoles <i>KiVa</i>	OR (IC à 95 %)
Victimisation auto-rapportée					
Intervention	16,6 %	16,4 %	8,9 %	-30 %	1,47 (1,10-1,96)
Contrôle	16,8 %	19,1 %	12,7 %		Rapporté par les pairs : 1,83
Harcèlement auto-rapporté					
Contrôle	8 %	4,6 %	3,1 %	-17 %	1,22 (0,78-1,90)
Intervention	7,9 %	6,9 %	3,8 %		Rapporté par les pairs : 1,29

Inspiré de Kärnä et al. (2011b, p. 323)

Dans leur étude, Salmivalli et al. (2011) ont investigué les effets de *KiVa* sur différentes formes de victimisation. Il semble que le programme a eu un effet positif sur toutes les formes de victimisation. Cependant cet effet n'est pas significatif pour l'exclusion, la dégradation ou le vol du matériel et les menaces. Il faut toutefois savoir que, leur fréquence plutôt basse, l'intervalle de confiance pour ces formes de victimisation est assez large. L'absence d'effet significatif pourrait être due à un manque de précision.

Tableau 3. Effets de *KiVa* sur la victimisation auto-rapportée

	Mai 2007	Mai 2008	Changemen t	Différence en faveur des écoles <i>KiVa</i>	OR (IC à 80 %)
Victimisation verbale					
Intervention	9,4 %	7,3 %	-22 %	-31 %	1.45 (1,14-1,83)
Contrôle	9,8 %	10,7 %	+9 %		
Victimisation par exclusion					
Intervention	5,3 %	4,4 %	-17 %	-25 %	1.32 (0,98-1,77)
Contrôle	6,4 %	6,9 %	+8 %		
Victimisation physique					
Intervention	4,3 %	2,6 %	-40 %	-53 %	1.94 (1,37-2,74)
Contrôle	4,5 %	5,1 %	+13 %		
Victimisation par manipulation					
Intervention	4,8 %	3,6 %	-25 %	-30 %	1.43 (1,05-1,97)
Contrôle	5,6 %	5,9 %	+5 %		
Victimisation matérielle					
Intervention	1.3 %	1,0 %	-30 %	-63 %	1.76 (0,96-3,22)
Contrôle	1,2 %	1,6 %	+33 %		

Annexe 1 Evaluation de KiVa

Victimisation par menace					
Intervention	2,0 %	1,0 %	-50 %	-21 %	1.52 (0,89-2,61)
Contrôle	2,4 %	1,7 %	-29 %		
Victimisation raciste					
Intervention	3,8 %	2,1 %	-45 %	-27 %	1.54 (1,02-2,32)
Contrôle	3,4 %	2,8 %	-18 %		
Victimisation sexuelle					
Intervention	6,0 %	5,0 %	-17 %	-32 %	1.42 (1,07-1,88)
Contrôle	6,5 %	7,5 %	+15 %		
Cyber Victimisation					
Intervention	2,2 %	1,4 %	-36 %	-50 %	1.80 (1,11-2,93)
Contrôle	2,2 %	2,5 %	+14 %		
Victimisation globale					
Intervention	15,1 %	9,0 %	-40 %	-29 %	1.56 (1,28-1,91)
Contrôle	16,1 %	14,3 %	-11 %		

Inspiré de Salmivalli et al. (2011, p. 409). Les OR significatifs sont en gras.

D'après les résultats obtenus par Salmivalli, Garandeau et Veenstra (2012, cité par Salmivalli et al., 2013), le programme *KiVa* permet également d'augmenter l'appréciation scolaire, la motivation académique et les performances académiques des élèves.

Garandeau et al. (2018) ont analysé l'effet de la réduction globale de la victimisation suite à l'implantation de *KiVa* sur l'ajustement social et psychologique des élèves qui sont resté·e·s victimes de harcèlement malgré l'intervention. Les élèves qui étaient déjà victimes de harcèlement avant l'implantation du programme et qui le sont toujours à la fin de l'année présentent davantage de symptômes de dépression et d'anxiété sociale lorsqu'ils ou elles se trouvent dans des classes où le niveau global de victimisation a diminué plutôt que dans des classes où il est resté stable ou a augmenté. De plus, ces élèves sont également moins aimé·e·s par leurs pairs et se sentent moins défendu·e·s. Le fait que les classes se trouvent dans une école *KiVa* ou une école contrôle n'a pas d'influence. Cela veut donc dire que, si *KiVa* permet de diminuer la prévalence de la victimisation, l'adaptation psychologique et sociale des élèves qui restent victimes de harcèlement malgré l'intervention peut empirer.

2. Les effets sont-ils différenciés ?

Au niveau du harcèlement rapporté par les pairs, il y a une différenciation des effets en fonction du niveau scolaire des élèves : la taille d'effet est plus importante chez les élèves plus âgé·e·s ($d=0.03$ en 4^{ème} contre $d=0.18$ en 6^{ème}). Le niveau scolaire a également un impact sur l'effet de *KiVa* concernant les scores de dépression et d'estime de soi. Chez les élèves de 4^{ème} et 5^{ème} année, il n'y a pas d'effet de l'intervention sur le niveau de dépression. Par contre chez les élèves de 6^{ème} année qui sont fréquemment victimes de harcèlement, le niveau de dépression est plus faible et l'estime de soi plus élevée dans les écoles *KiVa* que dans les écoles contrôles (Juvonen et al., 2016). Les mêmes auteurs ont montré que la fréquence de la victimisation subie avait un impact sur les effets de *KiVa* concernant la perception du climat scolaire : l'amélioration de la perception du climat de classe suite à l'implantation du programme est plus importante chez les enfants fréquemment victimes de harcèlement au prétest par rapport à ceux et celles l'étant moins.

Annexe 1 Evaluation de *KiVa*

fréquemment. Par contre, ce n'est pas le cas des attitudes envers l'école qui sont plus favorables dans la condition expérimentale, peu importe le niveau de victimisation.

Garandeau, Lee et Salmivalli (2014) ont également montré que *KiVa* a un effet sur le harcèlement rapporté par les pairs, mais uniquement chez les élèves qui, au prétest, étaient perçus comme peu ou moyennement populaire. Le programme ne semble donc pas être efficace pour réduire les comportements de harcèlement d'élèves perçus comme fortement populaires.

3. Des mesures de suivi ont-elles été prises ?

Les mesures ont uniquement été prises après 4 et 7 mois d'implantation. Le programme *KiVa* prévoit que les écoles prennent des mesures des taux de harcèlement et victimisation chaque année, mais, pour l'instant, ces données n'ont été publiées nulle part.

4. L'implantation a-t-elle été mesurée ?

Aucune mesure de l'implantation n'est rapportée.

Qualité de l'évaluation de *KiVa* lors de la phase d'évaluation RCT de 2008-2009 (années 1-3 et 7-9)

Article principal : Kärnä, A., Voeten, M., Little, T. D., Alanen, E., Poskiparta, E., & Salmivalli, C. (2013). Effectiveness of the kiva antibullying program: Grades 1-3 and 7-9. *Journal of Educational Psychology*, 105(2), 535-551. doi:10.1037/a0030417

1. Quel était le plan de recherche de l'évaluation ?

Le plan de recherche de l'étude de Kärnä et al. (2013) est le même que celui de l'étude de Kärnä et al. (2011b). Il s'agit d'une étude longitudinale avec un design RCT. De plus, pour les études de Garandeau, Poskiparta & Salmivalli (2014) et de Garandeau, Vartio, Poskiparta, & Salmivalli (2016) les écoles dans la condition expérimentale ont été aléatoirement réparties en deux groupes. Selon le groupe dans lequel se trouvait l'école, elle devait gérer les actions indiquées en utilisant une approche avec ou sans confrontation. Les mesures pré-test ont été récoltées avant l'implantation du programme (en mai 2008) et les mesures post-test ont été récoltées à deux reprises, 4 et 9 mois après le début de l'implantation (en décembre 2008-février 2009 et en mai 2009).

Les analyses statistiques effectuées sont également les mêmes que pour l'étude de Kärnä et al. (2011b), mis à part pour l'étude de Ahtola, Haataja, Kärnä, Poskiparta & Salmivalli (2012) pour laquelle seules des données post-test étaient disponibles (il s'agit donc d'une simple comparaison du groupe contrôle et expérimental).

2. Quelles étaient les caractéristiques de l'échantillon ?

Les écoles participant à l'étude avaient été contactées en 2006, en même temps et de la même manière que celles ayant participé à l'étude de Kärnä et al. (2011b). 125 écoles volontaires ont été aléatoirement

Annexe 1 Evaluation de KiVa

assignées à la condition expérimentale (47 écoles) ou à la condition contrôle (78 écoles). De plus, 31 écoles qui faisaient partie de la condition contrôle l'année précédente (pour les élèves de la 4^{ème} à la 6^{ème}) font ici partie de la condition expérimentale. Une dizaine d'écoles ont quitté l'étude avant la récolte des données. Il y a donc, au total, 74 écoles implantant le programme en 1^{er}, 2^{ème} et 3^{ème} année et 73 en 7^{ème}, 8^{ème} et 9^{ème} année (avec une seule école implantant le programme dans les deux cycles). Comme pour des raisons pratiques il n'était pas possible de prendre de mesure prétest pour les élèves de 1^{ère} et 7^{ème} année, les analyses ont uniquement été réalisées chez les 2^{ème} et 3^{ème} année (4704 élèves issus de 273 classes) et chez les 8^{ème} et 9^{ème} année (11070 élèves issus de 686 classes). Les élèves inclus-es dans les analyses ont reçu un consentement parental et n'ont pas changé d'école au cours de l'étude. Afin de tenir compte des données manquantes, une procédure FIML (*Full Information Maximum Likelihood*) sous le postulat *missing at random* a été effectuée.

L'étude de Ahtola et al. (2012) porte sur les enseignant·e·s de la 1^{ère} à la 3^{ème} année. Les données étaient disponibles pour 238 participant·e·s (128 dans la condition expérimentale et 110 dans la condition contrôle) travaillant dans 62 écoles. L'étude de Ahtola et al. (2013) porte sur les 93 enseignant·e·s de la 1^{ère} à la 3^{ème} année qui ont retourné le livret, donné toutes les leçons et complété le questionnaire en ligne.

Les études de Garandeau, Poskiparta & Salmivalli (2014); Williford et al. (2013); Haataja et al. (2014), Haataja, Ahtola, Poskiparta, & Salmivalli (2015), Garandeau et al. (2016) et Yang & Salmivalli (2015) portent à la fois sur les données issues de la phase d'évaluation RCT de 2007-2008 et sur celles issues de la phase d'évaluation RCT de 2008-2009.

Les analyses réalisées dans l'étude de Garandeau, Poskiparta & Salmivalli (2014) portent sur les données des élèves de la 1^{ère} à la 9^{ème} année pour lesquels une action indiquée a été mise en place. Sur les 117 écoles faisant partie de la condition expérimentale, seules 65 écoles, ne se distinguant pas des autres en termes de niveau de harcèlement, ont renvoyé les données concernant les actions indiquées. Les données faisaient lieu de 609 cas de harcèlement gérés, mais 62 cas impliquant un élève qui n'avait pas obtenu de consentement parental et 208 cas incomplets n'ont pas été inclus dans les analyses. Sur les cas restants, 160 situations ont été gérées avec une approche sans confrontation et 179 avec une approche avec confrontation. Les analyses réalisées par Garandeau et al. (2016) portent également sur l'analyse des actions indiquées, mais uniquement chez les élèves de la 4^{ème} à la 8^{ème} année. Sur les interventions analysées, 133 situations ont été gérées avec confrontation et 208 sans.

Pour l'étude de Williford et al. (2013), les analyses sont réalisées sur les données des élèves de la 4^{ème} à la 9^{ème} année (excepté les élèves de 7^{ème} année). L'échantillon est donc composé de 18412 élèves. Les analyses réalisées par Haataja et al. (2014) portent uniquement sur les élèves et enseignant·e·s de la 1^{ère} à la 6^{ème} année. En enlevant les participant·e·s qui n'avaient pas complété toutes les mesures (harcèlement et victimisation auto-rapportés pour les élèves, livret d'implantation pour les enseignant·e·s), l'échantillon est composé de 332 enseignant·e·s et de 7413 élèves. Les analyses réalisées par Haataja et al. (2015) ont porté sur 282 enseignant·e·s (sur les 439 de l'échantillon total), également de la 1^{ère} à la 6^{ème}, ayant renvoyé le livret et répondu au questionnaire en ligne. Dans l'étude de Yang & Salmivalli (2015), les analyses portent sur 23520 élèves de la 2^{ème} à la 9^{ème} année. L'étude de Haataja, Sainio, Turtonen et Salmivalli (2016) porte

Annexe 1 Evaluation de KiVa

uniquement sur les victimes stables de harcèlement à partir des données de 9075 élèves venant de 76 écoles, dont 348 victimes stables.

3. Par qui et où l'évaluation a-t-elle été menée ?

Les auteur·e·s de ces études sont Christina Salmivalli, Elisa Poskiparta, Antti Kärnä, Erkki Alanen, Annarilla Ahtola, Claire Garandeau et Anne Haataja (University of Turku), Todd Little, Aaron Boulton, Kathryn DePaolis et Christian Elledge (University of Kansas), Marinus Voeten (Radboud University), Ihno Lee (Stanford University) et Anne Williford (University of Texas). L'évaluation est donc menée par la même équipe que l'évaluation précédente.

4. Quelles et combien de mesures ont été prises ?

La récolte des données s'est faite en mai 2008 (temps 1), décembre 2008-février 2009 (temps 2) et mai 2009 (temps 3). La passation des questionnaires s'est faite exactement de la même façon que dans l'étude de Kärnä et al. (2011b). Pour les enfants de 7-9 ans, la définition du harcèlement était simplifiée, l'enseignant·e lisait toutes les questions à voix haute (elles n'étaient donc pas présentées de façon aléatoire), les échelles étaient accompagnées de couleur afin de faciliter les réponses et aucune mesure supplémentaire au OBVQ n'a été prise.

Les mesures qui ont été prises sont les suivantes :

- * **Harcèlement auto-rapporté** : Le harcèlement auto-rapporté a été mesuré via l'item global de la version révisée du *Olweus Bully/Victim Questionnaire* (Olweus, 1996; cité par Kärnä et al., 2013) : « *How often have you bullied others in the last couple of months ?* ». Les élèves y répondent via une échelle de Likert à 5 points (0=*Jamais*, 1=*Seulement une ou deux fois*, 2=*Deux ou trois fois par mois*, 3=*Environ une fois par semaine*, 4=*Plusieurs fois par semaine*). Pour calculer les OR, les élèves ont été catégorisé·e·s comme harceleur·euse s'ils ou elles avaient répondu 2, 3 ou 4. En plus de l'item global, quatre items mesuraient des formes spécifiques de harcèlement (verbal, physique, exclusion sociale et manipulation). Ces items ont été moyennés pour former un score global de harcèlement.
- * **Victimisation auto-rapportée** : La victimisation auto-rapportée a été mesurée via l'item global du OBVQ : « *How often have you been bullied in the last couple of months ?* ». Les élèves y ont répondu via une échelle à 5 points comme pour le harcèlement et la dichotomisation en victime/non victime s'est faite de la même manière que pour le harcèlement. En plus de l'item global, quatre items mesuraient des formes spécifiques de victimisation (verbale, physique, exclusion sociale et manipulation). Ces items ont été moyennés pour former un score global de victimisation. Au temps 3, il était également demandé aux élèves s'ils ou elles avaient parlé à un·e adulte de leur victimisation éventuelle.
- * **Rôles dans les situations de harcèlement rapportés par les pairs** : Les rôles dans les situations de harcèlement ont été mesurés via le *Participant Role Questionnaire* (Salmivalli et al., 1996; cité par Kärnä et al., 2013), exactement de la même façon que dans l'étude de Kärnä et al. (2011b). Les alphas se situent entre .73 et .90.

Annexe 1 Evaluation de KiVa

- * **Cyberharcèlement** : Le cyberharcèlement a été mesuré via l'item du *OBVQ* portant sur ce type de harcèlement (« *I bullied others by cell phone or through the Internet: I sent mean or hurtful messages, calls, or pictures.* »). Les élèves y ont répondu via une échelle de Likert à 5 points allant de *Jamais* à *Plusieurs fois par semaine*.
- * **Cybervictimisation** : La cybervictimisation a été mesurée via l'item du *OBVQ* portant sur ce type de harcèlement (« *I was bullied by cell phone or through the Internet: I received mean or hurtful messages, calls, or pictures.* »). Les élèves y ont répondu via une échelle de Likert à 5 points allant de *Jamais* à *Plusieurs fois par semaine*.
- * **Forme de victimisation (lors des actions indiquées)** : En se basant sur ce que l'élève pris·e en charge lui a raconté, le membre de l'équipe *KiVa* qui complète le rapport a dû noter à quelle forme de victimisation l'élève avait été confronté·e en choisissant parmi les formes suivantes : insultes, coups, exclusion sociale, rumeurs et mensonges répandus, harcèlement à contenu sexuel, menaces, vol/dommages, remarques racistes, cyberharcèlement via téléphone, cyberharcèlement via internet et autre. Pour les analyses seules les quatre premières formes de harcèlement, qui sont les plus fréquentes, ont été prises en compte.
- * **Durée de la victimisation (lors des actions indiquées)** : Le membre de l'équipe *KiVa* indiquait également depuis combien de temps le harcèlement durait parmi les possibilités suivantes : *Une semaine ou deux*, *Un mois*, *2 à 6 mois*, *6 à 12 mois* et *Plus d'un an*.
- * **Issue de la situation (lors des actions indiquées)** : Lors de la discussion de suivi, le membre de l'équipe *KiVa* demandait à l'élève victime de choisir une des options suivantes concernant l'issue de la situation : *Le harcèlement a totalement cessé*, *Le harcèlement a diminué*, *Le harcèlement est resté le même* ou *Le harcèlement a empiré*. Pour les analyses, l'issue de la situation a été transformée en variable binaire (1=*Le harcèlement a stoppé*, 0=*Le harcèlement continu*).
- * **Perception de l'intervention (lors des actions indiquées)** : Pour mesurer la responsivité aux deux approches, les élèves auteur·e-s de harcèlement ont dû, suite à l'intervention, répondre à des questions portant sur le sentiment d'avoir été blâmé, sur le sentiment d'avoir été poussé à ressentir de l'empathie et sur leurs intentions de changer de comportement.

Les enseignant·e-s ont également répondu à un questionnaire. Les mesures qui ont été prises sont les suivantes :

- * **Compétence à gérer les cas de harcèlement** : Le sentiment de compétence des enseignant·e-s à gérer les situations de harcèlement a été mesuré via deux items (« *How much, in your own opinion, do you know about school bullying ?* » et « *How good, in your own opinion, are your skills to reduce school bullying ?* ») auxquels les enseignant·e-s ont répondu via une échelle à 5 points allant de *Très peu de connaissances/compétences* à *Beaucoup de connaissances/compétences*. Les items ont été moyennés pour former un score global (l'alpha est égal à .64).
- * **Confiance envers le programme** : La confiance envers le programme a été mesurée via trois items (« *To what extent do you believe that the KiVa program will decrease bullying incidents ?* », « *To what extent do you believe that the KiVa program will enhance the well-being of victimized students ?* » et « *To what extent do you believe that the KiVa program will enhance school satisfaction of pupils* »).

Annexe 1 Evaluation de KiVa

?») auxquels les enseignant·e·s ont répondu via une échelle à 5 points allant de *Très peu* à *Très fortement*. Les items ont été moyennés pour former un score global. L'alpha est égal à .86

- * **Vision du harcèlement comme un phénomène malléable** : La compréhension des enseignant·e·s de la malléabilité du harcèlement a été mesurée via huit items auxquels ils et elles devaient marquer leur degré d'accord via une échelle à 5 points allant de *Pas du tout d'accord* à *Totalement d'accord*. Quatre des huit items insistent sur la possibilité d'intervention de la part de l'école (ex. « *Group dynamics are the reason for bullying, and school staff can have an influence* ») tandis que les quatre autres, dont le score est inversé, considèrent le harcèlement comme un phénomène sur lequel le personnel de l'école n'a pas prise (ex. « *School staff cannot help it, some kids just bully others* »). Les items ont été moyennés pour former un score global (l'alpha est égal à .72).
- * **Soutien de la direction dans la lutte anti-harcèlement** : Les enseignant·e·s ont évalué le support de la direction en répondant à 5 items (ex. « *The principal supports the antibullying work in our school* ») sur une échelle de Likert à 5 points allant de *Pas du tout d'accord* à *Totalement d'accord*. L'alpha est égal à .90.
- * **Soutien général de la direction** : Pour évaluer le soutien général de la direction, les enseignant·e·s ont répondu à trois questions (ex. « *Do you receive help and support from your closest superior when needed ?* ») sur une échelle à cinq point allant de *Presque jamais* à *Toujours*. L'alpha est égal à .80.
- * **Compétences de gestion de classe** : Les enseignant·e·s ont évalué leurs compétences de gestion de classe en répondant à 6 questions (ex. « *To what extent you are able to calm a disruptive and noisy children* ») sur une échelle à 9 points allant de *Pas du tout* à *Très bien*. L'alpha est égal à .85.

L'implantation du programme a été mesurée via un livret que les enseignant·e·s devaient remplir après chaque leçon (donc une fois par mois pendant 10 mois). Les mesures prises dans le livret sont les suivantes :

- * **Fréquence des leçons (quantité d'implantation)** : Les enseignant·e·s devaient indiquer quelles leçons ils et elles avaient effectivement données parmi les 10 prévues.
- * **Adhérence aux leçons (quantité d'implantation)** : Les enseignant·e·s devaient indiquer pour chaque activité s'ils et elles l'avaient réalisée ou non. Cela permet de calculer un pourcentage d'adhérence aux leçons.
- * **Durée des leçons (quantité d'implantation)** : Les enseignant·e·s devaient indiquer après chaque leçon le temps qu'elle avait duré en minutes.
- * **Temps de préparation des leçons (qualité d'implantation)** : Les enseignant·e·s devaient indiquer pour chaque leçon le temps en minutes qu'ils et elles avaient mis à la préparer.
- * **Engagement des élèves (qualité d'implantation)** : Les enseignant·e·s devaient évaluer la proportion d'élèves qui avaient participé avec enthousiasme à la leçon sur une échelle à 4 points (0-25 %, 25-50 %, 50-75 % ou 75-100 %).

Effets de *KiVa* lors de la phase d'évaluation RCT de 2008-2009 (années 1-3 et 7-9)

1. Des effets ont-ils été observés ?

Dans l'article de Kärnä et al. (2013), bien que des mesures aient également été prises après 4 mois, seuls les effets du programme après 7 mois d'implantation sont développés (voir tableau 1). Chez les élèves de 2^{ème} et 3^{ème} année, il y a un effet positif de l'intervention sur le harcèlement. Pour la victimisation, l'effet du programme dépend du genre de l'élève et de la proportion de garçons dans sa classe. La réduction de victimisation pour les élèves du groupe expérimental est uniquement présente chez les filles dans des classes avec une proportion moyenne ou élevée de garçons (chez les garçons et dans des classes avec peu de garçons, il n'y a pas d'effet). Chez les élèves de 8^{ème} et 9^{ème} année, l'intervention ne réduit ni la victimisation ni le harcèlement auto-rapporté. Elle réduit la victimisation rapportée par les pairs, mais uniquement chez les élèves les moins âgés (ayant au maximum l'âge moyen des élèves de 8^{ème} année). Le harcèlement rapporté par les pairs est réduit suite à l'intervention, mais uniquement pour les garçons et surtout dans les classes avec un nombre important de garçons. L'intervention amène une réduction des comportements d'assistance chez les filles comme chez les garçons, mais de façon plus importante chez ces derniers et permet de diminuer les comportements de renforcement uniquement chez les garçons. D'après les analyses de Williford et al. (2013), le programme permet de réduire la cybervictimisation, peu importe le genre et l'âge. Par contre, l'effet du programme sur le cyberharcèlement n'est plus significatif chez les enfants âgés de plus de 12,87 ans.

Il y a un effet significatif de l'intervention dans le sens opposé à celui prévu pour les comportements de défense. Ces comportements diminuent davantage dans le groupe expérimental. Cet effet augmente avec l'âge moyen de la classe dans laquelle l'élève se trouve.

Tableau 4. Taille des effets (OR) après 7 mois d'implantation de *KiVa*

Victimisation auto-rapportée	2 ^{ème} et 3 ^{ème} année	Garçons	1,04 (0,79-1,30)
		Filles	1,04 (0,79-1,30)
	8 ^{ème} et 9 ^{ème} année		1,04 (0,86-1,22)
Harcèlement auto-rapporté	2 ^{ème} et 3 ^{ème} année		1,43 (1,10-1,77)
	8 ^{ème} et 9 ^{ème} année		1,08 (0,88-1,28)
Cybervictimisation auto-rapportée	Ensemble des élèves		1,29 (1,05-1,57)
Cyberharcèlement auto-rapporté	Élèves d'environ 11,3 ans		1,69 (1,15-2,48)

Inspiré de Kärnä et al. (2013, pp. 546) et de Williford et al. (2013, pp. 827-828). Les OR significatifs sont en gras et leurs intervalles de confiance à 95 % sont indiqués entre parenthèses.

Annexe 1 Evaluation de *KiVa*

Yang & Salmivalli (2015) se sont intéressées à l'effet du programme *KiVa* sur les élèves auteur·e·s et victimes de harcèlement, mais également chez les élèves qui présentent à la fois des niveaux élevés de harcèlement et de victimisation. Selon, leurs résultats, qui porte sur l'ensemble des élèves y compris ceux de la 4^{ème} à la 6^{ème} année, le programme a un effet positif chez ces trois catégories d'élève et cet effet est même plus important pour les élèves auteur·e·s et victimes de harcèlement (voir tableau 5). Il semble donc que, malgré les difficultés plus importantes que présentent généralement ces élèves, le programme permet de les aider, même sans spécifiquement les viser.

Tableau 5. Évolution du nombre d'élèves catégorisé·e·s comme harceleur·euse·s, victimes et harceleur·euse·s-victimes suite à l'implantation de *KiVa* chez les élèves de la 2^{ème} à la 4^{ème} année

	Mai 2008	Mai 2009	Changement	Différence en faveur des écoles <i>KiVa</i>	OR (IC à 95 %)
Prévalence selon les mesures auto-rapportées					
Harceleur·euse·s					
Contrôle	3,3 %	2,6 %	-21 %	-12 %	1,26 (1,10-1,41)
Intervention	3,6 %	2,4 %	-33 %		
Victimes					
Contrôle	10,2 %	7,2 %	-29 %	-14 %	1,35 (1,23-1,46)
Intervention	10,1 %	5,8 %	-43 %		
Harceleur·euse·s -victimes					
Contrôle	2,6 %	1,9 %	-27 %	-8 %	1,51 (1,29-1,73)
Intervention	2,3 %	1,5 %	-35 %		
Prévalence selon les mesures hétéro-rapportées					
Harceleur·euse·s					
Contrôle	7,0 %	7,5 %	+7 %	-14 %	1,20 (1,06-1,34)
Intervention	7,4 %	6,9 %	-7 %		
Victimes					
Contrôle	7,1 %	7,2 %	+1 %	-6 %	1,24 (1,10-1,39)
Intervention	7,3 %	6,9 %	-5 %		
Harceleur·euse·s -victimes					
Contrôle	1,7 %	2,3 %	+35 %	-41 %	1,63 (1,33-1,92)
Intervention	1,6 %	1,5 %	-6 %		

Inspiré de Yang & Salmivalli (2015, pp. 86-87). Les OR significatifs sont en gras et leurs intervalles de confiance à 95 % sont indiqués entre parenthèses.

Concernant l'effet des actions indiquées (Garandeau, Poskiparta & Salmivalli 2014), les élèves pris-es en charge après s'être fait harceler rapportent dans 78,2 % des cas que le harcèlement a stoppé, dans 19,5 % des cas qu'il a diminué et dans 2,1 % des cas que la situation n'a pas changé (seul un élève rapporte que la

Annexe 1 Evaluation de *KiVa*

situation s'est empirée). L'approche avec confrontation est plus efficace, mais uniquement dans les écoles secondaires. Pour les élèves de primaire, il n'y a pas de différence entre les deux approches. Pour les cas gérés via l'approche avec confrontation, la durée de la victimisation a un impact sur la probabilité de réussite (l'approche est plus efficace pour les cas de victimisation qui ne dure pas depuis longtemps) alors que ce n'est pas le cas pour l'approche sans confrontation. Dans tous les cas, comparativement aux autres formes de harcèlement, il est plus probable que le harcèlement continue malgré une intervention s'il se marque par de l'exclusion sociale. L'intention de changer de comportement des élèves qui ont harcelé n'est pas influencée par l'approche utilisée. Par contre, le fait que l'élève soit en secondaire, perçoive que l'adulte a essayé de susciter son empathie ou perçoive que l'adulte a condamné son comportement est lié à une intention plus importante de changer de comportement.

Selon les résultats de Haataja et al. (2016), seulement 23,6 % des élèves qui ont rapporté être harcelés au prétest comme au post-test ont été identifié·e·s comme victime de harcèlement par les adultes de l'école et ont fait l'objet d'une intervention. 30 % de la variance au niveau de l'identification des élèves peut être attribuée à l'école et 70 % aux caractéristiques individuelles des élèves. Les élèves sont davantage pris en charge par les adultes s'ils et elles sont reconnu·e·s par leurs pairs comme étant victimisé·e·s et/ou ont parlé de leur situation à un adulte. Les élèves victimes sont également davantage identifié·e·s dans les écoles avec un engagement fort envers le programme. Par contre, le fait d'avoir un score élevé de harcèlement diminue la probabilité d'être reconnu comme victime. Les garçons sont plus souvent reconnus, mais uniquement en primaire. Le type de victimisation, l'âge de l'élève et la taille de l'école n'ont pas d'influence.

Ahtola et al. (2012) ont montré que les enseignant·e·s dans les écoles *KiVa* se sentaient plus compétent·e·s pour gérer les situations de harcèlement que ceux et celles dans les écoles contrôles (par contre, il n'y a pas de différence au niveau de la confiance envers le programme ou de la perception du harcèlement comme un phénomène malléable). Dans les écoles *KiVa*, les enseignant·e·s qui font partie de l'équipe *KiVa* ont une confiance plus élevée envers le programme que ceux et celles qui n'en font pas partie et les enseignant·e·s qui donnent des leçons *KiVa* aux élèves se sentent plus compétent·e·s et perçoivent davantage le harcèlement comme un phénomène malléable que les autres.

2. Les effets sont-ils différenciés ?

L'effet de l'intervention dépend de l'âge et du niveau scolaire des élèves, de leur genre, et de la composition des classes. Alors que, chez les élèves de 2^{ème} et 3^{ème} année, la réduction du harcèlement n'est pas influencée par ces facteurs, la réduction de la victimisation est, elle, uniquement présente chez les filles dans des classes avec une proportion moyenne ou élevée de garçons. Chez les élèves de 8^{ème} et 9^{ème} année, l'intervention réduit la victimisation rapportée par les pairs, mais uniquement chez les élèves les moins âgé·e·s tandis que le harcèlement rapporté par les pairs est uniquement réduit pour les garçons et surtout dans les classes avec un nombre important de garçons. L'effet de l'intervention sur les rôles adoptés par les témoins dépend également du genre avec une réduction de l'assistance plus importante chez les garçons ($d=0.18$, contre $d=0.06$ chez les filles) et une réduction du renforcement uniquement présente pour eux. Le cyberharcèlement diminue uniquement chez les élèves les plus jeunes.

3. Des mesures de suivi ont-elles été prises ?

Les mesures ont uniquement été prises après 4 et 7 mois d'implantation (ou deux semaines après l'intervention pour les actions indiquées). Des données ont normalement continué à être récoltées après cela, mais, pour l'instant, elles ne sont rapportées nulle part.

4. L'implantation a-t-elle été mesurée ?

L'implantation du programme a été mesurée après chaque leçon pour les enseignant·e·s du primaire ainsi qu'à la fin de l'année pour tou·te·s les enseignant·e·s. En moyenne 9 des 10 leçons ont été dispensées dans les écoles primaires et 5 des 6 composantes (en comptant l'introduction, les quatre leçons et la conclusion) l'ont été dans les écoles secondaires. La fidélité d'implantation est donc assez importante (Kärnä et al., 2013). Plus précisément, selon l'étude de Haataja et al. (2014) portant sur les enseignant·e·s de la 1^{ère} à la 6^{ème} année, les leçons ont en moyenne duré 79 minutes, mis 29 minutes à préparer et comprenaient 68 % des activités possibles. Selon Ahtola et al. (2013), 66,7 % des enseignant·e·s de la 1^{ère} à la 3^{ème} année ont implanté les dix leçons et environ 80 % du programme a été implanté avec une durée moyenne totale de 11 h (sur les 15 h prévues). Déjà durant la première année, le niveau d'implantation du programme a tendance à diminuer (Haataja et al., 2014).

Pour les actions indiquées, les chercheur·euse·s ont vérifié, au cours des réunions avec les réseaux d'écoles, que la méthode était implantée comme prévu. De plus, les questionnaires remplis par les élèves ayant participé à des actions indiquées suite à des comportements de harcèlement confirment qu'ils et elles ont plus eu le sentiment de se faire blâmer dans la condition avec confrontation.

Selon Ahtola et al. (2013), le support de la direction envers le programme anti-harcèlement est positivement corrélé à l'implantation du programme (par contre, ce n'est pas le cas du support général de la direction). Haataja et al. (2015) ont remarqué que certain·e·s enseignant·e·s (53 %) implantaient le programme de façon élevée, et ce tout au long de l'année. D'autres (30 %) implantaient le programme de façon modérée constamment au long de l'année. Et finalement, les dernier·ère·s (17 %) implantaient le programme de façon importante en début d'année, mais diminuaient nettement leurs efforts par la suite, de façon à n'avoir au final délivré qu'environ 50 % du contenu. Le support de la direction est associé à un maintien des efforts dans le temps tandis que les croyances personnelles sont associées au fait de commencer à implanter le programme de façon importante. La formation, les capacités de gestion de la classe et l'engagement des élèves ne sont associés ni à l'un ni à l'autre.

D'après les résultats de Haataja et al. (2014), la quantité et la qualité de l'implantation ont un effet sur la victimisation (mais pas sur le harcèlement). Une adhérence importante aux leçons et un temps de préparation élevé permettent une diminution plus importante de la victimisation. Pour chaque augmentation d'un écart-type dans le degré d'adhérence (c'est-à-dire environ 20 % des activités), la victimisation diminue de 0,17 écart-type supplémentaire et pour chaque augmentation d'un écart-type dans le temps de préparation (environ 16 min), la victimisation diminue de 0,21 écart-type supplémentaire. Par contre, la durée de leçons n'a pas d'effet sur la victimisation.

Qualité de l'évaluation de *KiVa* lors de la phase de dissémination nationale de 2009-2010

Article principal : Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Alanen, E., & Salmivalli, C. (2011a). Going to scale: A nonrandomized nationwide trial of the *KiVa* antibullying program for Grades 1–9. *Journal of Consulting and Clinical Psychology*, 79(6), 796–805. doi:10.1037/a0025740

1. Quel était le plan de recherche de l'évaluation ?

Le plan de recherche de l'étude de Kärnä et al. (2011a) est quasi-expérimental. Il s'agit d'une étude longitudinale par cohortes adjacentes. Le programme a été implanté pendant l'année scolaire 2009-2010, avec une mesure pré-test prise en mai 2009 et une mesure post-test en mai 2010. Les résultats post-test de chaque cohorte sont comparés aux résultats pré-test de la cohorte de l'année supérieure dans la même école. Par exemple, les résultats de mai 2010 des élèves qui étaient en 1^{ère} année en 2009-2010 sont comparés avec les résultats de mai 2009 des élèves qui étaient en 2^{ème} année en 2009-2010). Cela permet, en l'absence d'un groupe contrôle, que les effets de maturation ne viennent pas compliquer l'interprétation des résultats. Afin d'évaluer les effets du programme, les OR de victimisation et harcèlement ont été calculés pour chaque niveau scolaire.

2. Quelles étaient les caractéristiques de l'échantillon ?

Pour recruter les écoles participantes, des lettres décrivant le programme *KiVa* ont été envoyées en 2008 à toutes les 3218 écoles finlandaises n'ayant pas encore implanté le programme. 1827 écoles se sont portées volontaires et 1450 d'entre elles ont pu implanter le programme lors de l'année scolaire 2009-2010. Sur les 1450 écoles participantes, 888 écoles ont participé à l'évaluation au pré et au post-test et composent l'échantillon final. Les écoles ayant quitté l'étude présentent des taux de victimisation et de harcèlement au prétest légèrement plus importants que les autres écoles. L'échantillon est représentatif des écoles finlandaises au niveau de la proportion d'écoles primaires/secondaires et de langue finlandaise/suédoise ainsi que du nombre d'élèves et de classes. Certains élèves ont été exclus des analyses, car leurs réponses étaient clairement contradictoires et les élèves ayant des données manquantes sur une mesure n'ont pas été pris en compte dans l'analyse concernant la variable en question. Les analyses sur la victimisation ont été réalisées sur 156634 élèves dans la condition contrôle et 141103 élèves dans la condition expérimentale et celles sur le harcèlement sur 156629 élèves dans la condition contrôle et 141099 élèves dans la condition expérimentale.

3. Par qui et où l'évaluation a-t-elle été menée ?

Les auteurs de l'étude sont les mêmes que ceux ayant évalué l'efficacité du programme en 2007-2008 et 2008-2009 : Antti Kärnä, Elisa Poskiparta et Christina Salmivalli (University of Turku), Todd Little (University of Kansas) et Marinus Voeten (Radboud University). Le contexte est également le même avec toutefois un échantillon plus étendu.

4. Quelles et combien de mesures ont été prises ?

Le prétest a eu lieu en mai 2009 et le post-test en mai 2010. Les élèves ont rempli un questionnaire sur ordinateur, exactement comme pour les études de 2007-2008 et de 2008-2009. Les premières questions portaient sur des données sociodémographiques (genre, âge, classe, école). Ensuite, la définition du

Annexe 1 Evaluation de KiVa

harcèlement (comme proposée par Olweus, 1996; cité par Kärnä et al., 2011a) leur était présentée. Les élèves ont répondu aux deux items globaux du OBVQ («*How often have you been bullied at school in the last couple of months ?*» et «*How often have you bullied others at school in the last couple of months ?*») afin de mesurer la victimisation et le harcèlement. Les élèves y ont répondu via une échelle à 5 points (*Jamais, Seulement une ou deux fois, Deux ou trois fois par mois, Environ une fois par semaine, Plusieurs fois par semaine*). Les élèves ont été catégorisé·e·s comme victime et harceleur·euse en utilisant le point de coupure de *2 ou 3 fois par mois* suggéré par Solberg et Olweus (2003; cité par Kärnä et al., 2011a).

Effets de KiVa lors de la phase de dissémination nationale de 2009-2010

1. Des effets ont-ils été observés ?

Sur l'échantillon global (tout âge confondu), les OR sont significatifs et valent 1,21 pour la victimisation et 1,18 pour le harcèlement. Cela correspond à une réduction de 15 % de la victimisation et de 14 % du harcèlement. Toutefois, l'âge influence les effets du programme. À partir de la 1^{ère} année, les effets augmentent pour atteindre leur plus haut point en 4^{ème} année. Ensuite, ils diminuent pour devenir non significatifs à partir de la 6^{ème} (sauf pour la victimisation en 8^{ème} année). Les effets obtenus sont globalement moins importants que dans les études d'évaluation de 2007-2008 et de 2008-2009 (voir tableau 6).

Tableau 6. Effets de KiVa sur la victimisation et le harcèlement auto-rapporté en fonction de l'âge

	Mai 2009 (contrôle)	Mai 2010 (KiVa)	OR (IC à 95 %)
Victimisation auto-rapportée			
1 ^{ère} année	25,9 %	16,6 %	1,17 (1,08-1,26)
2 ^{ème} année	23,5 %	20,0 %	1,23 (1,13-1,34)
3 ^{ème} année	19,1 %	16,1 %	1,23 (1,13-1,35)
4 ^{ème} année	16,2 %	12,7 %	1,33 (1,20-1,46)
5 ^{ème} année	13,5 %	10,7 %	1,30 (1,18-1,44)
6 ^{ème} année	11,8 %	9,8 %	1,22 (1,09-1,37)
7 ^{ème} année	11,6 %	10,6 %	1,11 (0,97-1,26)
8 ^{ème} année	10,0 %	8,8 %	1,15 (1,00-1,32)
9 ^{ème} année	9,3 %	8,3 %	1,13 (0,98-1,31)
Total	16,3 %	13,8 %	1,21 (1,14-1,29)
Harcèlement auto-rapporté			
1 ^{ère} année	13,6 %	12,0 %	1,15 (1,04-1,28)
2 ^{ème} année	12,5 %	9,9 %	1,30 (1,16-1,47)
3 ^{ème} année	9,6 %	7,7 %	1,28 (1,12-1,46)
4 ^{ème} année	9,4 %	7,2 %	1,34 (1,15-1,55)
5 ^{ème} année	8,7 %	7,2 %	1,23 (1,06-1,42)
6 ^{ème} année	10,0 %	8,5 %	1,19 (1,03-1,38)
7 ^{ème} année	11,3 %	10,8 %	1,05 (0,90-1,23)

Annexe 1 Evaluation de KiVa

8 ^{ème} année	11,9 %	11,5 %	1,05 (0,90-1,21)
9 ^{ème} année	13,2 %	12,5 %	1,07 (0,93-1,23)
Total	11,1 %	9,6 %	1,18 (1,09-1,27)

Inspiré de Kärnä et al (2011a, p. 799) Les OR significatifs sont en gras et leurs intervalles de confiance à 95 % sont indiqués entre parenthèses.

Selon les estimations de Salmivalli et al. (2013), durant cette première année de dissémination nationale, le programme KiVa a permis d'éviter que 2300 élèves harcèlent et que 3900 élèves se fassent harcelé·e·s, ce qui, généralisé à l'ensemble de la population scolaire finlandaise, représente une réduction d'environ 7500 élèves harceleur·euse·s et 12500 élèves victimes.

2. Les effets sont-ils différenciés ?

L'effet de l'intervention sur la victimisation et le harcèlement dépend de l'âge des élèves. La diminution est la plus importante chez les enfants de 4^{ème} année et à partir de la 7^{ème} année, il n'y a plus d'effet significatif (mis à part pour la victimisation en 8^{ème} année). Il n'y a par contre pas d'influence du genre, du type d'éducation (enseignement ordinaire ou spécialisé) ou la langue d'enseignement (suédois ou finlandais).

3. Des mesures de suivi ont-elles été prises ?

L'étude rapporte uniquement des données prises à la fin de l'année scolaire. Toutefois, les chercheurs ont continué à récolter des données, mais pour le moment celles-ci ne sont pas publiées.

4. L'implantation a-t-elle été mesurée ?

Au niveau des actions de prévention universelle, les enseignant·e·s de primaire ont rapporté avoir implanté en moyenne 8 des 10 leçons prévues. Le jeu vidéo a été joué par les élèves dans 69,6 % des classes. Les enseignant·e·s de secondaire ont rapporté avoir passé 9 h au total sur les thèmes KiVa. Les élèves de secondaire ont été sur KiVaStreet dans 62,1 % des écoles. Les écoles ont rapporté avoir traité en moyenne 5,69 cas de harcèlement via une action indiquée. Globalement, le temps consacré au programme est donc inférieur au temps recommandé. De plus, il y a une variabilité importante d'une école et d'un·e enseignant·e à l'autre. Les analyses ont révélé une corrélation positive entre le temps dédié au programme et la réduction du harcèlement et de la victimisation, que ce soit pour les actions de prévention ou les actions indiquées. La fidélité d'implantation est plus faible lors de cette phase de dissémination nationale qu'elle ne l'était lors des études d'évaluation de 2007-2008 et 2008-2009 (Salmivalli et al., 2013).

Implantation de *KiVa* aux Pays-Bas

Le programme *KiVa* est progressivement introduit aux Pays-Bas dans l'enseignement primaire ordinaire. Son implantation est financée par le ministère de l'Éducation (Onderwijs Bewijs). Afin d'assurer une mise en œuvre réussie, un consortium *KiVa* a été créé en 2011. Il est composé de chercheurs de l'University of Groningen, d'équipe de consultation et de bureaux d'études.

L'étude sur l'efficacité de *KiVa* aux Pays-Bas a été réalisée dans un contexte particulier. En novembre et décembre 2012, deux élèves (Tim Ribberink et Fleur Bloemen) se sont suicidé·e·s suite au harcèlement qu'il et elle subissaient. Cela a déclenché de nombreuses discussions sur le harcèlement et a sans doute amené toutes les écoles à augmenter leurs efforts pour lutter contre le harcèlement. Cela expliquerait la diminution du harcèlement observée également dans les écoles contrôles entre 2012 et 2014. Ce qui est frappant en tout cas, c'est que l'amélioration absolue dans les écoles *KiVa* aux Pays-Bas (63 %) est meilleure qu'en Finlande (40 %), mais que les différences entre les écoles *KiVa* et les écoles contrôles sont moindres aux Pays-Bas (11 % de différence) qu'en Finlande (29 % de différence).

Qualité de l'évaluation de *KiVa* aux Pays-Bas

Article principal : Veenstra, R. (2015). *Signaleren en tegengaan van pesten: het KiVa antipestprogramma: Eindrapportage voor Onderwijs Bewijs*. Rijksuniversiteit Groningen.

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude longitudinale avec un design RCT. Après la prise des données prétest, le *Netherlands Bureau for Economic Policy Analysis* a aléatoirement assigné les écoles à la condition contrôle ou à l'une des deux conditions d'intervention en utilisant une procédure de stratification bloquée. Cela veut dire que les écoles ont été réparties par bloc de trois en fonction des niveaux moyens de victimisation, de harcèlement, d'attitudes envers le harcèlement et de bien-être à l'école. Ensuite dans chaque bloc, une école a été assignée par tirage au sort à la condition contrôle, une école à la condition *KiVa* et une école à la condition *KiVa+*. De cette façon, les différences au prétest entre les conditions sont minimisées. La condition *KiVa+* est une variante spécifique aux Pays-Bas : les enseignant·e·s reçoivent en début d'année un rapport portant sur les réseaux sociaux au sein de leur classe. Cela leur permet d'acquérir une compréhension détaillée des relations positives et négatives dans leur classe et de savoir quel·le·s élèves sont victimes de harcèlement. L'hypothèse derrière cette condition est qu'une meilleure connaissance de leur classe permettrait aux enseignant·e·s de lutter plus efficacement contre le harcèlement (qui passe souvent inaperçu autrement) et d'améliorer le climat de classe.

Les effets du programme ont été analysés via un calcul de l'*Intent-to-Treat* (différence entre les écoles assignées à la condition contrôle et celles assignées à la condition expérimentale) et du *Treatment-to-Treated* (différence entre les écoles de la condition contrôle et les écoles de la condition expérimentale dans lesquelles les leçons *KiVa* ont effectivement été données).

2. Quelles étaient les caractéristiques de l'échantillon ?

Pour recruter les écoles, des lettres décrivant le programme ont été envoyées à toutes les écoles primaires d'enseignement ordinaire des Pays-Bas en automne 2011. Sur les 6938 écoles contactées, 132 écoles se sont portées volontaires. Environ un tiers de ces écoles disposaient déjà d'un programme de lutte contre le harcèlement, mais ont jugé qu'elles avaient besoin d'un programme plus efficace. Finalement, seules 100 écoles ont participé au prétest, les autres ont laissé tomber par manque de temps ou de moyens. Une école a également quitté l'étude après l'assignation aux conditions. Les analyses ont donc été réalisées sur 99 écoles, 33 dans chaque condition. Les écoles proviennent de toutes les provinces des Pays-Bas, et sont situées dans des zones rurales ou semi-urbaines. 56 % des écoles sont des écoles publiques (ce qui est plus que la moyenne nationale de 38 %).

Tout·e·s les élèves de la quatrième à la septième année qui avaient reçu le consentement parental ont participé à l'étude. La 4ème année aux Pays-Bas correspond à la 2ème primaire en Belgique et la 7ème année à la 5ème. 10838 élèves (3536 dans la condition contrôle et 7302 dans une des conditions expérimentales) réparti·e·s dans 460 classes ont participé à l'étude. Les élèves ont entre 7 et 13 ans, environ 50 % sont des garçons et 93 % sont né·e·s aux Pays-Bas. Le nombre exact d'élèves inclus·es dans les analyses varie entre 6142 et 9343 suivant l'étude sauf pour l'étude d'Oldenburg, Van Duijn et Veenstra (2018) qui porte uniquement sur un sous-échantillon de sept classes de 3ème année (élèves de 8-9 ans) et celle de Huitsing et al. (2018) qui porte uniquement sur les 4356 élèves de 3ème et 4ème année.

Pour l'étude de van der Ploeg, Steglich & Veenstra (2016), seules les écoles dans lequel le programme a été implanté ont été prises en considération. L'approche par groupes de soutien a été utilisée pour 56 élèves victimes (30 filles et 26 garçons) dans 28 écoles. Ces élèves ont été apparié·e·s avec des élèves faisant partie des mêmes écoles, mais n'ayant pas été pris en charge via une action indiquée. L'appariement a été réalisé en fonction de la fréquence de la victimisation, des niveaux de dépression et de bien-être à l'école, du genre et du niveau scolaire. Les élèves pour lequel·le·s aucune correspondance exacte n'a été trouvée ainsi que ceux et celles n'ayant pas indiqué être victime de harcèlement au prétest ont été exclu·e·s de l'analyse, tout comme les cas pour lesquels l'issue de la situation était inconnue. L'échantillon final est composé de 38 élèves ayant fait l'objet d'une intervention (63,3 % de filles, âge moyen de 9,24 ans) et de 571 élèves apparié·e·s.

3. Par qui et où l'évaluation a-t-elle été menée ?

L'évaluation est menée au Pays-Bas par une équipe de chercheur·euse·s de l'University of Groningen (Gijs Huitsing, Tessa Kaufman, René Koens, Tina Kretschmer, Gerine Lodder, Femke Munniksmä, Beau Oldenburg, Christian Steglich, Miranda Sentse, Rozemarijn van der Ploeg, Marijtje Van Duijn, René Veenstra et Freek Velthausz) ainsi que par Jaana Juvonen (UCLA), Christina Salmivalli (University of Turku) et Hannah Schacter (University of Southern California). Certain·e·s de ces chercheur·euse·s étaient déjà impliqué·e·s dans les études portant sur les données finlandaises.

4. Quelles et combien de mesures ont été prises ?

Les données ont été recueillies à cinq reprises : en mai 2012, octobre 2012, mai 2013, octobre 2013 et mai 2014. Comme pour les études finlandaises, les élèves ont rempli les questionnaires en ligne en se connectant via un mot de passe durant les heures scolaires. Ils et elles étaient encadré-e-s par leurs enseignant-e-s qui avaient reçu des instructions détaillées et étaient encouragé-e-s à aider les élèves sans les influencer. Chaque élève lisait et répondait aux questions après avoir visionné des vidéos d'explications. Le questionnaire contenait cinq vidéos dans lesquelles une actrice professionnelle expliquait la raison d'être du questionnaire (recherche sur le bien-être des enfants à l'école), la façon de le remplir (avec un exemple) et un certain nombre de règles de base (ex. les élèves ne doivent pas se parler en remplissant le questionnaire). Il leur était également expliqué que leurs réponses resteraient confidentielles, mais que leur enseignant-e recevrait des commentaires pour s'assurer que tou-te-s les élèves aiment aller à l'école. Le questionnaire commençait par l'*OBVQ*, définition comprise. Pour le reste, l'ordre des questions et des échelles était aléatoire afin d'éviter tout effet d'ordre. Les dernières questions portaient sur des données sociodémographiques (genre, lieu et date de naissance, frères et sœurs, niveau scolaire).

Les mesures qui ont été prises sont les suivantes :

- **Victimisation** : La victimisation a été mesurée via l'item global du OBVQ (Olweus, 1996; cité par Veenstra, 2015). Les élèves ont dû indiquer combien de fois ils ou elles avaient été victimes de harcèlement depuis les vacances de Noël avec les choix de réponse suivants : *Jamais*, *Une ou deux fois*, *Deux ou trois fois par mois*, *Environ une fois par semaine* et *Plusieurs fois par semaine*. 11 items portant sur des formes spécifiques de victimisation ont également été utilisés, mais 3 d'entre eux (ceux portant sur le harcèlement raciste, le harcèlement sexuel et les autres type d'harcèlement) ont, par la suite, été ôtés de l'analyse étant donné leur fréquence très faible. Les élèves qui ont rapporté être victimes de harcèlement devaient également nommer les élèves qui les harcelaient ainsi que ceux et celles qui les défendaient. La différence entre le nombre de défenseurs en octobre et en mai a permis de calculer un indicateur de changement dans la défense allant de 0 (la défense s'est complètement arrêtée) à 3 (la défense a augmenté).
- **Harcèlement** : Le harcèlement a été mesuré de la même façon que la victimisation, mais avec les items formulés de manière active.
- **Anxiété sociale** : L'anxiété sociale a été mesurée via 7 items (ex. « *I am scared to talk to someone I don't know* ») dérivés du *Social Phobia Screening Questionnaire* développé par Furmark et al. (1999; cité par Kaufman, Kretschmer, Huitsing, & Veenstra, 2018).
- **Attitude face au harcèlement** : L'attitude face au harcèlement a été mesurée via 6 items venant d'une échelle développée par Rigby et Slee (1991; cité par Veenstra, 2015) : « *Je trouve ennuyeux que personne n'aide un enfant victime d'harcèlement* », « *Je pense que les personnes qui harcèlent sont lâches* », « *Je pense qu'il est bon d'aider les enfants qui ne peuvent pas se défendre* » ; « *Je pense que participer au harcèlement est une erreur* », « *Je me sens honteux quand je vois quelqu'un être victime d'harcèlement* » et « *Je pense que c'est une bonne chose si quelqu'un aide les enfants victimes de harcèlement* ». Les enfants ont répondu sur une échelle à 5 points allant de *Jamais* à *Toujours*.

Annexe 1 Evaluation de KiVa

- **Bien-être à l'école** : Le bien-être à l'école a été mesuré via les mêmes 7 items que Kärnä et al. (2011b).
- **Changement dans la victimisation** (lors des actions indiquées) : Lors de la réunion de suivi dans le cadre des actions indiquées, les élèves devaient dire si la victimisation avait totalement stoppé, diminué, était restée stable ou avait augmenté. Le changement dans la victimisation a également été calculé en faisant la différence entre les scores de fréquence de la victimisation et du nombre d'élèves par qui l'élève est harcelé·e obtenus en octobre et en mai.
- **Comportements externalisés** : Les comportements externalisés ont été mesurés via 13 items de l'échelle *Youth Self Report Conduct Problem Scale* (Achenbach et al., cité par Kaufman et al., 2018). Les élèves devaient répondre à chaque item (ex. « *I break rules at school or elsewhere* ») sur une échelle à 3 points allant de *Jamais* à *Souvent*.
- **Confiance en soi** : La confiance en soi a été mesurée via une échelle à 5 items développée par Rosenberg (1965; cité par Veenstra, 2015).
- **Contrôle de soi** : Le contrôle de soi a été mesuré via 8 items issus du *Temperament in Middle Childhood Questionnaire* (Simonds and Rothbart, 2004 ; cité par Kaufman et al., 2018). Les élèves devaient répondre à chaque item (ex. « *When someone tells me Stop, I can stop* ») sur une échelle à 5 points allant de *Jamais* à *Toujours*.
- **Estime de soi** : L'estime de soi a été mesurée via les items positifs de l'échelle *Rosenberg Self Esteem Scale* (Rosenberg, 1965; cité par Kaufman et al., 2018). Les élèves devaient répondre à chaque item (ex. « *On the whole, I am satisfied with myself* ») sur une échelle à 5 points allant de *Jamais* à *Toujours*.
- **Importance d'être un·e bon·ne ami·e** (objectifs sociaux des élèves) : L'importance d'être un·e bon·ne ami·e a été mesurée via trois items (« *Je veux savoir comment être un bon ami* », « *J'aime apprendre à mieux m'entendre avec mes amis* » et « *Je veux savoir comment rester les meilleurs amis* »).
- **Importance de la popularité** (objectifs sociaux des élèves) : L'importance de la popularité a été mesurée via trois items (« *Je veux que les autres enfants pensent que je suis populaire* », « *Je veux des amis qui sont populaires* » et « *Je veux être ami avec les enfants populaires* »).
- **Importance d'éviter de faire de mauvaises choses** (objectifs sociaux des élèves) : L'importance d'éviter de faire de mauvaises choses a été mesurée via trois items (« *J'essaie de ne pas faire de choses qui pourraient amener les autres enfants à m'harceler* », « *J'essaie de ne pas faire de choses qui attirent l'attention sur moi* » et « *Je ne veux pas que les autres enfants disent de mauvaises choses à mon sujet* »).
- **Perception de l'enseignant·e par les élèves** : La perception de l'enseignant·e a été mesurée à l'aide de 3 mêmes items que dans l'étude de Veenstra et al. (2014), c'est-à-dire « *What does your teacher think of bullying?* » (attitudes), « *How much can the teacher do in order to decrease bullying?* » (efficacité) et « *How much has the teacher done in order to decrease bullying since last autumn?* » (efforts). Les élèves ont dû répondre sur une échelle à 5 points allant de *C'est une bonne chose* à *C'est une très mauvaise chose* ou de *Rien* à *Vraiment beaucoup*.

Annexe 1 Evaluation de KiVa

- **Relations avec les parents** : La chaleur et le rejet parental ont été évalués via l'échelle *Warmth and Rejection Scale* (Arrindell et al., 1983; cité par Kaufman et al., 2018). Les élèves ont dû répondre à une série d'items concernant la chaleur et le rejet maternel et paternel (ex. « *Is your mother/father sometimes harsh and unkind to you ?* »). Étant donné leur corrélation importante, les items concernant les mères et pères ont été moyennés pour former un score global de chaleur parentale et un score global de rejet parental.
- **Relations dans la classe** : Les élèves ont dû donner les noms des élèves de leur classe qui étaient leurs meilleur·e·s ami·e·s ainsi que ceux et celles qu'ils et elles n'aimaient pas et ceux et celles qu'ils et elles percevaient comme étant les plus populaires. Dans les deux cas, un nombre illimité d'élèves de la classe pouvaient être nommé·e·s.
- **Sentiments dépressifs** : Les sentiments dépressifs ont été mesurés via une échelle à 9 items développée par Chorpita et al. (2000; cité par Veenstra, 2015).

Les alphas sont supérieurs à .70 pour toutes les mesures. D'autres mesures ont également été prises, mais sans être exploitées dans les études actuellement publiées. Il s'agit des rôles dans les situations de harcèlement, de l'efficacité personnelle à défendre, de l'empathie envers la victime, de la responsabilité, de l'acceptation par les pairs, du leadership, de l'utilisation d'internet ainsi que d'une évaluation des élèves par les enseignant·e·s et les parents (University of Groningen, 2015). Les élèves ont également répondu à un questionnaire portant sur les rôles joués par les pairs dans les situations de harcèlement. Leurs réponses ont permis de créer des cartes de réseau pour chaque classe. Concernant les actions indiquées, les membres de l'équipe KiVa de l'école ont dû remplir un rapport après chaque rencontre avec des élèves.

Effets de KiVa aux Pays-Bas

1. Des effets ont-ils été observés ?

Effets du programme général :

Que ce soit un ou deux ans après le prétest, la victimisation diminue dans les écoles KiVa. Une baisse est également observée dans les écoles contrôle, mais moins importante (voir tableau 7). La diminution de la victimisation est de plus grande ampleur pour les élèves se faisant plus fréquemment harceler. Le programme permet de diminuer la victimisation globale, verbale et relationnelle, mais pas la victimisation physique. Dans tous les cas, les effets sont d'assez petite taille (voir tableau 8).

Tableau 7. Évolution des taux de victimisation globale suite à l'implantation de KiVa aux Pays-Bas

Annexe 1 Evaluation de *KiVa*

	Évolution sur un an	Évolution sur deux ans	Différence en faveur des écoles <i>KiVa</i> sur un an	Différence en faveur des écoles <i>KiVa</i> sur deux ans
Victimisation chaque mois				
<i>KiVa</i>	-29 %	-58 %	-3 %	-15 %
Contrôle	-26 %	-43 %		
Victimisation chaque semaine				
<i>KiVa</i>	-40 %	-54 %	-51 %	-13 %
Contrôle	+11 %	-41 %		
Victimisation chaque jour				
<i>KiVa</i>	-60 %	-71 %	-15 %	-8 %
Contrôle	-45 %	-63 %		
Victimisation au moins une fois par mois				
<i>KiVa</i>	-51 %	-63 %	-16 %	-13 %
Contrôle	-35 %	-50 %		

Inspiré de Veenstra (2014, p. 10; 2015, p. 10). Analyses Intent-to-Treat.

Au niveau du harcèlement, le programme n'a un effet que sur le harcèlement global après deux ans (qui est moins élevé dans les écoles *KiVa* que dans les écoles contrôles). Au niveau des formes spécifiques et du score total de harcèlement, il n'y a aucune différence significative entre les conditions.

Le programme a un effet sur les perceptions qu'ont les élèves de leur enseignant·e. Alors que l'attention et les attitudes perçues de l'enseignant·e n'évoluent pas avec le temps dans la condition contrôle, les élèves des écoles *KiVa* perçoivent que leur enseignant·e porte plus d'attention au harcèlement et a des attitudes plus négatives face à celui-ci après l'implantation de *KiVa*. De plus, alors qu'au prétest il n'y avait pas de différence entre les groupes au niveau des compétences de gestion du harcèlement de l'enseignant·e, au post-test, les élèves de la condition *KiVa* perçoivent que leur enseignant·e est plus compétent·e pour gérer le harcèlement que ceux et celles de la condition contrôle. Ces effets sont présents après un an d'implantation du programme et se maintiennent l'année suivante.

KiVa a également un impact sur les attitudes des élèves et leurs buts sociaux. Après un an, tou·te·s les élèves ont des attitudes plus négatives envers le harcèlement, mais l'évolution est plus importante dans les écoles *KiVa* (toutefois les effets ne se maintiennent pas après deux ans). Que ce soit après un ou deux ans, les élèves des écoles *KiVa* accordent plus d'importance au fait d'être de bon·ne·s ami·e·s que ceux et celles des écoles contrôles. Le programme n'a pas d'effet sur les autres buts sociaux, l'anxiété sociale ou la confiance en soi. Les sentiments dépressifs sont moins élevés dans les écoles *KiVa* que dans les écoles contrôles après un an, mais cet effet ne se maintient pas dans le temps.

Tableau 8. Taille des effets (d de Cohen) après un et deux d'implantation de *KiVa*

Annexe 1 Evaluation de KiVa

	Taille des effets (d de Cohen)	
	Après un an (mai 2013)	Après deux ans (mai 2014)
Victimisation		
Victimisation globale	-0.14	-0,13
Victimisation verbale	-0.12	-0,09 (marginale- ment significatif)
Victimisation relationnelle	-0.15	-0,10
Victimisation physique	-0.03 (n.s.)	0.04 (n.s.)
Victimisation (échelle de 8 items)	-0,08 (n.s.)	-0,11
Harcèlement		
Harcèlement global	-0.06 (n.s.)	-0,11
Harcèlement verbal	-0.03 (n.s.)	-0.04 (n.s.)
Harcèlement relationnel	-0.05 (n.s.)	-0.06 (n.s.)
Harcèlement physique	-0.01 (n.s.)	-0,02 (n.s.)
Harcèlement (échelle de 8 items)	0.01 (n.s.)	-0,04 (n.s.)
Perception de l'enseignant-e		
Attention portée au harcèlement	0,58	0,46
Attitudes envers le harcèlement	0,19	0,15
Compétence pour gérer le harcèlement	0,25	0,19
Buts sociaux		
Importance d'être un bon ami	0,15	0,41
Importance de ne pas faire de mauvaises choses	-0.02 (n.s.)	-0,12 (n.s.)
Importance d'être populaire	-0.03 (n.s.)	0.12 (n.s.)
Caractéristiques des élèves		
Attitudes envers le harcèlement	0,09	0.10 (marginale- ment significatif)
Bien-être à l'école	0.11 (marginale- ment significatif)	0.03 (n.s.)
Anxiété sociale	-0.02 (n.s.)	-0,02 (n.s.)
Sentiments dépressifs	-0.08	-0,05 (n.s.)
Estime de soi	0.04 (n.s.)	0.06 (n.s.)

Inspiré de Veenstra (2015, p.9). Analyses Intent-to-Treat. Les effets significatifs sont en gras.

Les résultats de Garandeau et al. (2018) selon lesquels l'ajustement social et psychologique des élèves qui sont resté-e-s victimes de harcèlement malgré l'intervention empirique sont retrouvés par Huitsing et al. (2018). Après l'intervention, les élèves qui se font toujours harceler présentent des symptômes dépressifs plus importants et une estime de soi plus faible s'ils et elles se trouvent dans la condition expérimentale plutôt que dans la condition contrôle (il n'y a par contre pas d'effet significatif pour l'anxiété et le bien-être à l'école). Cela s'expliquerait par le fait que, dans une école où le niveau général de harcèlement diminue, les élèves qui se font toujours harceler auraient plus tendance à se blâmer et plus de difficultés à trouver des ami-e-s qui partagent leur situation.

Annexe 1 Evaluation de KiVa

Effets des actions indiquées :

van der Ploeg et al. (2016) ont étudié l'effet spécifique des actions indiquées sur la victimisation. A court-terme, les élèves étaient très positif·ve·s quant aux effets de l'action indiquée, la plupart ont rapporté que le harcèlement s'était arrêté ou avait diminué. Toutefois, à long terme, les élèves ont évalué leur situation de façon significativement plus négative (voir tableau 9).

Tableau 9. Effets des actions indiquées à court et long terme

	Arrêt total de la victimisation	Diminution de la victimisation	Pas de changement de la victimisation	Augmentation de la victimisation
Court-terme	29 %	55 %	16 %	0 %
Long-terme	26 %	34 %	8 %	32 %

Inspiré de van der Ploeg et al. (2016, p.11) Répartition en % des élèves ayant eu une action indiquée suivant l'issue de la situation.

À long terme, les élèves ayant eu une action indiquée présentent une augmentation du nombre de défenseur·euse·s ce qui n'est pas le cas des autres élèves victimes de harcèlement. Toutefois, le fait d'avoir eu ou non une action indiquée n'influence ni l'évolution du nombre de harceleur·euse·s ni celle du bien-être à l'école. De plus, il y a un effet des actions indiquées contraire aux hypothèses concernant la fréquence de la victimisation : la fréquence de victimisation a diminué chez les élèves n'ayant pas eu d'action indiquée alors qu'elle est restée stable chez ceux et celles en ayant eu.

2. Les effets sont-ils différenciés ?

Les auteurs ne rapportent pas d'analyse permettant de savoir si les effets sont différenciés en fonction des variables démographiques ou en fonction de la condition expérimentale (*KiVa* vs. *KiVa+*).

Il est supposé que le programme permette de réduire le harcèlement en incitant les témoins des situations de harcèlement à agir. Oldenburg et al. (2018) ont investigué l'effet modérateur des relations entre les élèves sur les comportements de défense. Il ressort de leur étude qu'il est plus probable qu'un·e élève victime de harcèlement se fasse défendre par un·e élève qu'il ou elle considère, ou qui le ou la considère, comme un·e ami·e. De même, la probabilité qu'un·e élève en défende un·e autre est plus importante lorsque, sans être eux ou elles-mêmes ami·e·s, ces deux élèves sont désigné·e·s comme des ami·e·s par les mêmes élèves. Il est peu probable qu'un·e élève se fasse défendre par quelqu'un qu'il ou elle n'aime pas, ou qui ne l'aime pas. Par contre, il est probable qu'un·e élève en défende un·e autre s'ils ou elles n'aiment pas les mêmes élèves. L'effet de *KiVa* est donc modéré par les relations qui existent entre les élèves d'une classe.

Kaufman et al. (2018) se sont intéressé·e·s aux caractéristiques des élèves qui restaient victimes de harcèlement malgré l'intervention. Selon leurs résultats, les élèves pour lequel·le·s l'implantation du programme ne permet pas une diminution ou un arrêt de la victimisation présentent, avant l'intervention, des niveaux particulièrement élevés de rejet par les pairs et d'anxiété sociale et une chaleur parentale plus

Annexe 1 Evaluation de *KiVa*

faible. Ce sont des caractéristiques qui semblent modérer l'effet de *KiVa* et qui sont à prendre en compte pour une amélioration du programme.

3. Des mesures de suivi ont-elles été prises ?

Des mesures de suivi ont été prises, que ce soit pour tester l'efficacité générale du programme ou l'effet des actions indiquées. Concernant l'efficacité générale du programme, la plupart des effets se maintiennent un an après le premier post-test (le programme était alors toujours implanté). Par contre, au niveau des actions indiquées, les résultats sont très différents une semaine après l'intervention et à la fin de l'année. L'effet de l'intervention, qui était positif après une semaine, devient négatif au suivi.

4. L'implantation a-t-elle été mesurée ?

L'implantation a été mesurée via le nombre de leçons *KiVa* dispensées. En mai 2013, seulement 36 % des enseignant·e·s ont dispensé toutes les dix leçons (y compris la leçon de clôture festive) et 58 % sont parvenus à terminer les neuf leçons thématiques. Il semble donc que la mise en œuvre du programme dans les temps soit difficile pour de nombreux·ses enseignant·e·s.

Les enseignant·e·s devaient uniquement rapporter le nombre de leçons données, mais, lors des rencontres d'expériences, ils et elles ont pu donner de nombreux feedbacks sur l'implantation du programme. Selon leurs dires, les leçons *KiVa* prennent plus de temps que prévu lorsqu'on laisse de la place aux interventions des élèves. L'interactivité des leçons est jugée très positivement par les enseignant·e·s (elle leur permet d'apprendre à mieux connaître leur classe), mais rend difficile l'intégration des leçons à l'horaire des cours. Beaucoup d'enseignant·e·s ont divisé les leçons en petits morceaux distribués sur plusieurs jours. Ils et elles apprécient avoir la possibilité de facilement glisser des exercices *KiVa* entre deux autres activités. Toujours selon les enseignant·e·s, les leçons *KiVa* créent une atmosphère de sécurité dans le groupe. Grâce à leur ton non pédagogique, elles sont appréciées par les élèves qui s'y impliquent et adoptent une attitude ouverte. Les élèves trouvent ce qu'ils et elles apprennent vraiment intéressant et intègrent le fait qu'ils et elles sont responsables de leur comportement. Les élèves apprécient le fait que, grâce à *KiVa*, il existe des règles de conduite claires dans l'école. Les enseignant·e·s remarquent que leurs élèves intègrent bien ce qu'ils et elles apprennent et recourent à leurs nouvelles connaissances en dehors des leçons *KiVa* (ex. ils et elles font remarquer quand un comportement n'est « *pas KiVa* »).

Il semble que les tailles d'effet soient plus importantes dans le cas des analyses *Treatment-on-Treated* (comparant les écoles contrôles aux écoles *KiVa* ayant implanté toutes les leçons) que dans celui des analyses *Intent-to-Treat* (comparant les écoles contrôles à toutes les écoles *KiVa*). Par exemple, dans les écoles *KiVa*, le score moyen de victimisation est de 1,76 et la diminution de la victimisation de 51 %. Toutefois, si toutes les écoles avaient implanté le programme de façon optimale, le score moyen de victimisation vaudrait 1,64 et la diminution de la victimisation passerait à 54 %. Mais aucune analyse n'a été effectuée pour confirmer que les effets sont plus importants quand davantage de leçons ont été dispensées.

Implantation de *KiVa* au Royaume-Uni

Le Royaume-Uni est, comme tout pays, confronté au harcèlement scolaire : 11 % des élèves rapportent se faire régulièrement harceler. C'est pourquoi les écoles sont légalement obligées d'adopter une politique anti-harcèlement. Toutefois, la loi donne uniquement des directives concernant ce qui est requis de faire. Le contenu du travail effectué par les écoles, tout comme sa qualité, est très variable. De nombreuses interventions visant à réduire le harcèlement sont utilisées, avec très peu d'évaluation systématique de leur qualité.

Suite à une conférence donnée par Christina Salmivalli (codirectrice de *KiVa*) à Cambridge, Judy Hutchings a présenté le programme au gouvernement gallois qui a accepté de subventionner son implantation dans certaines écoles à partir de l'année scolaire 2012-2013. L'évaluation de l'efficacité de cette implantation a été permise grâce à un financement européen. Suite aux résultats positifs de cette première évaluation, un fond a été obtenu pour réaliser une évaluation contrôlée de l'efficacité de *KiVa* au Pays de Galles. Le programme a été implanté dans 20 écoles lors de l'année scolaire 2013-2014. Les résultats de cette évaluation ne sont pas encore disponibles, seul le protocole de l'étude a été publié (Clarkson et al., 2016).

Qualité de l'évaluation de *KiVa* au Royaume-Uni

Article : Hutchings, J., & Clarkson, S. (2015). Introducing and piloting the *KiVa* bullying prevention programme in the UK. *Educational and Child Psychology*, 32(1), 49-61.

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une comparaison pré/post-test sans groupe contrôle. De plus, étant donné qu'il s'agit d'une étude de petite taille mise en place tardivement, les données pré et post-test des élèves n'ont pas pu être reliées. Les effets du programme ont donc été évalués via des tests-t comparant, au niveau des écoles, les proportions d'élèves victimes et auteur·e·s de harcèlement avant et après l'implantation de *KiVa*.

2. Quelles étaient les caractéristiques de l'échantillon ?

17 écoles primaires réparties à travers le Pays de Galles ont été recrutées. Suivant les écoles, le programme était dispensé en 5^{ème} et 6^{ème} année, uniquement en 5^{ème} ou uniquement en 6^{ème} (seule la traduction de l'unité 2 de *KiVa* était alors disponible en anglais). Quatre écoles ont été exclues de l'analyse parce que les données post-test de la majorité des élèves n'étaient pas disponibles. Les élèves ayant des données manquantes sur une des mesures ont été exclus de l'analyse portant sur la variable en question. L'échantillon final est composé de 13 écoles comptant 473 élèves au prétest (228 filles et 245 garçons) et 472 au post-test (225 filles et 247 garçons).

3. Par qui et où l'évaluation a-t-elle été menée ?

L'évaluation est menée au Pays de Galles par des chercheuses de la Bangor University (Judy Hutchings et Susan Clarkson) qui n'ont pas de lien avec l'équipe finlandaise.

Annexe 1 Evaluation de *KiVa*

4. Quelles et combien de mesures ont été prises ?

Les mesures prétest ont été prises au début de l'année scolaire 2012-2013 et les données post-test à la fin de celle-ci (les données n'ont donc pas été prises à la même période de l'année). Les élèves ont répondu à un questionnaire anonyme en ligne en étant encadré·e·s par leurs enseignant·e·s qui avaient reçu des consignes précises sur la façon de se comporter pendant la passation des questionnaires. Les items globaux du *Revised OBVQ* (Olweus, 1996; cité par Hutchings & Clarkson, 2015) ont été utilisés afin de mesurer les deux variables dépendantes : la victimisation et le harcèlement. La définition du harcèlement du *OBVQ* était présentée aux élèves avant les questions. De plus, une version raccourcie de la définition restait affichée sur l'écran. Les élèves ont été considéré·e·s comme victime et harceleur·euse sur base du point de coupure 2 à 3 fois par mois recommandé par Solberg et Olweus (cité par Hutchings & Clarkson, 2015).

Effets de *KiVa* au Royaume-Uni

1. Des effets ont-ils été observés ?

Lorsque l'on considère l'échantillon total, la victimisation comme le harcèlement diminuent significativement au cours de l'année. Pour les filles, il y a une réduction significative de la victimisation et du harcèlement tandis que pour les garçons, seul le harcèlement diminue. Les pourcentages d'élèves catégorisé·e·s en tant que victime et harceleur·euse dans chaque école au prétest et au post-test sont indiqués dans le tableau 9.

1. Les effets sont-ils différenciés ?

Les effets sont différenciés en termes de genre : la victimisation diminue uniquement chez les filles. Les auteures ne rapportent aucune autre analyse permettant de savoir si les effets sont différenciés.

2. Des mesures de suivi ont-elles été prises ?

Aucune mesure de suivi n'a été prise.

3. L'implantation a-t-elle été mesurée ?

Les enseignant·e·s ont répondu à un questionnaire en ligne sur leur expérience d'implantation du programme. 73,3 % d'entre eux et elles ont considéré que les leçons étaient faciles à donner. Ils et elles estimaient qu'entre 75 et 100 % des élèves étaient engagé·e·s et enthousiastes face aux leçons. Les enseignant·e·s ont évalué à quel point il leur semblait que le programme convenait pour réduire le harcèlement sur une échelle à 5 points allant de Ne convenant extrêmement pas à Convenant tout à fait. Le score moyen attribué à *KiVa* était de 4,47. Les enseignant·e·s ont également donné des retours très positifs concernant les discussions générées par les leçons et ont rapporté que tou·te·s les élèves appréciaient participer aux leçons *KiVa*.

Annexe 1 Evaluation de KiVa

Tableau 9. Pourcentage d'élèves catégorisé en tant que victime et harceleur-euse avant et après l'implantation de KiVa

	Filles				Garçons				Total			
	Victime		Harceleur-euse		Victime		Harceleur-euse		Victime		Harceleur-euse	
	Prétest	Post-test	Prétest	Post-test	Prétest	Post-test	Prétest	Post-test	Prétest	Post-test	Prétest	Post-test
École 1	25 %	4 %	0 %	3 %	16 %	3 %	9 %	3 %	20 %	3 %	5 %	3 %
Ecole 2	41 %	23 %	17 %	4 %	25 %	8 %	14 %	0 %	35 %	18 %	16 %	3 %
Ecole 3	11%	0%	0%	0%	5%	0%	0%	0%	8%	0%	0%	0%
Ecole 4	11%	0%	11%	0%	6%	12%	0%	0%	8%	8%	4%	0%
Ecole 5	8%	8%	23%	0%	0%	38%	8%	0%	4%	8%	16%	0%
Ecole 6	11%	0%	0%	0%	6%	0%	5%	5%	8%	0%	4%	4%
Ecole 7	22 %	0 %	22 %	0 %	27 %	20 %	0 %	7 %	25 %	13 %	9 %	4 %
Ecole 8	4 %	0 %	8 %	0 %	8 %	0 %	7 %	0 %	8 %	0 %	7 %	0 %
Ecole 9	14 %	0 %	8 %	0 %	33 %	7 %	12 %	12 %	22 %	3 %	10 %	5 %
École 1 1	50 %	14 %	27 %	0 %	8 %	0 %	9 %	0 %	31 %	8 %	19 %	0 %
École 1 2	18 %	7 %	22 %	0 %	17 %	6 %	14 %	0 %	17 %	6 %	18 %	0 %
École 1 3	13 %	0 %	6 %	0 %	0 %	0 %	14 %	0 %	6 %	0 %	10 %	0 %
École 1 5	11 %	11 %	14 %	0 %	12 %	12 %	15 %	4 %	11 %	11 %	14 %	2 %

Inspiré de (Hutchings & Clarkson, 2015, pp. 55-56)

Implantation de KiVa en Italie

Comme partout ailleurs, le harcèlement scolaire est un problème en Italie avec une prévalence située entre 10 et 25 % (Menesini & Nocentini, 2015 ; cité par Nocentini & Menesini, 2016). Pourtant, les seules interventions qui existent sont des actions non systématiques, généralement réalisées au niveau de la classe et rarement évaluées en termes d'efficacité. C'est face à ce besoin d'interventions plus systématiques et ayant fait leurs preuves que le programme KiVa a été adapté au contexte italien puis implanté en Toscane par l'University of Florence, avec l'aide du *Regional School Board Tuscany*.

Qualité de l'évaluation de KiVa en Italie

Article : Nocentini, A., & Menesini, E. (2016). KiVa Anti-Bullying Program in Italy: Evidence of Effectiveness in a Randomized Control Trial. *Prevention Science*, 17(8), 1012-1023. doi:10.1007/s11121-016-0690-z

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude longitudinale avec un design RCT. Les écoles assignées à la condition contrôle étaient placées sur une liste d'attente d'un an et devaient, en attendant, continuer à gérer le harcèlement comme à leur habitude.

L'efficacité du programme a été testée en prenant en compte la structure hiérarchique des données ainsi que la présence de données manquantes. Les tailles d'effet ont été calculées à la fois en *d* de Cohen et via les *odds ratios*.

2. Quelles étaient les caractéristiques de l'échantillon ?

Pour recruter les écoles, le *Regional School Board of Tuscany* a envoyé des lettres à 35 écoles de Toscane en mai 2013. Les écoles contactées comportaient à la fois de l'enseignement primaire et secondaire (avec un nombre suffisant de classes de 4^{ème} et 6^{ème} année) et étaient caractérisées par des performances académiques et un milieu socio-économique dans la moyenne. 13 écoles remplissant ces critères se sont portées volontaires. Ces écoles comptaient 97 classes, 126 enseignant-e-s et 2184 élèves concerné-e-s par le programme. Les élèves n'ayant pas reçu le consentement parental ont été exclu-e-s de l'analyse. De plus, les élèves ayant des données manquantes sur une des mesures ont été exclus de l'analyse portant sur la variable en question. Au temps 1, l'échantillon était composé de 2042 élèves dont 1039 élèves dans la condition expérimentale et 1003 élèves dans la condition contrôle. Au temps 2, il était composé de 1910 élèves dont 954 dans la condition expérimentale et 956 dans la condition contrôle.

3. Par qui et où l'évaluation a-t-elle été menée ?

L'évaluation a été menée en Italie par deux chercheuses de l'University of Florence (Annalaura Nocentini et Esilia Menesini) qui n'ont pas de lien avec l'équipe finlandaise.

4. Quelles et combien de mesures ont été prises ?

Les mesures pré-test ont été récoltées avant l'implantation du programme (en septembre-octobre 2013) et les mesures post-test ont été récoltées à la fin de l'année scolaire (en mai-juin 2014). Les élèves ont répondu à des questionnaires papier-crayon en étant encadré·e·s par des psychologues, chercheur·euse·s et étudiant·e·s. Chaque élève était associé·e à un code afin de pouvoir relier les données prétest et post-test.

Les mesures qui ont été prises sont les suivantes :

- **Victimisation** : La victimisation a été mesurée à l'aide de la *Florence Bullying and Victimization Scale* (Palladioni et al., 2016 ; cité par Nocentini & Menesini, 2016). Cette échelle comporte 14 items décrivant des comportements de victimisation (ex. « *I was threatened* »). Quatre items portent sur la victimisation physique, sept sur la victimisation verbale et trois sur la victimisation indirecte. Les élèves ont dû indiquer à quelle fréquence ils et elles subissaient ces types de comportements sur une échelle à 5 points allant de *Jamais* à *Plusieurs fois par semaine*. Les items ont été moyennés pour former un score global de victimisation. Les alphas se situent entre .82 et .86.
- **Harcèlement** : Le harcèlement a été mesuré via les mêmes items de la *Florence Bullying and Victimization Scale* Palladioni et al., 2016; cité par Nocentini & Menesini, 2016), mais formulés à l'actif. Les élèves ont également complété l'item global du *OBVQ*, afin de pouvoir calculer des OR comparables aux autres études portant sur *KiVa*. Les alphas se situent entre .82 et .86.
- **Attitudes pro-harcèlement et pro-victime** : Les attitudes pro-harcèlement et pro-victime ont été mesurées via une version révisée du *Questionnaire on Attitudes toward Bullying* (Menesini et al., 2003 ; cité par Nocentini & Menesini, 2016), une échelle composée de 6 items définissant des attitudes pro-harcèlement (ex. « *It's ok to call some kids nasty names* ») et de 6 items définissant des attitudes pro-victimes (ex. « *One should try to help the bullied victim* »). Les élèves devaient marquer leur accord à chaque item via une échelle à 5 points allant de *Désaccord complet* à *Accord complet*. Un score global a été calculé pour les attitudes pro-harcèlement et un autre pour les attitudes pro-victime. Les alphas se situent entre .62 et .71.
- **Empathie envers la victime** : L'empathie envers la victime a été mesurée via l'échelle à 7 items utilisée dans les études finlandaises évaluant *KiVa* (Pöyhönen, Kärnä et Salmivalli, 2008). Les items ont été moyennés pour former un score global. Les alphas sont égaux à .83 au temps 1 et .82 au temps 2.

Effets de *KiVa* en Italie

1. Des effets ont-ils été observés ?

En primaire (4^{ème} et 5^{ème} année) comme en secondaire (6^{ème} année), le programme a un effet sur la victimisation et le harcèlement. Alors que, dans les écoles contrôles, les scores de victimisation et harcèlement restent stables ou augmentent, dans les écoles *KiVa*, ils diminuent. De plus, les attitudes pro-victimes et l'empathie envers les victimes augmentent davantage dans les écoles primaires *KiVa* que dans les écoles primaires contrôles. De même, les attitudes pro-harcèlement diminuent davantage dans les écoles primaires *KiVa* que dans les écoles primaires contrôles. En secondaire, les attitudes pro-harcèlement diminuent dans les écoles *KiVa* alors qu'elles augmentent dans les écoles contrôles (par contre, il n'y a pas d'effet de l'intervention sur l'empathie envers la victime et les attitudes pro-victimes). La taille des effets varie entre .20 et .40 (voir tableau 10).

Tableau 10. Taille des effets (d de Cohen) en primaire et secondaire suite à l'implantation de *KiVa* en Italie.

Primaire	
Harcèlement	0,24
Victimisation	0,38
Empathie envers la victime	0,40
Attitudes pro-victimes	0,24
Attitudes pro-harcèlement	0,30
Secondaire	
Harcèlement	0,21
Victimisation	0,23
Empathie envers la victime	0.08 (n.s)
Attitudes pro-victimes	0.05 (n.s)
Attitudes pro-harcèlement	0,20

Inspiré de Nocentini & Menesini (2016, p. 1018). Les effets significatifs sont en gras.

Les chercheuses ont également calculé les OR afin de pouvoir comparer les résultats obtenus avec ceux d'autres études. Seul l'OR de la victimisation dans les écoles primaires est significatif (voir tableau 11).

Annexe 1 Evaluation de *KiVa*

Tableau 11. Évolution des taux de victimisation et harcèlement suite à l'implantation de *KiVa* en Italie

	Septembre 2013	Mai 2014	Changement	Différence en faveur des écoles <i>KiVa</i>	OR (IC à 95 %)
Victimisation					
Primaire					
Contrôle	21,9 %	19,1 %	-12,8 %	38,2 %	1,93 (1.744-2.125)
Intervention	22,2 %	10,9 %	-51 %		
Secondaire					
Contrôle	6,0 %	7,3 %	+22 %	34,9 %	1216 (0.938-1.576) (n.s)
Intervention	7,0 %	6,1 %	-12,9 %		
Harcèlement					
Primaire					
Contrôle	7,0 %	5,0 %	-28,6 %	32,4 %	1,31 (0.930-1.826) (n.s)
Intervention	9,0 %	4,4 %	-51 %		
Secondaire					
Contrôle	3,9 %	3,3 %	-15,4 %	36,5 %	1325 (0.927-1.892) (n.s)
Intervention	4,3 %	2,5 %	-41,9 %		

Inspiré de Nocentini & Menesini (2016, p. 1019). Les effets significatifs sont en gras.

2. Les effets sont-ils différenciés ?

Les effets sont moins importants en secondaire qu'en primaire. Les auteurs ne rapportent aucune autre analyse permettant de savoir si les effets sont différenciés.

3. Des mesures de suivi ont-elles été prises ?

Aucune mesure de suivi n'a été prise.

4. L'implantation a-t-elle été mesurée ?

Aucune mesure de l'implantation n'est rapportée dans l'étude.

Résumé des effets de *KiVa*

La Finlande, les Pays-Bas, le Royaume-Uni et l'Italie ne sont pas les seuls pays où des études visant à évaluer l'efficacité de *KiVa* sont menées. Aux États-Unis, des données ont été récoltées dans des écoles primaires implantant *KiVa* au cours des années scolaires 2012-2013 et 2013-2014. Swift et al. (2017) ont publié une étude portant sur les facteurs liés aux enseignant·e·s qui impactent l'implantation du programme. Mais, à ce jour, aucune étude portant sur l'efficacité de cette implantation n'a été publiée. Au Chili, *KiVa* est implanté et évalué depuis l'année scolaire 2016-2017, mais, pour l'instant, seul un protocole d'étude a été publié (Gaete et al., 2017). Des rapports sur l'efficacité du programme dans d'autres pays (en Nouvelle-Zélande par exemple) sont présents sur le site de *KiVa* mais uniquement accessibles aux personnes disposant une licence. De plus des informations sont parfois postées sur la page Facebook de *KiVa* (ex. graphique présentant l'évolution du pourcentage d'élèves harcelés dans les écoles estoniennes implantant *KiVa*), mais sans être accompagnées d'explications ou d'analyses plus élaborées.

Les effets observés sont-ils constants d'une étude à l'autre ?

D'une étude et d'un pays à l'autre, les effets du programme *KiVa* sont assez semblables. Le programme a des effets positifs sur le harcèlement et la victimisation dans tous les cas. Toutefois, ces effets ne concernent pas toujours tou·te·s les élèves. Par exemple, au Royaume-Uni, l'implantation de *KiVa* est associée à une réduction de la victimisation chez les filles, mais pas chez les garçons (Hutchings & Clarkson, 2015) et, en Finlande, le programme n'a pas d'effet sur les élèves les plus âgés (Kärnä et al., 2013; Kärnä et al., 2011a). Des effets ne sont pas non plus retrouvés via tous les instruments de mesure utilisés. Par exemple, Nocentini & Menesini (2016) retrouvent des effets sur le harcèlement en utilisant la *Florence Bullying and Victimisation Scale*, mais pas en utilisant l'item global du *OBVQ*. Le programme semble également efficace pour diminuer la cybervictimisation et le cyberharcèlement (du moins chez les élèves de moins de 13 ans), mais cette question n'est investiguée que dans une étude (Williford et al., 2013).

Dans certaines études, des variables secondaires ont également été testées. Au niveau des rôles adoptés par les élèves dans les situations de harcèlement, que ce soit en Finlande ou aux Pays-Bas (Kärnä et al., 2013; Kärnä et al., 2011b; Veenstra, 2015), une diminution des comportements d'assistance et de renforcement est observée, mais sans être accompagnée d'une augmentation des comportements de défense (ils diminuent même dans les écoles *KiVa* en Finlande). En Italie, des effets sont observés sur les attitudes anti-harcèlement et l'empathie envers les victimes, mais pas en Finlande (Kärnä et al., 2011b; Nocentini & Menesini, 2016). Les autres mesures prises ne sont pas les mêmes d'une étude à l'autre, il n'est donc pas possible de savoir si les effets observés sont constants.

Les effets sont presque toujours de petite taille. Les tailles d'effet sont reportées en termes de d de Cohen en Finlande, aux Pays-Bas et en Italie (Kärnä et al., 2011b; Nocentini & Menesini, 2016; Veenstra, 2015). Les effets significatifs varient entre un d de .08 et un d de .58. Les effets les plus importants sont observés pour des mesures telles que les perceptions qu'ont les élèves de leur enseignant·e ou l'empathie envers les victimes. Concernant la réduction du harcèlement et de la victimisation, les effets sont inférieurs à .20, sauf en Italie où tous les effets observés sont de plus grande taille (voir tableau 12).

Tableau 12. Taille des effets (d de Cohen) dans différents pays

		Taille des effets (d de Cohen)	
		Harcèlement	Victimisation
Finlande		.10	.17
Pays-Bas		.11	.13
Italie	Primaire	.24	.38
	Secondaire	.21	.23

Inspiré de Kärnä et al. (2011b, p. 323); Nocentini & Menesini (2016, p. 1018); Veenstra (2015, p. 10)

Pour les études réalisées en Finlande et en Italie, les OR sont également rapportés. L'analyse des OR confirme que les effets de KiVa sur le harcèlement et la victimisation sont de petite taille. Le tableau 13 reprend les OR observés en Finlande (Kärnä et al., 2013; Kärnä et al., 2011a; Kärnä et al., 2011b). En résumé, après l'implantation du programme, les probabilités d'harcéler ou de se faire harceler sont de 1,2 à 1,3 fois plus importante dans les écoles contrôles que dans les écoles KiVa.

Tableau 13. Taille des effets de KiVa en Finlande

	OR (IC à 95 %)			
	Victimisation auto-rapportée		Harcèlement auto-rapportée	
	Évaluation RCT	Dissémination nationale	Évaluation RCT	Dissémination nationale
1ère année	1,33	1,17	1,50	1,15
2ème année		1,23		1,30
3ème année		1,23		1,28
4ème année	1,53	1,33	1,41	1,34
5ème année		1,30		1,23
6ème année		1,22		1,19
7ème année	1,13	1,11	1,21	1,05
8ème année		1,15		1,05
9ème année		1,13		1,07
Total	1,28	1,21	1,30	1,18

Inspiré de Kärnä et al. (2013, p. 546); Kärnä et al. (2011a, p. 802) et Kärnä et al. (2011b, p. 323).

Concernant l'effet spécifique des actions indiquées, il est très positif lorsqu'il est mesuré à court terme et de façon directe par l'équipe *KiVa* mais l'est beaucoup moins lorsqu'il est mesuré à long terme et de façon confidentielle (Garandeau, Poskiparta & Salmivalli, 2014; Veenstra, 2015).

Le programme semble-t-il plus efficace pour une certaine population ?

Des différences en fonction du genre sont parfois observées, mais il n'en ressort aucun pattern de résultats clair. Il semble donc que le programme soit efficace pour les filles comme pour les garçons. Par contre, au niveau de l'âge, il semble clair que l'efficacité du programme s'estompe après la 4^{ème} année jusqu'à être quasiment nul après la 6^{ème} année (donc chez les élèves de 12-13 ans en Finlande et 11-12 ans dans les autres pays). Il ne semble pas que ce soit le passage en secondaire qui soit lié à cette inefficacité du programme vu qu'en Italie (Nocentini & Menesini, 2016), où la transition se fait entre la 5^{ème} et la 6^{ème} plutôt qu'entre la 6^{ème} et la 7^{ème}, des effets sont toujours observés chez les élèves de 6^{ème} année (quoique de façon moins importante que chez les 4^{ème} et 5^{ème} année). Mis à part l'âge et le genre, en Finlande les effets sont également modérés par la composition des classes (avec des effets plus importants lorsqu'il y a une grande proportion de garçons). Il semble également que *KiVa* soit moins efficace pour diminuer le harcèlement des élèves populaires et la victimisation des élèves rejeté·e·s par leurs pairs, socialement anxieux·euses et ayant peu de chaleur dans la relation avec leurs parents (Garandeau, Lee, et al., 2014; Kaufman et al., 2018). Toutefois, ces modérateurs n'ont pas été testés dans les autres études.

En conclusion, l'efficacité de *KiVa* est soutenue par de nombreuses données pour les élèves jusqu'à la 6^{ème} année. Toutefois, les effets du programme sont de petite taille et les actions indiquées ne semblent pas efficaces pour stopper les cas de harcèlement avérés.

Annexe 2 : Evaluation de *Friendly Schools*

Les évaluations de *Friendly Schools* publiées jusqu'à présent ont toujours été effectuées en Australie. Selon Rolls (2007), le programme a également été évalué à Harlem (New York). Les données devraient avoir été collectées en mai 2007, mais il ne semble pas qu'elles aient été publiées.

Une évaluation de la première version de *Friendly Schools* a été effectuée en 2000-2002. Suite à cela, les composantes familiales du programme ont été étoffées. *Friendly Schools Friendly Families* a été évalué en 2002-2004. La partie du programme concernant le cyberharcèlement, *Cyber Friendly Schools* a été évaluée de 2010 à 2012. Aucune évaluation de la dernière version du programme, *Friendly Schools Plus*, n'a été publiée.

Implantation de *Friendly Schools*

Qualité de l'évaluation de *Friendly Schools* (2000-2002)

Article principal : Cross, D., Monks, H., Hall, M., Shaw, T., Pintabona, Y., Erceg, E., ... Lester, L. (2011). Three-year results of the friendly schools whole-of-school intervention on children's bullying behaviour. *British Educational Research Journal*, 17(1), 105-129. doi:10.1080/01411920903420024

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude longitudinale avec un design RCT. L'étude a débuté en avril 2000 et a duré trois années scolaires, de 2000 à 2002 (en Australie, l'année scolaire s'étend de février à novembre). Les écoles de la condition expérimentale devaient remplacer leur curriculum standard d'éducation à la santé par *Friendly Schools*. Un soutien actif leur était apporté lors des deux premières années et un soutien passif lors de la troisième. Les écoles de la condition contrôle étaient encouragées à enseigner le curriculum standard. Pour éviter un effet Hawthorne, elles ont reçu tout le matériel et les formations nécessaires à l'implantation d'un autre programme, visant à promouvoir la sécurité routière.

Lors de chaque post-test, des modèles ANCOVA prenant en compte les données pré-test, le genre et la taille de l'école ont été utilisés pour tester l'effet du groupe. Il n'est expliqué nulle part comment les données manquantes ont été gérées.

2. Quelles étaient les caractéristiques de l'échantillon ?

Sur base des résultats d'une analyse de puissance, les auteur·e·s ont décidé de sélectionner aléatoirement 29 écoles parmi les 305 écoles primaires gouvernementales de la ville de Perth. Toutes les écoles sélectionnées ont accepté de participer. Avant d'être aléatoirement assignées à la condition expérimentale (15 écoles) ou contrôle (14 écoles), les écoles ont été classées en fonction de leur taille et de leur statut socio-économique.

Annexe 2 – Evaluation de *Friendly Schools*

Les élèves ont été suivi·e·s de la 4^{ème} à la 6^{ème} primaire. 95,2 % des élèves de 4^{ème} année, soit 1968 élèves, ont rempli le questionnaire au pré-test (les autres élèves étaient absent·e·s le jour de la récolte des données ou n'avaient pas reçu le consentement parental). Environ deux tiers des élèves ayant répondu au pré-test ont participé à tous les temps de mesure. Les élèves n'ayant pas répondu au post-test 3 avaient, au pré-test, une probabilité plus élevée de se faire régulièrement harceler et moins élevée de rapporter à quelqu'un le harcèlement vécu. Les élèves du groupe contrôle et du groupe expérimental ayant quitté l'étude ne se distinguaient sur aucune des variables d'intérêt.

Au pré-test, les deux conditions ne se distinguaient pas au niveau des scores de harcèlement et de victimisation ou de la probabilité d'avoir parlé du harcèlement, mais davantage d'élèves de la condition expérimentale avaient vu un·e autre élève se faire harceler.

3. Par qui et où l'évaluation a-t-elle été menée ?

L'étude a été menée à Perth, dans le nord-ouest de l'Australie, par une équipe de chercheur·euse·s du *Child Health Promotion Research Center* de l'Edith Cowan University (Donna Cross, Helen Monks, Marg Hall, Thérèse Shaw, Yolanda Pintabona, Erin Erceg, Greg Hamilton, Stacey Waters et Leanne Lester) ainsi que par Clare Roberts de la Curtin University of Technology. L'évaluation est donc réalisée par l'équipe ayant développé le programme.

4. Quelles et combien de mesures ont été prises ?

Les données pré-test ont été récoltées en avril 2000 tandis que les données post-test ont été récoltées en novembre 2000, 2001 et 2002. Le pré-test a donc eu lieu à une période de l'année différente des post-tests. Les élèves ont répondu aux questionnaires pendant les heures scolaires en étant encadré·e·s par des assistant·e·s de recherche formé·e·s (et, pour la plupart, aveugles aux conditions de l'étude). Tous les items étaient lus à voix haute. Les élèves étaient identifié·e·s par des numéros afin de pouvoir relier les données pré et post-test tout en assurant leur confidentialité. Des sessions de rattrapage étaient organisées par les enseignant·e·s pour les élèves absent·e·s le jour de la récolte des données.

Le questionnaire débutait par une définition du harcèlement, adaptée de Olweus (1994; cité par Cross et al., 2011) : « *Bullying is when a person is ignored, left out on purpose, or not allowed to join in; hit, kicked or pushed around; has lies or nasty stories told about them to make other kids not like them; is made afraid of getting hurt and/or made fun of and teased in a mean and hurtful way. When teasing is done in a friendly and playful way it is not called bullying, and while fighting is not a good thing to do, it is not bullying when two students who are as strong as each other get into a fight. If two people who are normally friends have an argument, this isn't bullying.* » (Cross et al., 2011, p; 112). Comme la définition avait été jugée difficile à comprendre lors d'une étude pilote, elle était accompagnée de pictogrammes. Tous les items du questionnaire sont basés sur des items développés dans des études précédentes (Olweus, 1996; Rigby, 1998a; cités par Cross et al., 2011) et ont été testés dans une étude pilote auprès d'élèves similaires afin de s'assurer de leur fidélité et validité faciale.

Annexe 2 – Evaluation de *Friendly Schools*

Les mesures qui ont été prises sont les suivantes :

- **Victimisation** : Afin de mesurer la victimisation, les élèves ont dû indiquer à quelle fréquence ils et elles s'étaient fait harceler au cours du dernier trimestre parmi les possibilités de réponse suivantes : *Presque chaque jour*, *La plupart des jours*, *Une fois par semaine*, *Toutes les quelques semaines*, *Une ou deux fois* et *Jamais*. Les odd ratios ont été calculés en dichotomisant les élèves de deux manières différentes afin de tester les effets du programme sur la présence du harcèlement peu importe sa fréquence (les élèves ayant répondu *Jamais* sont opposé-e-s aux autres) et sur la présence du harcèlement fréquent (les élèves ayant répondu *Toutes les quelques semaines* ou moins sont opposé-e-s aux autres).
- **Harcèlement** : Afin de mesurer le harcèlement, les élèves ont dû indiquer à quelle fréquence ils et elles avaient harcelé d'autres élèves au cours du dernier trimestre. Les possibilités de réponse, tout comme les points de coupure, étaient les mêmes que pour la victimisation.
- **Rapport de la victimisation vécue** : Afin de mesurer le rapport de la victimisation vécue, il était demandé aux élèves s'ils et elles avaient parlé à quelqu'un du harcèlement éventuellement subi. Pour le calcul des odd ratios, les élèves ayant parlé de leur victimisation et ceux et celles n'étant pas victimes de harcèlement ont été opposé-e-s aux élèves n'ayant pas parlé de la victimisation qu'ils et elles vivaient.
- **Victimisation observée** : Afin d'évaluer la victimisation observée, les élèves ont dû indiquer s'ils et elles avaient vu des élèves de leur âge ou plus jeunes se faire harceler au cours du dernier trimestre. Il y avait plusieurs possibilités de réponses, mais les élèves ont simplement été dichotomisé-e-s suivant le fait qu'ils et elles aient ou non assisté à des scènes de harcèlement.
- **Rapport de la victimisation observée** : Afin de mesurer le rapport de la victimisation observée, il était demandé aux élèves s'ils et elles parleraient à quelqu'un du fait d'avoir vu un autre élève se faire harceler. L'odd ratio a été calculé en opposant la réponse *J'en parlerai à quelqu'un* aux autres réponses.
- **Attitudes anti-harcèlement** : Les attitudes anti-harcèlement ont été évaluées via 9 items basés sur la *Pro-victim Scale* (Rigby & Slee ; cité par Cross et al., 2004a).
- **Solitude** : La solitude a été évaluée via 7 items adaptés du *Loneliness and Social Dissatisfaction Questionnaire* (Asher & Wheeler, 1985 ; cité par Cross et al., 2004a).
- **Soutien des pairs** : La perception de soutien des pairs a été évaluée via 17 items adaptés de la *Friendship Features Interview for Young Children* (Ladd, Kochenderfer & Coleman, 1996 ; cité par Cross et al., 2004a).
- **Compétences sociales** : Les compétences sociales ont été évaluées via les 10 items du *Students' Social Competence with Peers Pupil Questionnaire* (Spence, 1995 ; cité par Cross et al., 2004a).

Annexe 2 – Evaluation de *Friendly Schools*

- **Nombre d'ami·e·s :** Le nombre d'ami·e·s a été évalué via 4 items adaptés du *Peer Relation Questionnaire* (Rigby & Slee, 1998; cité par Cross et al., 2004a)
- **Appréciation de l'école :** L'appréciation de l'école a été évaluée via 4 items développés spécifiquement pour l'étude.
- **Conséquences attendues en cas de harcèlement :** Les conséquences attendues en cas de harcèlement ont été évaluées via des items basés sur un questionnaire développé par Rigby (1997; cité par Cross et al., 2004) visant à mesurer les perceptions des élèves de ce qui pourrait arriver s'ils ou elles harcelaient un autre élève. L'analyse factorielle a révélé deux facteurs, l'un relevant des conséquences positives et l'autre des conséquences négatives du harcèlement.

Effets de *Friendly Schools*

1. Des effets ont-ils été observés ?

Comme prévu, le programme a un effet sur la victimisation auto-rapportée. Toutefois, cet effet est uniquement présent au post-test 1 pour la victimisation totale et au post-test 3 pour la victimisation fréquente (voir tableau 1). Dans ces deux cas, la probabilité d'être victime de harcèlement est plus importante dans les écoles de la condition contrôle. Il y a également un effet significatif concernant le fait de rapporter la victimisation vécue : les élèves de la condition expérimentale ont une probabilité plus importante de rapporter le harcèlement qu'ils et elles subissent par rapport aux élèves de la condition contrôle. Cet effet est uniquement présent au premier post-test ce qui, d'après les auteur·e·s, pourrait s'expliquer par le fait que, par la suite, les élèves se sentent davantage capables de gérer le harcèlement d'eux ou d'elles-mêmes.

Il n'y a aucun effet du programme sur le harcèlement auto-rapporté, que ce soit le harcèlement total ou fréquent. Par contre, aux trois post-test, les élèves du groupe contrôle ont une probabilité plus importante que ceux et celles du groupe expérimental d'avoir vu d'autres élèves de leur âge ou plus jeunes se faire harceler. De plus, au post-test 2 (c'est-à-dire le seul pour lequel ces données ont été rapportées), les élèves de la condition expérimentale sont plus enclin·e·s à rapporter avoir une situation de harcèlement observée que ceux et celles de la condition expérimentale.

Au niveau des variables secondaires, les résultats sont uniquement rapportés par Cross et al. (2004a) pour le post-test 2. L'intervention n'a aucun effet significatif sur les attitudes anti-harcèlement, la solitude, les compétences sociales, le nombre d'ami·e·s ou les conséquences positives attendues en cas de harcèlement. Toutefois, les élèves de la condition expérimentale ont une probabilité plus importante de rapporter aimer l'école et de s'attendre à des conséquences négatives en cas de harcèlement. Il y a également un effet de l'intervention sur le soutien perçu de la part des pairs, mais celui-ci va dans le sens inverse des hypothèses : les élèves ont une probabilité plus importante de percevoir un soutien élevé dans la condition contrôle que dans la condition expérimentale.

Annexe 2 – Evaluation de *Friendly Schools*

Tableau 1. Taille d'effet (OR) après un, deux et trois ans d'implantation de *Friendly Schools*

Victimisation auto-rapportée	Post-test 1 (novembre 2000)	1.49 (1,14-1,94)
	Post-test 2 (novembre 2001)	1.25 (0,95-1,64)
	Post-test 3 (novembre 2002)	1.26 (0,93-1,71)
Victimisation fréquente auto-rapportée	Post-test 1 (novembre 2000)	1.16 (0,87-1,54)
	Post-test 2 (novembre 2001)	1.00 (0,70-1,43)
	Post-test 3 (novembre 2002)	1.50 (1,10-2,05)
Harcèlement auto-rapporté	Post-test 1 (novembre 2000)	0.81 (0,54-1,20)
	Post-test 2 (novembre 2001)	1.15 (0,81-1,63)
	Post-test 3 (novembre 2002)	1.02 (0,75-1,40)
Harcèlement fréquent auto-rapporté	Post-test 1 (novembre 2000)	0.66 (0,39-1,13)
	Post-test 2 (novembre 2001)	1.14 (0,71-1,83)
	Post-test 3 (novembre 2002)	0.87 (0,48-1,59)
Rapport de la victimisation vécue	Post-test 1 (novembre 2000)	1.60 (1,12-2,13)
	Post-test 2 (novembre 2001)	1.28 (0,95-1,74)
	Post-test 3 (novembre 2002)	1.29 (0,94-1,77)
Victimisation observée	Post-test 1 (novembre 2000)	1.36 (1,03-1,81)
	Post-test 2 (novembre 2001)	1.48 (1,14-1,92)
	Post-test 3 (novembre 2002)	1.67 (1,25-2,24)
Rapport de la victimisation observée	Post-test 2 (novembre 2001)	3.5 (1.5-8.2)

Inspiré de Cross et al. (2011, pp. 119-120) et de Cross et al. (2004a, p. 203). Les effets significatifs sont indiqués en gras.

2. Les effets sont-ils différenciés ?

Alors qu'un effet est observé au post-test 1 et au post-test 3, aucun effet n'est observé sur la victimisation au post-test 2, lorsque les élèves sont en 5^{ème} année. Cependant, il n'est pas possible de savoir si l'absence d'effet est due au niveau scolaire des élèves, celui-ci étant confondu avec la période d'implantation du programme.

3. Des mesures de suivi ont-elles été prises ?

Le programme a été implanté sur trois ans. Les mesures ont été prises à plusieurs reprises avec des post-test ayant lieu 7, 19 et 31 mois après le pré-test. Les effets sont uniquement constants pour la victimisation observée. Pour la victimisation vécue, un effet est observé au niveau de la victimisation totale au post-test 1, mais n'apparaît plus par la suite et un effet au niveau de la victimisation fréquente apparaît uniquement au dernier post-test. Les résultats concernant certaines mesures sont uniquement rapportés pour le post-test 2, il n'est donc pas possible de savoir si les effets trouvés étaient déjà présents auparavant et s'ils se maintiennent par la suite.

4. L'implantation a-t-elle été mesurée ?

L'implantation du programme a été mesurée via des journaux de bord que les enseignant·e·s devaient compléter. En 2000 (enseignant·e·s de 4^{ème} année), la médiane du nombre de leçons que les enseignant·e·s ont rapporté avoir dispensées est de 8 sur 9 (pour une médiane de 9,5h). En 2001 (enseignant·e·s de 5^{ème} année), elle est de 6 leçons (pour une médiane de 8 h, ce qui donne donc 17,5h sur les deux années). La médiane du nombre d'activités que les parents rapportent avoir complétées est de deux sur neuf en 2000 et d'une sur neuf en 2001.

La plupart des écoles ont mis au moins un an pour établir leur comité anti-harcèlement, consulter la communauté scolaire et revoir leur politique anti-harcèlement, ce qui était plus lent qu'anticipé. Selon les données récoltées auprès de membres des comités pilotant *Friendly Schools* dans leur école, dans la majorité des écoles expérimentales moins de 30 % du programme a été complété au niveau de l'école. Au niveau des classes 67 % du programme a été implanté en moyenne. Selon les comités, la faible implantation du programme au niveau de l'école serait due à un manque de capacité, les écoles manqueraient des compétences, structures, ressources et de l'engagement nécessaire à une implantation optimale. Le plus dur serait d'impliquer les parents, à cause de certaines barrières telles qu'un manque de ressource (spécialement pour les parents dont l'anglais n'est pas la langue maternelle), un manque de temps pour construire une relation avec les parents, un manque d'appréciation des parents comme des enseignant·e·s des bénéficiaires à tirer d'une collaboration et une perception de non-intérêt ou manque de motivation des parents. Beaucoup d'enseignant·e·s ont rapporté que le programme était excellent, mais qu'il prenait plus de temps que prévu. La plupart d'entre eux et elles ont indiqué avoir pris près du double du temps indiqué pour mener les activités (ce qui serait dû à leur nature interactive et au grand intérêt des élèves).

Implantation de *Friendly Schools Friendly Families*

Qualité de l'évaluation de *Friendly Schools Friendly Families* (2002-2004)

Article principal : Cross, D., Waters, S., Pearce, N., Shaw, T., Hall, M., Erceg, E., ... Hamilton, G. (2012). The Friendly Schools Friendly Families programme: Three-year bullying behaviour outcomes in primary school children. *International Journal of Educational Research*, 53, 394-406. doi:10.1016/j.ijer.2012.05.004

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude longitudinale avec un design RCT. Les écoles participantes ont été aléatoirement réparties parmi une des trois conditions expérimentales : intensité élevée, intensité modérée et intensité faible. Les conditions *intensité élevée* et *intensité modérée* sont fort semblables. La seule différence est que, dans la condition *intensité élevée*, le programme était renforcé au niveau familial avec une formation additionnelle de 3 h destinée au comité de l'école et visant à augmenter leur capacité à impliquer les parents et une série d'activités visant à conscientiser et augmenter les compétences des parents et familles (workshop de 2 h, feuilles de communication parents-enfants et activités classe-famille). Aucune formation n'était dispensée aux écoles de la condition *intensité faible*, mais celles-ci pouvaient, comme toutes les écoles de la ville, avoir accès au matériel *Friendly Schools*.

L'évaluation porte sur l'implantation du programme lors des années scolaires 2002, 2003 et 2004. Lors des deux premières années, le programme était activement implanté dans les conditions *intensité élevée* et *intensité modérée* tandis que l'année 2004 est davantage une année de suivi. De plus, à partir de 2004, le programme a été introduit avec une intensité importante dans la condition *intensité faible*.

A chaque post-test des modèles de régression logique ont été utilisés. Ces modèles tiennent compte de la présence de données nichées et contrôlent l'effet des données pré-test, du genre des élèves et de la taille des écoles. Toutes les données disponibles ont été incluses dans les analyses et aucune mesure particulière n'a été prise pour tenir compte des données manquantes.

Au niveau des données portant sur les parents, Cross et al. (2016) ont utilisé des modèles multiniveaux pour déterminer les effets de l'intervention. Les modèles contrôlent l'effet des perceptions des parents de leur engagement dans la vie de l'école, de leur implication dans le développement de la politique scolaire, et de l'efficacité des stratégies de l'école pour contrecarrer et gérer les comportements de harcèlement. Dans l'étude de Lester et al. (2017), des modèles à mesures répétées ont été utilisés pour déterminer les effets de l'intervention et des chi-carré pour déterminer les différences entre les mères et pères.

2. Quelles étaient les caractéristiques de l'échantillon ?

Sur base des résultats d'une analyse de puissance, les auteur·e·s ont décidé de sélectionner 20 écoles de la ville de Perth pour participer à l'étude. Les écoles ont été sélectionnées aléatoirement, mais seulement après avoir été classées en fonction de leur taille et de leur statut socio-économique. Davantage d'écoles au statut socio-économique moyen ont été sélectionnées afin d'avoir un échantillon représentatif de la ville

Annexe 2 – Evaluation de *Friendly Schools*

de Perth. Les écoles ont toutes accepté de participer et d’être aléatoirement réparties parmi les conditions *intensité élevée* (7 écoles), *intensité modérée* (6 écoles) et *intensité faible* (7 écoles).

Au pré-test, les élèves participant à l’étude étaient en 2^{ème}, 4^{ème} et 6^{ème} année. Cependant les données des élèves de 2^{ème} année, qui étaient des données hétéro et non auto-rapportées, n’ont pas été prises en compte dans l’article de Cross et al. (2012). Il y avait au pré-test 2552 élèves de 4^{ème} et 6^{ème} année (879 dans la condition *intensité élevée*, 712 dans la condition *intensité modérée* et 961 dans la condition *intensité faible*). Les élèves de 4^{ème} année ont été suivi·e·s sur trois années scolaires (de la 4^{ème} à la 7^{ème}) et ceux et celles de 6^{ème} année sur deux années scolaires (en 6^{ème} et 7^{ème}, ils et elles sont ensuite passé·e·s en secondaire). Au post-test 3, il n’y avait plus que 741 élèves, soit 57,2 % des élèves de 4^{ème} année. Le taux d’abandon est donc important. Il n’y a pas de différence entre les élèves ayant quitté l’étude et les autres mis à part au niveau du harcèlement auto-rapporté : la probabilité de rapporter avoir harcelé d’autres élèves est moins élevée parmi les élèves ayant quitté l’étude.

Les trois groupes expérimentaux ne se distinguaient pas en termes de genre des élèves ou de tailles des écoles. Cependant, davantage de parents d’élèves du groupe *intensité modérée* possédaient un diplôme technique ou de qualification et un nombre moins important de parents du groupe *intensité faible* possédaient un diplôme universitaire. Il n’y avait pas de différence sur les variables d’intérêt au pré-test excepté pour le harcèlement en 6^{ème} année : la probabilité d’avoir harcelé d’autres élèves était moins élevée dans le groupe *intensité faible* que dans les deux autres groupes.

L’étude de Cross et al. (2016a) porte sur les parents des élèves. L’échantillon est composé des parents des élèves de 2^{ème}, 4^{ème} et 6^{ème} année (un parent par élève, celui avec lequel l’élève parle le plus du harcèlement et des questions sociales). 83 % des participant·e·s étaient des mères, 14 % des pères et les 3 % des beaux-parents ou tuteur·trice·s. 3211 parents ont répondu au pré-test, dont 1444 (45 %) ont participé aux trois temps de mesure (pré-test, post-test 1 et post-test 2). Aucune différence n’a été trouvée au niveau des variables démographiques entre les parents ayant répondu à tous les temps de mesure et les autres. Par contre, davantage de parents ont été perdus au suivi dans les conditions *intensité faible* et *intensité élevée* que dans la condition *intensité modérée*. Il n’y avait pas de différence sur les variables d’intérêt au pré-test excepté pour la perception des priorités de l’école chez les parents d’élèves de 2^{ème} année : ceux de la condition *intensité faible* perçoivent davantage que la lutte contre le harcèlement est une priorité pour l’école de leur enfant que ceux des deux autres conditions. L’étude de Lester et al. (2017) porte également sur les parents des élèves, mais, cette fois, les deux parents étaient invités à participer. L’échantillon est composé de 1429 parents ayant complété le questionnaire au pré-test. Au post-test 2, il n’y avait plus que 60 % des parents de 2^{ème}, 65 % de 4^{ème} et 52 % de 6^{ème} qui ont participé.

3. Par qui et où l’évaluation a-t-elle été menée ?

L’étude a été menée dans la même ville que l’étude précédente (Perth), par une équipe essentiellement composée des mêmes chercheur·euse·s : Donna Cross, Stacey Waters, Natasha Pearce, Thérèse Shaw, Marg Hall, Erin Erceg et Greg Hamilton (l’Edith Cowan University), Sharyn Burns et Clare Roberts (Curtin University of Technology).

4. Quelles et combien de mesures ont été prises ?

Les données ont été récoltées en mars 2002 (pré-test), novembre 2002 (post-test 1), octobre 2003 (post-test 2) et octobre 2004 (post-test 3). Le pré-test n'a donc pas été effectué au même moment de l'année que les post-tests. Les élèves ont répondu au questionnaire pendant les heures scolaires, en étant encadré·e·s par des assistant·e·s de recherche formé·e·s lisant tous les items à voix haute. Le questionnaire débutait par une définition du harcèlement accompagnée de pictogrammes. La fidélité test/re-test des items avait été testée en faisant passer le questionnaire deux fois à un intervalle de deux semaines à des élèves de 4^{ème} année venant de trois autres écoles. La concordance des résultats est moyenne (kappas entre .41 et .61).

Les mesures qui ont été prises sont les suivantes :

- **Victimisation** : La victimisation a été évaluée via un item adapté du *OBVQ* (Olweus, 1996; cité par Cross et al., 2012) et du *Australian Peer Relations Assesment Questionnaire* (Rigby & Slee, 1991 ; cité par Cross et al., 2012) et portant sur la fréquence de victimisation subie au cours du dernier trimestre. Les possibilités de réponses étaient *Presque chaque jour, La plupart des jours, Une fois par semaine, Toutes les quelques semaines, Une ou deux fois* et *Jamais*. Les odd ratios ont été calculés en dichotomisant les élèves de deux manières différentes afin de tester les effets du programme sur la présence du harcèlement peu importe sa fréquence (les élèves ayant répondu *Jamais* sont opposé·e·s aux autres) et sur la présence du harcèlement fréquent (les élèves ayant répondu *Toutes les quelques semaines* ou moins sont opposé·e·s aux autres).
- **Harcèlement** : Le harcèlement a été évalué avec le même item que pour la victimisation, mais formulé de manière active. Les possibilités de réponse et points de coupure étaient également les mêmes.
- **Rapport de la victimisation** : Le rapport de la victimisation a été mesuré en demandant aux élèves s'ils et elles avaient parlé à quelqu'un du harcèlement éventuellement subi. Afin de calculer les odd ratios, les élèves ayant parlé de leur victimisation et ceux et celles n'étant pas victimes de harcèlement ont été opposé·e·s aux élèves n'ayant pas parlé de la victimisation qu'ils et elles vivaient.

Les parents d'élèves ont également répondu à des questionnaires. Les données ont été récoltées au pré-test ainsi que 10 et 22 mois plus tard. Deux questionnaires ont été développés. Le questionnaire le plus long était destiné au parent avec lequel l'enfant parlait le plus du harcèlement et des questions sociales en général tandis qu'une version plus courte était donnée à l'autre parent (les parents répondaient à la même version du questionnaire aux trois temps de mesure). L'étude de Cross, et al. (2016a) porte uniquement sur la version longue des questionnaires tandis que celle de Lester et al. (2017) porte sur les deux versions. Une série de stratégies ont été utilisées pour maximiser la participation des parents (personnalisation des lettres par leur enfant, petite récompense donnée à l'enfant lorsqu'il ou elle ramène l'enveloppe scellée...). Afin de vérifier leur validité faciale et de contenu, tous les items ont été revus par un panel d'expert·e·s composé d'éducateur·trice·s, de chercheur·euse·s spécialisé·e·s dans la prévention du harcèlement, de psychologue et de spécialistes de l'apprentissage des adultes.

Annexe 2 – Evaluation de *Friendly Schools*

Les mesures qui ont été prises dans la version courte du questionnaire sont les suivantes :

- **Fréquence de la communication parent-enfant sur le harcèlement :** La fréquence de la communication parent-enfant sur le harcèlement a été évaluée en demandant aux parents combien de fois ils avaient discuté des sujets suivants avec leur enfant : ce qu'est le harcèlement, comment réagir si l'on se fait harceler, pourquoi il ne faut pas harceler, comment réagir si l'on voit d'autres personnes se faire harceler, à qui parler si l'on se fait harceler et comment se faire des ami-e-s. Les possibilités de réponses étaient *Je ne m'en souviens plus, Jamais, Une ou deux fois, Trois à cinq fois* et *Plus de cinq fois*. Les réponses ont été moyennées pour former un score global (l'alpha est égal à .93).
- **Conseils donnés à leur enfant :** Afin de savoir quels conseils les parents donnaient à leur enfant pour réagir en cas de harcèlement, il leur était demandé de choisir une des réponses suivantes : *Aucun, ce n'est pas nécessaire, Aucun, je ne sais pas quoi dire, Ignorer le harcèlement, S'en aller, Riposter, L'accepter, Dire aux élèves qui harcèlent d'arrêter, Répondre avec humour, Demander l'aide d'autres élèves, En parler à un-e enseignant-e* et *En parler à quelqu'un à la maison*.
- **Influence sur la réponse de leur enfant face au harcèlement :** Afin de mesurer la perception des parents de l'influence qu'ils avaient sur la façon dont leur enfant répond face au harcèlement, il leur était demandé d'évaluer cette influence sur une échelle à quatre points allant de *Aucune influence* à *Beaucoup d'influence*.
- **Influence sur la décision de leur enfant de harceler :** Afin de mesurer la perception des parents de l'influence qu'ils avaient sur la décision de leur enfant de harceler ou non, il leur était demandé d'évaluer cette influence sur une échelle à quatre points allant de *Aucune influence* à *Beaucoup d'influence*.

Les mesures supplémentaires qui ont été prises dans la version longue du questionnaire sont les suivantes :

- **Connaissances liées au harcèlement :** Les connaissances des parents liées au harcèlement ont été évaluées via 9 items adaptés de Boulton et Underwood (1992 ; cité par Cross et al., 2016a). Un score de connaissance a été calculé en faisant la somme des réponses correctes (le score varie donc entre 0 et 9).
- **Efficacité personnelle à parler du harcèlement :** L'efficacité personnelle des parents pour discuter du harcèlement avec leur enfant a été évaluée en leur demandant s'ils avaient le sentiment d'avoir assez de connaissances pour discuter de ce qu'est le harcèlement, comment réagir si l'on se fait harceler, pourquoi il ne faut pas harceler, comment réagir si l'on voit d'autres personnes se faire harceler, à qui parler si l'on se fait harceler et comment se faire des ami-e-s. Une analyse factorielle a confirmé que les 6 items mesuraient une seule dimension. Un score d'efficacité personnelle a donc été calculé en comptant le nombre d'items pour lesquels les parents indiquent avoir suffisamment de connaissance (score entre 0 et 6).

Annexe 2 – Evaluation de *Friendly Schools*

- **Attitudes liées au harcèlement** : Les attitudes liées au harcèlement ont été évaluées à l'aide d'une échelle à 10 items adaptée de la *Provicim scale* de Rigby & Slee (1991 ; cité par Cross et al., 2016a). Les parents devaient répondre sur une échelle de 5 points allant de *Désaccord complet* à *Accord Complet* à des items tels que « *Bullying toughens students up* ». Un score positif est calculé en moyennant les six items reflétant une attitude anti-harcèlement (l'alpha est égal à .56) et un score négatif en moyennant les 4 items reflétant une attitude pro-harcèlement (l'alpha est égal à .65).
- **Habilité à aider son enfant à répondre au harcèlement** : Afin d'évaluer la perception des parents de leur habilité à aider leur enfant à répondre au harcèlement, ils ont dû indiquer sur une échelle à 5 points (allant de *Je ne sais pas* à *Beaucoup*) à quel point ils étaient capables d'encourager leur enfant à aider les autres, d'encourager leur enfant à rapporter le harcèlement aux adultes, de décourager leur enfant à harceler d'autres élèves, d'influencer la façon dont leur enfant répond au harcèlement, d'aider leur enfant à répondre au harcèlement d'une manière qui fait qu'il ne se reproduira plus et de davantage aider leur enfant à faire face au harcèlement que l'école ne le peut. Les réponses ont été moyennées de façon à former un score global (l'alpha vaut .73).
- **Perception que le harcèlement est une priorité pour l'école** : Il était demandé aux parents dans quelle mesure ils pensaient que le harcèlement était une priorité pour l'école de leur enfant sur une échelle de 4 points allant de *Je ne sais pas* à *Une priorité élevée*.

Effets de *Friendly Schools Friendly Families* selon l'évaluation de 2002-2004

1. Des effets ont-ils été observés ?

Le programme a un effet sur la victimisation, mais cet effet n'est significatif qu'à certains post-test. Pour les élèves de 4^{ème} année, la probabilité de se faire fréquemment harceler est plus importante dans le groupe *intensité faible* que dans le groupe *intensité élevée*, mais uniquement au post-test 1 (voir le tableau 2 pour les tailles d'effet). Par ailleurs, il n'y a pas d'effet de l'intervention sur la probabilité de se faire harceler ou non (sans tenir compte de la fréquence de ce harcèlement). Pour les élèves de 6^{ème} année, au post-test 1 la probabilité de se faire harceler (peu importe la fréquence) est plus importante dans le groupe *intensité faible* que dans le groupe *intensité élevée* sans qu'il y ait d'effet sur la fréquence du harcèlement. A l'inverse, au post-test 2, il n'y a pas d'effet sur le fait de se faire harceler ou non, mais la probabilité de se faire fréquemment harceler est plus importante dans le groupe *intensité modérée* que dans le groupe *intensité élevée*.

Le programme n'a aucun effet sur le harcèlement chez les élèves de 6^{ème} année. Par contre, chez les élèves de 4^{ème} année, bien qu'il n'y ait aucun effet au post-test 1, la probabilité de se faire fréquemment harceler est plus importante pour le groupe *intensité modérée* que pour le groupe *intensité élevée* au post-test 2 et la probabilité de se faire harceler est plus importante pour le groupe *intensité faible* que pour le groupe *intensité élevée* au post-test 3 (les autres effets ne sont pas significatifs).

Annexe 2 – Evaluation de *Friendly Schools*

Concernant le rapport de la victimisation, le programme n’a aucun effet chez les élèves de 4^{ème} année. Par contre, chez les élèves de 6^{ème} année, que ce soit au post-test 1 ou 2, la probabilité de rapporter la victimisation vécue est plus importante dans le groupe *intensité élevée* que dans le groupe *intensité faible*.

Les groupes *intensité faible* et *intensité modérée* ne diffèrent sur aucune variable.

Tableau 2. Taille d’effet (OR) chez les élèves de 4^{ème} et 6^{ème} année

			Élèves de 4 ^{ème} année	Élèves de 6 ^{ème} année
Victimisation auto-rapportée	Post-test 1 (novembre 2000)	<i>Modérée</i> vs.	1.27 (0,82-1,97)	1.17 (0,84-1,62)
		<i>Faible</i> vs. <i>Elevée</i>	1.48 (0,97-2,28)	1.54 (1,14-2,08)
	Post-test 2 (octobre 2003)	<i>Modérée</i> vs.	1.16 (0,85-1,60)	1.16 (0,74-1,81)
		<i>Faible</i> vs. <i>Elevée</i>	1.39 (1,02-1,91)	1.36 (0,87-2,13)
	Post-test 3 (octobre 2004)	<i>Modérée</i> vs.	1.13 (0,78-1,63)	/
		<i>Faible</i> vs. <i>Elevée</i>	1.64 (1,06-2,53)	/
Victimisation fréquente auto-rapportée	Post-test 1 (novembre 2000)	<i>Modérée</i> vs.	1.51 (1,01-2,26)	1.39 (0,92-2,12)
		<i>Faible</i> vs. <i>Elevée</i>	1.76 (1,20-2,59)	1.48 (1-2,17)
	Post-test 2 (octobre 2003)	<i>Modérée</i> vs.	1.07 (0,72-1,58)	2.13 (1,19-3,83)
		<i>Faible</i> vs. <i>Elevée</i>	1.14 (0,78-1,67)	1.43 (0,78-2,60)
	Post-test 3 (octobre 2004)	<i>Modérée</i> vs.	0.99 (0,61-1,61)	/
		<i>Faible</i> vs. <i>Elevée</i>	1.26 (0,74-2,16)	/
Harcèlement auto-rapporté	Post-test 1 (novembre 2000)	<i>Modérée</i> vs.	1.31 (0,91-1,88)	1.13 (0,81-1,57)
		<i>Faible</i> vs. <i>Elevée</i>	1.37 (0,96-1,94)	1.34 (0,99-1,81)
	Post-test 2 (octobre 2003)	<i>Modérée</i> vs.	1.32 (0,84-2,05)	0.93 (0,60-1,42)
		<i>Faible</i> vs. <i>Elevée</i>	1.33 (0,86-2,05)	1.00 (0,65-1,53)

Annexe 2 – Evaluation de *Friendly Schools*

	Post-test 3 (octobre 2004)	<i>Modérée</i> vs. <i>Elevée</i>	1.34 (0,90- 1,99)	/
		<i>Faible</i> vs. <i>Elevée</i>	1.74 (1,09- 1,99)	/
Harcèlement fréquent auto-rapporté	Post-test 1 (novembre 2000)	<i>Modérée</i> vs. <i>Elevée</i>	1.08 (0,47- 2,46)	0.74 (0,43- 1,29)
		<i>Faible</i> vs. <i>Elevée</i>	1.12 (0,51- 2,48)	0.72 (0,43- 1,20)
	Post-test 2 (octobre 2003)	<i>Modérée</i> vs. <i>Elevée</i>	2.79 (1,23- 6,35)	1.40 (0,75- 2,63)
		<i>Faible</i> vs. <i>Elevée</i>	2.07 (0,90- 4,76)	1.30 (0,68- 2,48)
	Post-test 3 (octobre 2004)	<i>Modérée</i> vs. <i>Elevée</i>	1.05 (0,51- 2,18)	/
		<i>Faible</i> vs. <i>Elevée</i>	0.95 (0,42- 2,14)	/
Rapport de la victimisation	Post-test 1 (novembre 2000)	<i>Modérée</i> vs. <i>Elevée</i>	1.05 (0,69- 1,60)	1.19 (0,79- 1,76)
		<i>Faible</i> vs. <i>Elevée</i>	1.29 (0,86- 1,90)	1.56 (1,09- 2,23)
	Post-test 2 (octobre 2003)	<i>Modérée</i> vs. <i>Elevée</i>	1.15 (0,70- 1,91)	1.42 (0,95- 2,14)
		<i>Faible</i> vs. <i>Elevée</i>	1.04 (0,63- 1,71)	1.78 (1,21- 2,62)
	Post-test 3 (octobre 2004)	<i>Modérée</i> vs. <i>Elevée</i>	0.87 (0,52- 1,44)	/
		<i>Faible</i> vs. <i>Elevée</i>	0.70 (0,39- 1,26)	/

Inspiré de (Cross et al., 2012, pp. 402-403). Les effets significatifs sont indiqués en gras.

Le programme a également certains effets sur les parents des élèves, comme l'indiquent les résultats de Lester et al. (2017) qui ont analysé les versions courtes des questionnaires. Le programme a un effet significatif sur la fréquence des discussions qui augmentent entre le pré-test et les post-test dans le groupe *intensité élevé* par rapport au groupe *intensité faible*. Cette augmentation est présente pour les mères comme pour les pères. *Friendly Schools Friendly Families* a également un effet sur la perception des parents de leur influence sur la décision de leur enfant de harceler ou non. Cependant, alors que cet effet va dans le sens attendu chez les pères (ceux du groupe *intensité élevée* perçoivent une augmentation de leur influence par rapport à ceux du groupe *intensité faible*), il va dans le sens contraire aux hypothèses chez les mères (la perception d'influence diminue dans le groupe *intensité élevée* par rapport au groupe *intensité faible*).

Annexe 2 – Evaluation de *Friendly Schools*

Le programme n'a pas d'effet significatif concernant la perception des parents de l'influence qu'ils ont sur la réponse de leur enfant face au harcèlement. A tous les temps de mesure approximativement la moitié des pères et mères perçoivent qu'ils et elles ont beaucoup d'influence sur la réponse de leur enfant. Le programme n'a pas non plus d'effet sur les conseils donnés. Que ce soit chez les mères ou les pères, les conseils que les parents donnent à leur enfant ne changent pas avec le temps (il s'agit principalement des conseils de recherche d'aide, prosociaux ou passifs, avec quelques différences entre les pères et mères).

Cross et al. (2016a) ont analysé les données issues des versions longues des questionnaires. Au post-test 1, les parents des élèves de 2^{ème} année discutent davantage du harcèlement avec leur enfant dans la condition *intensité élevée* que dans les conditions *intensité faible* et *intensité modérée*. Cet effet se maintient au post-test 2, mais uniquement pour la différence entre les conditions *intensité élevée* et *intensité faible*. Chez les parents des élèves de 4^{ème} année, cet effet n'est présent qu'au post-test 1 et ne concerne que la différence entre les groupes *intensité élevée* et *intensité faible*, tandis que chez les parents des élèves de 6^{ème} année le programme n'a aucun effet significatif sur la fréquence des discussions autour du harcèlement. Le programme a également un effet significatif sur l'efficacité personnelle des parents, mais cet effet concerne uniquement les parents des élèves de 4^{ème} année au post-test 1 : les parents du groupe *intensité élevée* rapportent une plus grande efficacité personnelle à discuter du harcèlement avec leur enfant que ceux du groupe *intensité modérée*. Les attitudes pro-harcèlement sont également uniquement impactées par le programme au post-test 1 chez les parents des élèves de 4^{ème} année : les parents du groupe *intensité élevée* rapportent moins d'attitudes pro-harcèlement envers le harcèlement que ceux des groupes *intensité modérée* et *intensité faible*.

Le programme n'a aucun effet significatif sur les connaissances liées au harcèlement, l'habilité à aider son enfant à répondre au harcèlement, les attitudes positives liées au harcèlement et la perception que la lutte contre le harcèlement est une priorité pour l'école de leur enfant.

2. Les effets sont-ils différenciés ?

Les effets observés chez les élèves de 4^{ème} et de 6^{ème} année ne sont pas exactement les mêmes. Au niveau de la victimisation, des effets sont observés aux post-test 1 et 2 chez les élèves de 6^{ème} année, mais uniquement au post-test 1 chez ceux et celles de 4^{ème} année. Alors que le programme a certains effets sur le harcèlement chez les élèves de 4^{ème} année, il n'en a aucun chez ceux et celles de 6^{ème} année. A l'inverse le programme n'a aucun effet sur le rapport de la victimisation chez les élèves de 4^{ème} année, mais en a chez les élèves de 6^{ème} année.

Concernant les parents, il semble que les effets du programme soient moins importants chez les parents des élèves de 6^{ème} année que chez ceux des élèves de 2^{ème} ou 4^{ème} année. Une différence entre les pères et les mères est observée concernant l'effet du programme sur l'influence des parents sur la décision de leur enfant de harceler ou non. Alors que l'effet va dans le sens attendu pour les pères, il va dans le sens contraire des hypothèses pour les mères.

Les différences observées se situent entre le groupe *intensité élevée* et les groupes *intensité faible* et *intensité modérée*. Entre ces deux derniers, aucune différence n'est retrouvée. Il semble donc que le renforcement des facteurs familiaux soit bien nécessaire pour observer des effets du programme.

3. Des mesures de suivi ont-elles été prises ?

Le programme a été implanté sur trois ans, mais de façon plus active les deux premières années. Les mesures ont été prises lors de plusieurs post-test. Pour les élèves, une mesure de suivi a été prise 19 mois après le pré-test (post-test 2), ainsi qu'une autre 31 mois après le pré-test (post-test 3) pour les élèves de 4^{ème} année. Les résultats sont variables d'un post-test à l'autre : certains effets ne se maintiennent pas tandis que d'autres apparaissent uniquement au second post-test. Pour les parents, une mesure de suivi (post-test 2) a été prise 22 mois après le pré-test. Certains effets se maintiennent dans le temps, mais ce n'est pas le cas de la majorité d'entre eux.

4. L'implantation a-t-elle été mesurée ?

L'implantation de *Friendly Schools Friendly Families* a été mesurée via des questionnaires et des interviews tout au long de l'étude. À la fin de chaque année, dans chaque école des groupes *intensité élevée* et *intensité modérée*, le comité de l'école rapportait l'implantation au niveau de l'école, les enseignant·e·s au niveau des classes et les élèves et leurs parents au niveau des familles. En moyenne, sur les trois ans, le niveau d'implantation était de 63 % au niveau de l'école (groupes *intensité élevée* et *modérée*), de 55 % au niveau des classes (groupes *intensité élevée* et *modérée*) et de 22 % au niveau des familles (groupe *intensité élevée* uniquement).

Selon Cross et al. (2016a), une plus grande proportion de parents dans les groupes *intensité élevée* (86 %) et *intensité modérée* (87 %) rapportent se souvenir avoir vu des newsletters que dans le groupe *intensité faible* (60 %). La majorité des parents (environ 70 %) ont trouvé ces newsletters utiles. 84 % des parents du groupe *intensité élevée* se souviennent avoir reçu les activités à faire à la maison et 78 % les avoir trouvées utiles. La plupart des parents trouvent que ces ressources leur permettent d'augmenter leur compréhension des réponses de l'école ainsi que d'augmenter leurs connaissances, confiance et compétence pour aborder le harcèlement. De plus, environ un quart des parents de ce groupe rapportent un changement dans leurs attitudes ou une amélioration de leur propre comportement. Bien qu'aucune analyse n'ait été menée, il semble que, dans le groupe *intensité modérée*, le programme ait uniquement permis d'augmenter les connaissances des parents et leur compréhension des réponses de l'école sans impacter leur confiance et compétence à aborder le harcèlement.

Implantation de *Cyber Friendly Schools*

Qualité de l'évaluation de *Cyber Friendly Schools* (2010-2012)

Article principal : Cross, D., Shaw, T., Hadwen, K., Cardoso, P., Slee, P., Roberts, C., ... Barnes, A. (2016b). Longitudinal impact of the Cyber Friendly Schools program on adolescents' cyberbullying behavior. *Aggressive Behavior*, 42, 166-180. doi:10.1002/ab.21609

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude longitudinale avec un design RCT. Les écoles ont été aléatoirement assignées à la condition expérimentale ou contrôle (utilisation habituelle de programme de prévention du harcèlement). L'étude s'est déroulée de 2010 à 2012 pour la partie principale et de 2010 à 2011 pour la partie portant sur les cyberleaders et rapportée dans l'article de Cross et al. (2015b).

Pour tenir compte du fait que les données ne se distribuaient pas normalement (la plupart des élèves n'étant pas impliqué-e-s dans les situations de cyberharcèlement) des modèles de croissance en deux parties ont été utilisés. Ces modèles tiennent compte de la distribution asymétrique des scores et renseignent tant sur la probabilité d'être victime ou auteur-e de cyberharcèlement que sur la gravité de la situation. Comme l'étude a été réalisée sur plusieurs années et que les classes ont changé, les modèles tiennent uniquement compte du fait que les élèves sont nichées dans les écoles.

Des procédures de *Full Information Maximum Likelihood* sous le postulat MAR (*missing at random*) ont été utilisées pour tenir compte des données manquantes. Toutefois certains cas (181) avec des données manquantes sur au moins un des prédicteurs au pré-test ou avec des données manquantes sur le cyberharcèlement ou la cybervictimisation à tous les points de mesure ont été exclus de l'analyse.

2. Quelles étaient les caractéristiques de l'échantillon ?

Toutes les écoles secondaires non gouvernementales de la ville de Perth comprenant au moins 90 élèves de 8^{ème} année ont été invitées à participer à l'étude. Sur les 53 écoles éligibles, 35 ont accepté de participer. Les écoles ont été classées par type (mixte ou non mixte), statut socio-économique et nombres d'élèves de 8^{ème} année avant d'être réparties entre les conditions (19 écoles dans la condition expérimentale, 16 dans la condition contrôle). Sur les 35 écoles, 7 étaient des écoles de filles et 3 des écoles de garçons, les écoles restantes étant mixtes.

L'échantillon est composé de 3382 élèves de 8^{ème} année au pré-test. 2940 élèves ont participé au post-test 1 et 2874 au post-test 2 (ce qui donne un taux d'abandon de 15 %). Les élèves venaient plutôt de familles au statut socio-économique élevé, 75 % d'entre eux et elles vivant dans des banlieues plus aisées que la moyenne, avec seulement 14 % de familles monoparentales. Tou-te-s les élèves ont accepté de participer, mais 7 % n'ont pas reçu le consentement parental. Il n'y avait aucune différence entre les deux groupes au pré-test, que ce soit au niveau des variables démographiques ou des variables d'intérêt.

Annexe 2 – Evaluation de *Friendly Schools*

L'étude de Cross et al., (2015b) porte sur les élèves qui ont joué un rôle de cyberleaders en 2010 ou en 2011. Ces élèves étaient en 10^{ème} année et avaient 14-15 ans. Chaque année, chaque école de la condition expérimentale a recruté entre 4 et 6 élèves pour devenir cyberleader. Le personnel scolaire sélectionnait des élèves intéressé·e·s par la technologie et étant des leaders sociaux positifs dans l'école. 87 élèves ont accepté de prendre ce rôle en 2010 et 138 en 2011. Ils et elles ont tou·te·s accepté de participer à l'étude et reçu le consentement parental, mais seulement 157 d'entre eux et elles ont répondu aux questionnaires (sans qu'il ne soit précisé pourquoi). La moitié des participant·e·s sont des filles, ils et elles viennent davantage de milieux favorisés et n'étaient, pour la majorité, pas victime de cyberharcèlement.

3. Par qui et où l'évaluation a-t-elle été menée ?

L'étude, comme les deux précédentes, a été réalisée à Perth. Elle a également été menée essentiellement par les mêmes chercheur·euse·s : Donna Cross, Thérèse Shaw, Leanne Lester et Amy Barnes (The University of Western Australia), Kate Hadwen, Patricia Cardoso et Laura Thomas (Edith Cowan University), Philip Slee (Flinders University) et Clare Roberts (Curtin University of Technology).

4. Quelles et combien de mesures ont été prises ?

Les élèves ont complété le questionnaire au début de leur 8^{ème} année (pré-test) ainsi qu'à la fin de leur 9^{ème} et 10^{ème} année (post-test 1 et 2). Les moments de mesures ne sont donc pas les mêmes pour le pré-test et les post-test. Les données ont été récoltées durant les heures scolaires par des assistant·e·s de recherche formé·e·s.

Au niveau des mesures démographiques, il était demandé aux élèves leur âge, genre, résultats académiques ainsi que s'ils et elles vivaient ou non avec leurs deux parents. Mis à part cela, seuls le cyberharcèlement et la cybervictimisation ont été mesurés. Ils ont été évalués via une échelle de 11 items basée sur celle développée par Smith, Mahdavi, Carvalho et Tippett (2006; cité par Cross et al., 2016b) ainsi que sur un travail formatif conduit par l'équipe de recherche auprès de jeunes. L'échelle débute par des définitions du harcèlement et cyberharcèlement (basées sur celles de Olweus, 1996 et Smith et al., 2006; cités par Cross, et al., 2016b), accompagnées d'exemples et de pictogrammes. Il était ensuite demandé « *Last term, how often were you cyberbullied in the following ways ?* » aux élèves qui devaient indiquer pour chaque item (ex. « *I had pictures/video clips about me sent to others' mobile phone to embarrass, hurt or upset me.* ») à quelle fréquence cela leur était arrivé parmi les possibilités de réponse suivantes : *Cela ne m'est pas arrivé, Une ou deux fois, Toutes les quelques semaines, Environ une fois par semaine, Quelques fois par semaine ou plus*. L'échelle de cyberharcèlement était identique, mais formulée de manière active. Les alphas sont égaux à .86 pour la cybervictimisation et .91 pour le cyberharcèlement.

Les cyberleaders ainsi que les membres du personnel qui les encadraient ont complété des questionnaires en ligne et, pour certain·e·s, participé à des entretiens. Les questionnaires (pré et post-formation) mesuraient l'usage d'internet, la cybervictimisation et le cyberharcèlement, le soutien des pairs, les liens avec l'école, l'utilité de la formation, l'engagement dans leur rôle, l'aide apportée par le personnel scolaire, l'efficacité personnelle et les compétences des cyberleaders. Les entretiens ont été menés avec un moins un leader dans chaque école avant et après chaque année d'étude (soit 12 participant·e·s en 2010 et 30 en 2011).

Effets de *Cyber Friendly Schools* selon l'évaluation de 2010-2012

1. Des effets ont-ils été observés ?

En règle générale, sans prendre en compte l'effet du groupe, la probabilité d'être concerné par la cybervictimisation ou le cyberharcèlement a tendance à diminuer avec le temps tandis que l'intensité du problème pour les élèves concerné-e-s a tendance à augmenter. Le programme a des effets significatifs sur la probabilité d'être concerné par la cybervictimisation ou le cyberharcèlement : entre le pré-test et le post-test 1, la probabilité d'être victime ou auteur-e de cyberharcèlement diminue davantage dans la condition expérimentale que dans la condition contrôle (voir tableau 3). Par contre, il n'y a pas de différence entre les conditions concernant l'évolution du post-test 1 au post-test 2 et, pour les élèves concerné-e-s, il n'y a aucune différence d'intensité du problème entre les deux groupes.

Tableau 3. Évolution des taux de cybervictimisation et cyberharcèlement (% d'élèves concerné-e-s)

	Pré-test	Post-test 1	Post-test 2
Cybervictimisation			
Condition contrôle	27 %	26 %	24 %
Condition expérimentale	30 %	25 %	23 %
Cyberharcèlement			
Condition contrôle	9 %	10 %	7 %
Condition expérimentale	11 %	8 %	7 %

Inspiré de Cross et al. (2016b, p. 172)

2. Les effets sont-ils différenciés ?

Les auteur-e-s n'ont réalisé aucune analyse permettant de savoir si les effets sont différenciés.

3. Des mesures de suivi ont-elles été prises ?

Le programme a été implanté sur deux ans. Un deuxième post-test a été réalisé un an après le premier. Alors qu'entre le pré-test et le post-test 1, une diminution plus importante de la cybervictimisation et du cyberharcèlement est observée dans la condition expérimentale, au post-test 2, il n'y a plus aucune différence entre les deux groupes. Les effets positifs de *Cyber Friendly Schools* ne se maintiennent donc pas dans le temps.

4. L'implantation a-t-elle été mesurée ?

Les enseignant-e-s de 8^{ème} et 9^{ème} année ont rapporté avoir implanté moins d'un tiers des neuf modules (par manque de temps et manque de confiance en leurs capacités).

La plupart des cyberleaders ont rapporté que la formation leur était utile (88 %), qu'ils et elles étaient engagé-e-s dans leur rôle (94 %), recevaient du soutien de la part de l'école (93 %) et pouvaient demander de l'aide au personnel scolaire (93 %). Après la formation, la plupart des élèves (environ 90 %) se sentaient compétent-e-s, ce qui ressort également des entretiens. Les élèves mettent aussi en avant le fait que

Annexe 2 – Evaluation de *Friendly Schools*

L'information qu'ils et elles fournissent est plus crédible et pertinente que si elle était venue d'un adulte. Les facteurs perçus comme augmentant leur efficacité sont la reconnaissance formelle de leur rôle par l'école, le soutien et le respect reçus de la part des autres élèves, du personnel et des parents, l'augmentation de confiance dans leur rôle et de leurs compétences de leadership, le fait d'avoir un groupe cohésif et coopératif de cyberleaders, et le fait d'avoir suffisamment de temps à l'école pour travailler en équipe. Au contraire, les facteurs inhibant leur efficacité sont le fait de ne pas avoir de soutien de la part du personnel et des parents, le fait que les élèves n'utilisent pas les ressources qui leur sont données, un manque d'intérêt dans le projet de la part des autres élèves, une pression négative des pairs, et un manque de communication entre les différentes parties impliquées.

Résumé des effets de *Friendly Schools*

Les effets observés sont-ils constants d'une étude à l'autre ?

Les effets ne sont pas les mêmes d'une étude à l'autre, ce qui est normal étant donné qu'il s'agit à chaque fois de l'évaluation d'une version différente du programme. *Friendly School* a permis une réduction de la victimisation auto-rapportée mais seulement en la mesurant d'une certaine manière et à certains points de mesure. Le programme n'a eu aucun effet sur le harcèlement rapporté, mais a permis de réduire la victimisation observée. Au niveau des variables secondaires, seuls trois effets sont significatifs sur les huit mesures prises dont un allant dans le sens contraire des hypothèses.

Il était attendu que les effets de *Friendly Schools Friendly Families* soient plus importants étant donné que le programme avait été renforcé. Le programme est bien associé à une diminution de la victimisation. Toutefois, sur les vingt différences testées seules trois sont significatives. Le programme a également un effet sur le harcèlement, mais uniquement pour les élèves de 4^{ème} année (et là encore, seules deux différences sont significatives sur les douze testées). Il permet également de réduire le rapport de la victimisation vécue, mais seulement pour les élèves de 6^{ème} année. *Friendly Schools Friendly Families* est aussi associé à des effets positifs au niveau des parents d'élèves. Il permet d'augmenter la fréquence des discussions parent-enfant autour du harcèlement dans les familles des élèves de 2^{ème} et 4^{ème} année (mais pas de 6^{ème} année) et a un impact positif sur l'efficacité personnelle et les attitudes des parents de 4^{ème} année. Toutefois, le programme n'a pas d'effets sur une série d'autres variables investiguées au niveau des parents.

Les résultats de l'évaluation de *Cyber Friendly Schools* sont plus cohérents : le programme permet de réduire la probabilité d'être concerné par le cyberharcèlement ou la cybervictimisation, mais pas la fréquence de ces phénomènes. De plus, les effets ne se maintiennent pas dans le temps.

La dernière version du programme *Friendly Schools Plus* n'a jamais été évaluée ce qui pose problème même si on peut supposer que cette version est au moins aussi efficace que les précédentes.

Le programme semble-t-il plus efficace pour une certaine population ?

Les auteur·e·s ne rapportent pas vraiment d'analyses permettant de savoir si le programme est plus efficace pour certain·e·s élèves que pour d'autres. La seule différence investiguée concerne le niveau scolaire, mais elle l'est uniquement dans une étude (celle de Cross et al., 2012) et les résultats ne sont pas très clairs (effets sur la probabilité de se faire harceler chez les 6^{ème} et sur la fréquence du harcèlement chez les 4^{ème}, sur le harcèlement auto-rapporté chez les 4^{ème} et sur le rapport de la victimisation chez les 6^{ème}).

Les trois études ont été réalisées dans des écoles de la ville de Perth, ce qui limite fortement la possibilité de généraliser les résultats obtenus à d'autres contextes.

Annexe 3 : Evaluation de *Zero*

Le programme a été évalué à deux reprises en Norvège. Tout d’abord, de 1992 à 1994, une étude pilote a été réalisée pour évaluer les effets de l’implantation des aspects indirects qui seront repris dans le programme *Zero*. Les composantes implantées à cette occasion sont donc uniquement celles qui concernent la façon dont l’enseignant·e gère sa classe et non celles qui concernent directement le harcèlement. Le programme complet a ensuite été évalué lors de son implantation en 2003-2004. Une version dérivée du programme a également été implantée en Irlande. Ce sont ces trois évaluations qui seront développées dans la suite du chapitre.

Implantation de *Zero* en Norvège

Le développement du programme *Zero* est lié de près au travail de lutte contre le harcèlement scolaire effectué en Norvège depuis 1983. En 1983, les cas non liés de deux adolescents s’étant suicidés suite au harcèlement qu’ils subissaient ont été rapportés dans les médias. Suite à cela, le ministre de l’Éducation a établi un comité composé entre autres de deux chercheurs, Dan Olweus et Erling Roland, afin de préparer une initiative au niveau national. Ces chercheurs ont écrit un livre sur le harcèlement scolaire, comprenant une définition du phénomène qui sert maintenant de référence, et rédigé le scénario d’un film (Roland, 2011). C’est ce matériel qui a servi de base pour l’effort national de lutte contre le harcèlement scolaire débutant en 1983 (Roland & Midthassel, 2012).

Le programme *Zero*, tout comme le programme *Olweus* contre le harcèlement, tire ses racines de cette première initiative nationale. Bien que le nom «*Zero*» n’ait été adopté qu’en 2003, le programme est plus ancien. Il date de 1996, c’est-à-dire de l’année du deuxième effort national de lutte contre le harcèlement. Le ministre de l’Éducation avait alors donné au *Center for Behavioural Research* de l’University of Stavanger la responsabilité de combattre le harcèlement scolaire. C’est en 1996 qu’une première version du guide de l’enseignant·e est sortie (Roland & Midthassel, 2012).

Le Manifeste Contre le Harcèlement a été annoncé par le Premier ministre ainsi que les acteurs clés de l’enseignement en septembre 2002 (il a pris fin et été évalué en septembre 2004). Le ministre de l’Éducation a alors recommandé deux programmes pour réduire le harcèlement : le programme *Olweus* et le programme *Zero*. Ces deux programmes ont été offerts à toutes les écoles norvégiennes. Les écoles devaient choisir un des deux programmes ou mettre en place leur propre projet. De nombreuses écoles ont choisi d’implanter l’un des programmes (Roland et al., 2010). L’état norvégien apporte un soutien financier aux développeurs des programmes *Olweus* et *Zero*, les écoles ne doivent donc que se charger des frais de fonctionnement (Stephens, 2011). C’est suite à ce manifeste, qui faisait de la tolérance zéro un élément clé de lutte contre le harcèlement, que le programme a été rebaptisé «*Zero*» en 2003. Le matériel de 1996 a été revu, mis à jour et complété. De plus, une emphase plus importante a été mise sur l’engagement des écoles dans l’implantation du programme et le support de l’implantation dans les écoles a été renforcé (Roland & Midthassel, 2012).

La Norvège a donc une longue histoire de lutte contre le harcèlement. Toutefois, malgré les efforts réalisés et la diminution des taux de harcèlement et de victimisation entre 2001 et 2004, des données plus récentes

Annexe 3 – Evaluation de Zero

montrent une augmentation importante, d'environ 70 %, du harcèlement entre 2004 et 2008 (Roland & Midthassel, 2012) alors qu'un deuxième manifeste a été lancé par le gouvernement en janvier 2006 (Roland, 2011). Il semble que le cyberharcèlement explique cette augmentation (Roland et Auestad, 2009; cité par Roland & Midthassel, 2012).

Qualité de l'évaluation pilote de Zero

Article : Galloway, D., & Roland, E. (2004). Is the direct approach to reducing bullying always the best? In P. K. Smith, D. J. Pepler, & K. Rigby (Eds.), *Bullying in schools : how successful can interventions be?* (pp. 37–53). Cambridge University Press.

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude quasi-expérimentale. Les neuf écoles expérimentales ont été choisies par un responsable de l'enseignement de la ville ayant accepté de sélectionner un échantillon représentatif d'écoles. Les neuf autres écoles de la ville ont été invitées à servir de groupe contrôle et six d'entre elles ont accepté. Le programme a été implanté lors de l'année scolaire 1992-1993 pour une partie des écoles et lors de l'année scolaire 1993-1994 pour l'autre partie. L'étude comporte deux groupes contrôles. Le premier est composé d'élèves de 1^{ère} année des écoles expérimentales qui ont complété le questionnaire à la fin de l'année précédant l'implantation du programme (1991-1992). Le second est composé des élèves des écoles contrôles (qui ont complété le questionnaire à la fin de l'année scolaire 1993-1994).

Étant donné que l'étude porte sur des élèves de 1^{ère} année, qui n'étaient donc pas encore présent·e·s dans l'école l'année précédente, aucun pré-test n'a pu être réalisé. Les données des écoles contrôles et celles des écoles expérimentales ont été comparées via des tests-t.

2. Quelles étaient les caractéristiques de l'échantillon ?

L'étude s'est déroulée dans une ville du sud de la Norvège comptant 18 écoles primaires. Comme elles différaient très peu, les données des élèves des deux groupes expérimentaux (celui de 1992-1993 et celui de 1993-1994), tout comme celles des deux groupes contrôles (celui de 1991-1992 et celui de 1993-1994), ont été combinées dans les analyses. Le groupe expérimental comporte entre 600 et 700 élèves et le groupe contrôle entre 450 et 500 élèves. 20 enseignant·e·s ont participé au programme de développement professionnel en 1992-1993 et 20 autres en 1993-1994.

Aucune information sur la présence de données manquantes et la façon dont elles ont été gérées n'est rapportée dans l'article.

3. Par qui et où l'évaluation a-t-elle été menée ?

L'étude pilote a été menée en Norvège par Erling Roland, le concepteur de Zero, ainsi que par David Galloway (également de l'University of Stavanger).

4. Quelles et combien de mesures ont été prises ?

Les enseignant·e·s ont répondu au questionnaire à la fin de l'année d'implantation du programme (juin 1993 ou 1994 suivant les écoles), tout comme les élèves de la condition expérimentale. Les élèves de la condition contrôle ont également rempli le questionnaire en fin d'année (1992 pour le premier groupe contrôle, 1994 pour le second). Lorsque les élèves ont complété le questionnaire anonyme, ils et elles étaient encadré·e·s par leur titulaire de classe qui lisait chaque question à voix haute. Le questionnaire était composé de 10 items auxquels les élèves devaient répondre via une échelle à 3 ou 4 points. Les 7 premiers items portaient sur les sentiments des élèves envers l'école (ex. «*Do you sometimes not want to go to school ?*») et sur leur comportement social (ex. «*Do you help other pupils when they need it ?*»). Le huitième item servait de contrôle («*Do you feel sad at home ?*»). Venait ensuite une définition du harcèlement suivi par un item portant sur la victimisation («*Does it happen that you are bullied by other pupils at school ?*») et un autre portant sur le harcèlement («*Does it happen that you take part in bullying other pupils at school ?*»). Pour ces deux items, les réponses disponibles étaient *Jamais*, *De temps en temps*, *Chaque semaine* et *Plusieurs fois par semaine*. Les points de coupure utilisés pour calculer les taux de harcèlement et de victimisation ne sont pas précisés.

Effets de Zero selon l'évaluation pilote

1. Des effets ont-ils été observés ?

Comme prévu par les hypothèses, les élèves dont les enseignant·e·s ont reçu le programme ont des valeurs plus positives que les autres élèves pour tous les items, excepté celui servant de contrôle. Le taux de harcèlement était 15 % moins élevé dans le groupe expérimental que dans le groupe contrôle et celui de victimisation 18,7 %. Ces données soutiennent l'intérêt d'une approche holistique pour lutter contre le harcèlement.

2. Les effets sont-ils différenciés ?

Les auteurs ne rapportent aucune analyse permettant de savoir si les effets sont différenciés.

3. Des mesures de suivi ont-elles été prises ?

Les auteurs font l'hypothèse que les effets de cette intervention devraient mieux se maintenir que les effets de programmes ciblant directement le harcèlement. Mais, comme ils n'ont pris aucune mesure de suivi, il n'est pas possible de savoir si c'est effectivement le cas.

4. L'implantation a-t-elle été mesurée ?

Aucune mesure de l'implantation n'est rapportée dans l'étude.

Qualité de l'évaluation de 2003-2004

Article principal : Roland, E., Bru, E., Midthassel, U. V., & Vaaland, G. S. (2010). The Zero programme against bullying: Effects of the programme in the context of the Norwegian manifesto against bullying. *Social Psychology of Education*, 13(1), 41-55. doi:10.1007/s11218-009-9096-0

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude longitudinale sans réel groupe contrôle. Le programme a été implanté lors de l'année scolaire 2003-2004, avec un pré-test réalisé en mai-juin 2003 et un post-test en mai-juin 2004. Étant donné que toutes les écoles norvégiennes devaient implanter un projet anti-harcèlement à partir de cette année, il n'était pas possible d'avoir un groupe contrôle. Le changement entre le pré et le post-test observé dans les écoles de l'échantillon a été comparé aux données d'une autre étude mesurant le changement des taux de harcèlement et victimisation entre 2001 et 2004.

Au niveau des analyses statistiques, des régressions et analyses de covariances ont été effectuées, en contrôlant pour le genre, le niveau scolaire, ainsi que les changements pré/post-test attribués au genre et au niveau scolaire.

2. Quelles étaient les caractéristiques de l'échantillon ?

Toutes les écoles norvégiennes avaient la possibilité d'implanter le programme *Zero* dans le cadre du Manifeste Contre le Harcèlement. 151 écoles primaires ont posé leur candidature et ont toutes été acceptées. Toutefois, 6 écoles ont quitté le programme au cours de l'année d'implantation. L'échantillon total est donc composé de 146 écoles, réparties dans toutes les provinces de Norvège. Les écoles sont de différentes tailles et situées dans des régions au niveau d'urbanisation varié. Au pré-test, les données ont été récoltées chez 20446 élèves de la 2^{ème} à la 7^{ème} année. Au post-test, elles ont été récoltées chez 20430 élèves, toujours de la 2^{ème} à la 7^{ème} année, afin que les résultats ne soient pas biaisés par l'effet de l'âge.

Les données contrôles viennent d'enquêtes sur l'environnement scolaire réalisées par le *Centre for Behavioural Research* en mai-juin 2001 et mai-juin 2004. Les deux échantillons sont indépendants et comprennent 3979 et 2828 élèves de 5^{ème}, 6^{ème} et 7^{ème} année. Les deux enquêtes ont été réalisées dans des provinces considérées comme représentatives de la Norvège (Statistics Norway, 1994; cité par Roland et al., 2010). Toutes ces écoles étaient supposées mettre en place un projet de lutte contre le harcèlement, que ce soit le programme *Zero*, le programme *Olweus* ou un autre projet. De plus, le personnel, les élèves et leurs parents étaient massivement exposés à la lutte nationale anti-harcèlement, principalement via les médias, et de nombreuses formations sur le harcèlement étaient offertes au personnel des écoles. Toutefois, aucune information sur ce que les écoles ont effectivement mis en place pour réduire le harcèlement n'est disponible. La seule question qui leur a été posée concerne les changements effectués entre la période de 1998-2001 et l'année 2003-2004 concernant la lutte anti-harcèlement. 75 % des écoles rapportent avoir augmenté leurs efforts pour réduire le harcèlement. Aucune information sur la présence de données manquantes et la façon dont elles ont été gérées n'est rapportée dans l'article.

L'étude de Midthassel, Bru et Idsoe (2008) porte sur le maintien des effets du programme à plus long terme. Toutes les écoles ayant implanté *Zero* en 2003-2004 ont été invitées à participer à l'étude. L'étude porte sur

Annexe 3 – Evaluation de Zero

72 écoles dans lesquelles les questionnaires ont été complétés en 2003, 2004 et 2006. Ces écoles ont été aléatoirement divisées en trois groupes. Dans le premier groupe, les écoles ont reçu, durant les années 2004-2005 et 2005-2006, du matériel lié à la lutte anti-harcèlement (des leçons à destination du personnel scolaire, du matériel à utiliser en classe...). Dans le deuxième groupe, en plus de recevoir ce matériel, chaque école était liée à un·e chercheur·euse du *Centre for Behavioural Research* qui leur téléphonait régulièrement et était joignable en cas de problème. Dans le troisième groupe, aucun soutien n'était apporté après la première année d'implantation.

3. Par qui et où l'évaluation a-t-elle été menée ?

L'évaluation est menée en Norvège par Erling Roland, Edvin Bru, Unni Vere Midthassel, Sigrun Ertesvåg et Grete Vaaland. Ils et elles font tou-te-s parti du *Centre for Behavioural Research* de l'University of Stavanger, là où *Zero* a été développé.

4. Quelles et combien de mesures ont été prises ?

Les données pré-test ont été prises en mai-juin 2003 pour les écoles implantant *Zero* et en mai-juin 2001 pour les écoles participant à l'enquête sur l'environnement scolaire. Toutes les données post-test ont été prises en mai-juin 2004. Dans les écoles concernées, des données ont été prises en fin d'année scolaire 2005-2006.

La procédure était pratiquement identique dans les écoles implantant *Zero* et celles participant à l'enquête sur l'environnement scolaire. Tou-te-s les élèves ayant obtenu le consentement parental et acceptant de participer ont répondu au questionnaire. Ils et elles étaient encadré·e-s par leur titulaire qui avait reçu des instructions écrites sur la façon de faire passer le questionnaire. L'enseignant·e précisait bien que les réponses des élèves resteraient anonymes et lisait chaque question à voix haute. Toutes les classes d'une même école devaient répondre au questionnaire au même moment. Pour les écoles participant à *Zero*, le taux de réponse était de 94 % au pré-test et 93 % au post-test. Pour l'enquête sur l'environnement scolaire, il était de 87 % au pré-test et 81 % au post-test.

Les mesures qui ont été prises sont les suivantes :

- **Harcèlement** : Le harcèlement a été mesuré via un item général (« *This school year how often have you bullied/hassled other pupils at school?* ») et trois items portant sur des formes spécifiques de harcèlement (« *Bullied/hassled other pupils by teasing them?* », « *Bullied/hassled other pupils at school by isolating/shutting them out?* » et « *bullied/hassled other pupils at school by hitting, kicking or shoving them?* »). À chaque item, les élèves ont répondu sur une échelle à 4 points (*Jamais, De temps en temps, Chaque semaine* et *Chaque jour*).
- **Victimisation** : La victimisation a été mesurée de la même manière que le harcèlement, mais avec les items formulés de manière passive.

Les alphas sont supérieurs à .70 dans tous les cas. Pour le calcul des pourcentages, les élèves ayant répondu *Chaque semaine* et *Chaque jour* ont été catégorisé·e-s comme victime ou auteur·e de harcèlement.

Effets de Zero selon l'évaluation de 2003-2004

1. Des effets ont-ils été observés ?

Dans les écoles où Zero a été implanté, il y a une réduction significative du score moyen de victimisation entre le pré et le post-test (et cette réduction est plus importante parmi les élèves de 2-4^{ème} année que parmi ceux de 5-7^{ème} année). Toutefois une réduction de même ampleur est observée dans le groupe contrôle. Il y a également une réduction significative du pourcentage d'élèves victimes dans la condition expérimentale, mais, à nouveau, cette réduction n'est pas différente de celle observée dans le groupe contrôle (voir tableau 1).

Dans les écoles ayant implanté le programme Zero, il y a une réduction significative du score moyen de harcèlement entre les deux temps de mesure (et cette réduction est plus importante parmi les élèves de 5-7^{ème} année que parmi ceux de 2-4^{ème} année). Toutefois une réduction de même ampleur est observée dans le groupe contrôle. Concernant le pourcentage d'élèves auteur·e·s de harcèlement, bien que les taux de prévalence vont dans le sens prévu, la différence entre le pré et le post-test n'est pas significative. Et il n'y a pas de différence de changement entre la condition contrôle et la condition expérimentale (voir tableau 1).

Tableau 1. Evolution des taux de victimisation et harcèlement chez les élèves de la 5^{ème} à la 7^{ème} année.

	Pré-test	Post-test	Pourcentage de changement
Victimisation			
Contrôle	6,8 %	5,4 %	-20,59 %
Intervention	7 %	5,7 %	-18,57 %
Harcèlement			
Contrôle	1,8 %	1,6 %	-11,11 %
Intervention	1,2 %	1,1 %	-8,33 %

Inspiré de Roland et al. (2010, p. 49). Les changements significatifs sont en gras.

En résumé, une diminution du harcèlement et de la victimisation est observée suite à l'implantation de Zero dans les écoles. Toutefois, cette diminution n'est pas plus importante que dans les autres écoles norvégiennes. Il n'est donc pas possible de conclure qu'elle est spécifiquement due au programme.

2. Les effets sont-ils différenciés ?

Les effets ne sont pas différenciés en fonction du genre. Par contre, concernant le niveau scolaire, la réduction du harcèlement est plus importante chez les élèves les plus âgé·e·s tandis que celle de la victimisation est plus importante chez les élèves plus jeunes. Les auteur·e·s ne rapportent aucune autre analyse permettant de savoir si les effets sont différenciés.

3. Des mesures de suivi ont-elles été prises ?

Une mesure de suivi a été prise par Midthassel et al. (2008) en juin 2006. Aucune différence n'est retrouvée entre le post-test et le suivi concernant la victimisation. Concernant le harcèlement, un déclin léger, mais significatif des scores est retrouvé. Les améliorations observées après un an d'implantation se maintiennent donc deux ans plus tard. Contrairement aux hypothèses des auteur·e·s, les résultats sont identiques pour les trois groupes (aucun soutien, soutien via matériel et soutien via matériel et téléphone). Selon les auteur·e·s ce serait soit parce que le soutien additionnel n'apporte rien de plus (les écoles étaient déjà assez compétentes suite à la première année d'implantation) soit parce que le matériel reçu pourrait ne pas avoir été utilisé comme recommandé.

4. L'implantation a-t-elle été mesurée ?

L'implantation n'a pas été mesurée de manière quantitative. Par contre, Midthassel et Ertesvåg (2008) ont réalisé des entretiens qualitatifs dans six écoles primaires faisant partie de l'échantillon de cette étude. Elles ont remarqué que les écoles qu'elles avaient considérées comme totalement prêtes à implanter un programme anti-harcèlement au premier entretien avaient réussi l'implantation du programme avec succès tandis que les écoles qu'elles avaient considérées comme partiellement prêtes ont obtenu des résultats plus mitigés (ex. des routines ne sont pas établies, le travail anti-harcèlement n'est pas intégré à l'organisation générale de l'école...).

Les chercheuses ont également constaté que le leadership était un facteur important, avec deux dimensions qui semblent particulièrement liées à une implantation réussie. Tout d'abord, il faut que le ou la directeur·trice ait un plan et le suive en intégrant le travail de lutte contre le harcèlement à la vie quotidienne de l'école. Ensuite, il faut qu'il ou elle possède la capacité d'inspirer le personnel afin que tou·te·s s'engagent à suivre le plan. Cela implique que le personnel reconnaisse que le ou la chef·fe d'établissement détient le leadership. Dans les trois écoles qui ont mené à bien la mise en œuvre du programme, les deux dimensions étaient présentes. En revanche, les trois écoles qui ont eu des résultats plus mitigés avaient des problèmes au niveau du leadership, que ce soit lié à l'une ou l'autre dimension.

Deux autres facteurs que les auteures ont identifiés comme étant associé à une implantation réussie sont la compatibilité entre le programme et le travail déjà effectué dans les écoles et l'énergie nécessaire pour implanter le programme (les deux étant liées vu que lorsque le programme est compatible avec ce qui se fait déjà, son implantation demande moins d'énergie).

Implantation de Zero en Irlande

Une étude à grande échelle réalisée en Irlande en 1997 a révélé que le harcèlement était un problème répandu. 31,3 % des élèves de primaires et 15,6 % des élèves de secondaire rapportent avoir été victimes de harcèlement au cours du dernier trimestre tandis que 26,5 % des élèves de primaire et 14,9 % des élèves de secondaire rapportent avoir harcelé d'autres élèves (O'Moore et al, 1997; cité par O'Moore & Minton, 2005). Il a donc à l'époque semblé important que l'Irlande dispose d'un programme anti-harcèlement implanté dans tout le pays (O'Moore & Minton, 2005).

O'Moore & Minton (2005), chercheur·euse·s au Trinity College, ont implanté un programme basé sur l'ancêtre de Zero, qui avait été implanté dans des écoles norvégiennes en 1996 (Roland & Munthe, 1997).

Quatre éléments clés du programme ont été réutilisés (O'Moore & Minton, 2005) :

1. Les écoles étaient accompagnées par une aide extérieure. 11 enseignant·e·s ont été formé·e·s lors de séminaires et workshops durant un total de 12 jours. Leur formation portait sur la définition du harcèlement, les profils d'élèves concerné·e·s, les signes révélateurs de victimisation et de harcèlement, les conséquences négatives du harcèlement, la lutte anti-harcèlement au niveau de l'école dans son entièreté, les stratégies d'interventions au niveau de la classe et au niveau individuel et la façon de gérer les parents des élèves concerné·e·s par le harcèlement. Des techniques de présentation étaient également enseignées. Les personnes ainsi formées étaient responsables de 3 à 5 écoles chacun·e : ils et elles formaient les enseignant·e·s, informaient les parents et assistaient les écoles.
2. Un pack contenant des informations sur le harcèlement a été donné aux formateur·trice·s afin qu'ils et elles puissent l'utiliser lors des formations aux enseignant·e·s et de l'assistance aux écoles
3. Un dépliant produit par l'*Anti-Bullying Research and Ressource Centre* du Trinity College a cette occasion a été distribué à tous les parents. Il contient des informations sur la prévalence, les types, les effets et les indicateurs de harcèlement.
4. Les élèves ont eu accès à des manuels adaptés à leur âge comprenant des idées pour lutter contre le harcèlement dans leurs classes et écoles.

Qualité de l'évaluation de Zero en Irlande

Article principal : O'Moore, A. M., & Minton, S. J. (2005). Evaluation of the effectiveness of an anti-bullying program in primary schools. *Aggressive Behavior*, 31(6), 609-622. doi:10.1002/ab.20098

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude longitudinale sans groupe contrôle. Le programme a été implanté lors de l'année 1999-2000. Le pré-test a été réalisé au cours des deux premiers trimestres de l'année 1998-1999 et le post-test lors de l'été 2000. Au niveau des analyses statistiques, des chi-carré ont été réalisés afin de voir si la proportion d'élèves ayant choisi l'une ou l'autre réponse aux items différait entre le pré et le post-test.

2. Quelles étaient les caractéristiques de l'échantillon ?

Les 100 écoles primaires du premier district d'un comté d'Irlande ont été invitées à participer à l'étude. 42 écoles ont accepté de participer, mais les données pré et post-test sont uniquement disponibles pour 22 écoles. Ces écoles comptaient entre 21 et 280 élèves (avec une seule école comptant plus de 200 élèves) et peuvent donc être considérées comme petites. Le nombre d'élèves par classe variait entre 3 et 28 et, mis à part une école située dans une petite ville, les écoles étaient toutes situées dans un village.

126 enseignant·e·s ont participé au pré-test et 83 au post-test. Les deux échantillons ne se distinguaient pas en termes de genre, d'âge, d'expérience d'enseignement et de poste occupé.

Aucune information détaillée sur la présence de données manquantes et la façon dont elles ont été gérées n'est rapportée dans l'article.

3. Par qui et où l'évaluation a-t-elle été menée ?

L'évaluation a été menée en Irlande par Astrid Mona O'Moore et Stephen James Minton du Trinity College (qui n'ont pas de lien avec l'University of Stavanger).

4. Quelles et combien de mesures ont été prises ?

Les élèves ont répondu au questionnaire anonyme durant les heures scolaires en étant encadré par leurs enseignant·e·s. Toutes les classes d'une même école passaient le questionnaire en même temps.

Afin de mesurer le harcèlement et la victimisation, les élèves ont répondu à une version modifiée du *OBVQ* (Olweus, 1983 ; cité par O'Moore & Minton, 2005) non détaillée dans l'article. Le questionnaire débutait par une définition du harcèlement suivie de questions portant sur le harcèlement subi et effectué au cours du dernier trimestre ainsi qu'au cours des cinq derniers jours. Des items portaient également sur les réactions des enseignant·e·s, de l'élève même et des autres élèves face au harcèlement.

Le questionnaire à destination des enseignant·e·s, le *Rigby Questionnaire* (Rigby, 1997 ; cité par O'Moore & Minton, 2005) débute également par une définition du harcèlement et comprend les items suivants : « *In your view, is your school a safe place for young people who find it hard to defend themselves from attacks by other students ?* », « *Do you think that teachers at your school are interested in trying to stop bullying?* », « *Do you personally try to stop bullying when you see it happening?* », « *Do you agree that stopping bullying is a matter for all staff?* » et « *Do you agree that more rigorous monitoring by staff of school bullying is needed?* ».

Effets de Zero selon l'évaluation en Irlande

1. Des effets ont-ils été observés ?

Les élèves rapportent significativement moins de victimisation au post-test qu'au pré-test (que ce soit au cours du dernier trimestre ou des cinq derniers jours). Au niveau du harcèlement, la réduction entre le pré et le post-test n'est pas significative pour le harcèlement rapporté au cours du dernier trimestre, mais elle l'est pour le harcèlement rapporté au cours des derniers jours (voir tableau 2).

Tableau 2. Evolution des taux de victimisation et harcèlement.

	Temps 1	Temps 2	Pourcentage de changement
Victimisation			
Au moins une fois au cours du dernier trimestre	36,7 %	29,5 %	-19,6 %
Fréquemment au cours du dernier trimestre	7,3 %	3,6 %	-50 %
Au moins une fois au cours des 5 derniers jours	22,1 %	12,6 %	-43 %
Harcèlement			
Au moins une fois au cours du dernier trimestre	27,1 %	22,4 %	-17,3 % (n.s.)
Au moins une fois au cours des 5 derniers jours	13,7 %	6,6 %	-48,2 %

Inspiré de O'Moore & Minton (2005, p. 615-617). Les changements significatifs sont en gras.

La proportion d'élèves qui perçoivent que les enseignant·e·s et pairs font quelque chose pour mettre fin au harcèlement n'est pas significativement différente entre le pré et le post-test, tout comme les réponses des enseignant·e·s à leur questionnaire. Par contre, il y a un changement significatif au niveau de la réaction auto-rapportée des élèves face au harcèlement. Alors qu'au pré-test 61,8 % des élèves répondent qu'ils ou elles essaieraient d'aider la victime face à une situation de harcèlement et 15,9 % qu'ils ou elles ne feraient rien, les proportions sont de 71,4 % et 9,2 % au post-test.

En conclusion, une réduction entre le pré et le post-test est observée concernant le harcèlement et la victimisation, mais pas concernant les autres variables (exceptée la réaction face au harcèlement). Comme il n'y a pas de groupe contrôle dans l'étude, il n'est pas possible de savoir si ces réductions sont dues à l'effet du programme. Des chi-carrés ont été réalisés et montrent que les taux de harcèlement ne diffèrent pas entre les élèves de 3^{ème}, 4^{ème} et 5^{ème} année, que ce soit au pré ou au post-test. Les auteurs en concluent donc que la réduction observée n'est pas due à un effet de l'âge.

2. Les effets sont-ils différenciés ?

Les auteur·e·s n'ont pas trouvé d'effet du niveau scolaire et, mis à part cela, ne rapportent aucune analyse permettant de savoir si les effets sont différenciés.

3. Des mesures de suivi ont-elles été prises ?

Aucune mesure de suivi n'a été prise.

4. L'implantation a-t-elle été mesurée ?

Aucune mesure de l'implantation n'est rapportée dans l'article.

Résumé des effets de *Zero*

Le programme a été implanté à d'autres occasions en Norvège. En 1996, l'ancêtre de *Zero* a été implanté à grande échelle et, selon Roland & Munthe (1997), les effets du programme étaient positifs. Toutefois, aucune donnée quantitative sur l'efficacité du programme en 1996-1997 ne semble avoir été publiée. Midthassel et Roland (2011; cité par Roland & Midthassel, 2012) ont étudié les effets de *Zero* dans six écoles primaires d'une communauté norvégienne et rapportent que la prévalence d'élèves impliqué·e·s dans des situations de harcèlement au moins une fois par semaine, que ce soit en tant que victime ou auteur·e, a diminué entre 40 et 70 %. Ces résultats ont uniquement été publiés dans le chapitre d'un livre en norvégien.

Dans l'étude pilote, portant uniquement sur l'aspect de gestion de la classe du programme, des résultats positifs ont été observés. Concernant les deux études réalisées à plus grande échelle, les résultats vont plutôt dans le sens d'une absence d'effet du programme en Norvège et d'un effet du programme en Irlande. Toutefois, il faut prendre en compte le fait que l'étude norvégienne a été réalisée dans un contexte de mobilisation nationale contre le harcèlement et que des programmes anti-harcèlement (y compris *Zero*) étaient également implantés dans les écoles contrôles. Il n'est donc pas possible de conclure suite à ces résultats que *Zero* est inefficace. En fait, étant donné les limites méthodologiques importantes des deux études, il est difficile de conclure quoi que ce soit. Aucune des deux études ne comprend de réel groupe contrôle : en Irlande il n'y a pas du tout de groupe contrôle et, en Norvège, les données du groupe expérimental sont comparées à des données obtenues sur un autre intervalle de temps et aucune information n'est disponible concernant les programmes anti-harcèlement utilisés dans les écoles du groupe qui sert de contrôle. De plus, l'étude irlandaise présente d'autres limites méthodologiques : les données pré et post-test n'ont pas été prises à la même période de l'année et le taux d'attrition est très important (les données pré et post-test sont uniquement disponibles pour 22 des 42 écoles participantes).

En conclusion, bien que l'implantation de programme *Zero* soit systématiquement associée à une diminution des taux de harcèlement et de victimisation, que ce soit en Norvège ou en Irlande, il n'est pas possible de conclure que les effets observés sont bien dus au programme.

Annexe 4 : Evaluation de ViSC

ViSC a tout d'abord été évalué en Autriche et en Allemagne puis également à Chypre, en Roumanie et en Turquie. Dans la suite du chapitre, les évaluations du programme seront présentées les unes à la suite des autres, pays par pays.

Implantation de ViSC en Autriche

L'implantation du programme ViSC en Autriche fait partie d'une stratégie nationale pour lutter contre la violence dans les écoles publiques (Spiel & Strohmeier, 2011). Depuis la fin des années 90, le thème de la violence scolaire attire beaucoup l'attention publique en Autriche. Par conséquent, le gouvernement a déclaré son intention de mettre en place des initiatives pour prévenir la violence (Spiel et al., 2016) et, depuis 1997, il fournit un soutien financier à certains projets d'intervention et de prévention en lien avec la violence dans les écoles. Toutefois l'efficacité de ces projets est très rarement évaluée. En 2003, un rapport d'Atria et Spiel (2003, cité par Strohmeier, Hoffmann, Schiller, Stefanek, & Spiel, 2012) rapporte que les initiatives réalisées par les enseignant·e·s pour prévenir la violence ne sont ni basées sur des preuves, ni efficaces. En 2007, suite à une série d'événements spectaculaires dans les écoles et à une discussion publique sur les taux élevés de harcèlement et de victimisation, le ministère de l'éducation autrichien a mandaté des chercheuses de l'University of Vienna pour qu'elles développent une stratégie nationale de prévention de la violence dans le système scolaire publique (Spiel & Strohmeier, 2011). Le but de cette stratégie est de développer une procédure qui fait de la prévention contre le harcèlement un but commun pour le plus grand nombre d'acteurs possibles, qui tire des apprentissages de la recherche et des exemples internationaux et qui implante des programmes basés sur des preuves dans les écoles. Un des six domaines d'activités de cette stratégie concerne la prévention et l'intervention dans les écoles. Parmi les activités de ce domaine, on retrouve le soutien envers le programme de prévention de la violence *Faustlos* (basé sur le programme *Second Step*) pour les écoles primaires et maternelles et l'implantation du programme ViSC dans les écoles secondaires. Un autre des domaines est l'évaluation et la recherche, avec comme activité prévue l'évaluation de l'implantation du programme ViSC et de ses effets à court et long terme (Spiel & Strohmeier, 2011). En 2008, la stratégie nationale est devenue une partie de l'accord de coalition entre les membres du gouvernement et il a été décidé qu'elle durerait jusqu'à la fin de la période de législation en septembre 2013. Par conséquent, pendant cette période, un budget a été dévoué à la stratégie nationale ce qui a permis de financer l'implantation et l'évaluation de ViSC. L'implantation de la stratégie nationale a continué durant la période législative allant de 2013 à 2018 (Spiel et al., 2016).

Lorsqu'il a été intégré à la stratégie nationale, le programme ViSC avait déjà été évalué une fois en Autriche et deux fois en Allemagne. Sur base des résultats de ces études, le projet de classe ViSC a été amélioré. Le matériel et les feuilles d'exercices disponibles dans le manuel ont été changés pour être plus facilement compris et appliqués par les enseignant·e·s et des suggestions concrètes pour des activités à base de jeux ont été ajoutées. Afin de mieux correspondre aux besoins des élèves et à l'organisation scolaire, la phase d'impulsion est passée de 6 à 8 leçons tandis que la phase d'action est passée de 6 à 4 leçons (Spiel & Strohmeier, 2011). Du plus, le programme s'est enrichi d'éléments au niveau de l'école dans son entièreté et l'âge cible des élèves a été modifié.

Qualité de l'évaluation de 2000-2001

Article principal : Atria, M., & Spiel, C. (2007). The Viennese Social Competence (ViSC) training for students: Program and evaluation. In & C. A. M. J. E. Zins, M. J. Elias (Ed.), *Bullying, victimization and peer harassment: A handbook of prevention and intervention* (pp. 179–197). New York : The Haworth Press.

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude longitudinale avec un design RCT. Les classes ont été aléatoirement assignées à la condition expérimentale ou contrôle. Le programme a été implanté de novembre 2000 à juin 2001.

Pour chaque période de mesure (juin 2000, juin 2001 et octobre 2001), le groupe contrôle et le groupe expérimental ont été comparés via des MANOVA. Les données pré et post-test ont donc été analysées séparément.

2. Quelles étaient les caractéristiques de l'échantillon ?

Les données pré-test ont été récoltées en juin et en octobre 2000 et les données post-test en juin et en octobre 2001. Deux classes d'une école secondaire supérieur ont participé au programme ViSC et deux autres classes de la même école ont servi de groupe contrôle. L'échantillon de l'étude est composé de 112 élèves (57 garçons et 55 filles) de 9^{ème} et 10^{ème} année. Il y avait 55 élèves dans la condition expérimentale et 57 dans la condition contrôle. Les élèves étaient âgé·e·s de 15 à 21 ans (avec un âge médian de 17 ans). Pour 47 % d'entre eux et elles l'allemand n'était pas la langue maternelle, 20 % avaient un père n'ayant pas eu de diplôme de secondaire et 68 % une mère qui était femme au foyer. Il n'y avait aucune différence en termes de genre, d'âge et de composition familiale entre le groupe contrôle et le groupe expérimental. Il n'y avait pas non plus de différence entre les groupes sur les variables d'intérêt au pré-test.

Aucune information sur la présence de données manquantes ou sur la façon dont elles ont été traitées n'est mentionnée.

3. Par qui et où l'évaluation a-t-elle été menée ?

L'évaluation est menée en Autriche par Moira Atria et Christiane Spiel, de l'University of Vienna. Christiane Spiel est la principale développeuse du programme ViSC.

4. Quelles et combien de mesures ont été prises ?

Les données ont été collectées à l'aveugle par des étudiant·e·s de troisième cycle. Les mesures suivantes ont été prises :

- **Niveau de démocratie** : Le niveau de démocratie perçu a été mesuré en utilisant un questionnaire de démocratie basé sur une version de Eder (1998; cité par Atria & Spiel, 2007) composé de quatre sous-échelles : égalité, information, participation et diversité d'opinion.

Annexe 4 – Evaluation de ViSC

- **Niveau d'agression** : Le niveau d'agression perçu a été mesuré en utilisant une version modifiée de l'*OBVQ* (Olweus, 1983; cité par Atria & Spiel, 2007) incluant quatre sous-échelles : harcèlement envers les pairs, victimisation subie par les pairs, agression envers les enseignant·e·s et victimisation subie par les enseignant·e·s.

Effets de ViSC selon l'évaluation de 2000-2001

1. Des effets ont-ils été observés ?

Juste après l'intervention, une différence significative entre le groupe contrôle et le groupe expérimental est observée pour le niveau de démocratie perçu : les élèves ayant eu l'intervention perçoivent un niveau plus élevé de démocratie que ceux et celles ne l'ayant pas eu. Cette différence n'est toutefois plus présente quatre mois après la fin du programme. Concernant l'agression, que ce soit en juin ou en octobre 2001, l'effet du groupe n'est pas significatif même si, au niveau descriptif, les moyennes vont dans le sens des hypothèses.

2. Les effets sont-ils différenciés ?

Les auteures ne rapportent aucune analyse permettant de savoir si les effets sont différenciés.

3. Des mesures de suivi ont-elles été prises ?

Une mesure de suivi a été prise en octobre 2001. L'effet de l'intervention n'est alors plus significatif.

4. L'implantation a-t-elle été mesurée ?

Aucune analyse statistique sur l'implantation n'a été réalisée. Toutefois des données qualitatives sont rapportées. La formatrice qui a dispensé les leçons ViSC dans les deux classes devait documenter chaque leçon. Elle rapporte que, lors de la première leçon, les élèves étaient principalement curieux·euses, mais certain·e·s se sont montrés sceptiques et, au cours des deux leçons suivantes, ont essayé de la « tester ». Toujours d'après elle, suite à sa réaction adéquate, les élèves ont coopéré de façon importante et se sont fortement impliqué·e·s dans les leçons. À partir de la 4^{ème} leçon, ils et elles se sont mis à aborder des problèmes personnels et à demander de les travailler. Lors de la phase de réflexion, les élèves ont rapporté un changement de perspective suite aux leçons, avec une capacité accrue à relativiser leurs propres intérêts et à voir la situation de point de vue de quelqu'un d'autre. Une limite à l'implantation du programme était le taux élevé d'absentéisme scolaire parmi ces classes. Toutefois, les élèves étaient fortement motivé·e·s à participer aux leçons ViSC, loupant parfois les cours habituels dispensés juste avant et après les leçons tout en étant présent·e·s pour celles-ci.

Qualité de l'évaluation de 2009-2010

Article principal : Yanagida, T., Strohmeier, D., & Spiel, C. (2016). Dynamic Change of Aggressive Behavior and Victimization Among Adolescents: Effectiveness of the ViSC Program. *Journal of Clinical Child & Adolescent Psychology*, 1–15. doi:10.1080/15374416.2016.1233498

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude longitudinale avec un design RCT. Toutefois, un nombre très important d'écoles ont refusé de participer après avoir appris qu'elles avaient été assignées à la condition contrôle ce qui fausse le design de l'étude. Le programme a été implanté de septembre 2009 à juin 2010.

Dans leur analyse, les auteur·e·s tiennent compte des changements dynamiques. Cela veut dire que la covariation entre les scores de changement de comportement agressif et de victimisation ainsi que l'impact des scores pré-test de la même variable (ex. score pré-test de la victimisation sur le changement dans la victimisation) et de l'autre variable (ex. score pré-test de l'agression sur le changement dans la victimisation) sont pris en compte. Une approche de modélisation par équation structurelle a été utilisée pour tester les hypothèses de l'étude. Différents modèles ont été utilisés afin de comparer les deux groupes d'abord indépendamment du genre puis en tenant compte de celui-ci.

Dans l'étude de Gradinger et al. (2015) portant sur le cyberharcèlement et la cybervictimisation, trois modèles ont été utilisés afin de tenir compte des formes traditionnelles d'agression et de victimisation ainsi que de l'âge et du genre.

2. Quelles étaient les caractéristiques de l'échantillon ?

En décembre 2008, toutes les écoles secondaires situées à Vienne ont été invitées à porter leur candidature pour l'implantation de ViSC. Sur les 155 écoles contactées, 34 ont demandé à participer à l'étude et 26 d'entre elles remplissaient les critères de participation (ex. accepter de prendre part à l'évaluation). Parmi ces 26 écoles, 13 ont été aléatoirement sélectionnées pour implanter le programme et 13 ont été assignées à la condition contrôle. Sur ces 13 écoles, seules 5 ont accepté de participer en tant qu'écoles contrôles (les 8 autres se sont retirées de l'étude). 2042 élèves (1377 dans la condition expérimentale et 665 dans la condition contrôle) ont participé à au moins un moment de mesure et composent donc l'échantillon final.

Au temps 1 (mai 2009), les élèves étaient en 5^{ème}, 6^{ème}, ou 7^{ème} année et avaient 11,7 ans en moyenne. L'échantillon comporte 1639 élèves (47,6 % de filles) venant de 103 classes (50 classes de 5^{ème} année, 51 de 6^{ème} et 2 de 7^{ème}). 46,4 % des élèves sont d'origine autrichienne, 14,3 % viennent des pays de l'ex-Yougoslavie, 14,3 % de Turquie et 19,1 % d'autres pays. Concernant le statut socio-économique, 9,4 % des élèves ont indiqué que leur famille avait moins d'argent que les autres familles, 71,7 % autant et 18,9 % plus. 65,2 % ont indiqué utiliser des téléphones portables et 37,5 % utiliser internet plusieurs fois par jour. Le groupe contrôle et le groupe expérimental ne se distinguaient sur aucune mesure sociodémographique ni sur aucune variable d'intérêt au pré-test. Le taux de consentement parental était supérieur à 90 % dans toutes les écoles. Au total, les données ont été collectées pour 71,1 % des élèves (consentant à participer et présent·e·s le jour de l'étude).

Annexe 4 – Evaluation de ViSC

La composition des classes est restée très constante du temps 1 au temps 2. 515 élèves ont uniquement participé au pré-test et 403 uniquement au post-test. Aucune différence significative n'a été trouvée entre ces élèves et ceux et celles qui ont participé aux deux mesures. Une procédure d'imputation multiple sous le postulat MAR (*missing at random*) a été utilisée pour gérer les données manquantes.

Pour évaluer la durabilité du programme, dans l'étude de Gradinger, Yanagida, Strohmeier, & Spiel (2016), des données ont été prises dans six des écoles ayant participé à l'étude. Les analyses ont donc été réalisées sur 659 élèves (319 dans la condition expérimentale et 340 dans la condition contrôle). Les données de l'étude de Schultes et al. (2014) portent sur les 13 écoles de la condition expérimentale ainsi que sur 7 écoles ayant implanté le programme l'année précédente. Sur les 708 enseignant·e·s ayant participé·e·s au programme, 370 ont accepté de participer à l'étude. 77,3 % de ces enseignant·e·s sont des femmes, 70,1 % ont plus de 40 ans et 48,2 % ont plus de 20 ans d'expérience. 125 enseignant·e·s ont répondu aux deux mesures, 126 seulement au pré-test et 119 seulement au post-test.

3. Par qui et où l'évaluation a-t-elle été menée ?

L'évaluation est menée en Autriche par Christiane Spiel, Christine Hoffmann et Marie-Therese Schultes (University of Vienna), Petra Gradinger, Takuya Yanagida, Elisabeth Stefanek et Dagmar Strohmeier (University of Applied Sciences Upper Austria), Eva-Maria Schiller (University of Münster) et Rens van de Schoot (Utrecht University). L'équipe du projet ViSC est composée de Christophe Burger, Katharina Derndarsky, Petra Gardinger, Christine Hoffman, Bianca Pollhammer, Eva-Maria Schiller, Marie-Therese Schultes, Christiane Spiel, Elisabeth Stefanek, Dagmar Strohmeier et Takuya Yanagida. Les personnes qui ont développé le programme et l'ont évalué sont donc essentiellement les mêmes.

4. Quelles et combien de mesures ont été prises ?

Les données ont été collectées en mai 2009 et mai 2010. Pour l'étude de Gradinger, Yanagida, Strohmeier, & Spiel (2016), des données ont également été prises en novembre/décembre 2010 (follow-up) dans trois écoles contrôles et trois écoles ayant eu le programme. Les élèves ont rempli les questionnaires en ligne, durant les heures scolaires. Ils et elles étaient supervisés par deux assistant·e·s de recherche qui leur précisait bien que leurs données resteraient confidentielles. Pour éviter tout effet d'ordre, les items et échelles ont été contrebalancés entre les participants.

Pour toutes les mesures, les élèves ont répondu sur une échelle de 5 points (0=*Pas du tout*, 1=*Une ou deux fois*, 2=*Deux ou trois fois par mois*, 3=*Une fois par semaine* et 4=*Presque tous les jours*). Les mesures suivantes ont été prises :

- **Harcèlement** : Le harcèlement a été mesuré via des items développés pour l'étude PISA 2009 en Autriche (Strohmeier et al., 2012a; cité par Yanagida et al., 2016). L'échelle de harcèlement auto-rapportée est composée d'un item global («*How often have you insulted or hurt other student during the last 2 months ?*») et de trois items spécifiques décrivant des comportements de harcèlement verbal, physique et relationnel. Les alphas sont égaux à .82 (au pré-test) et .83 (au post-test).

- **Victimisation due au harcèlement** : L'échelle de victimisation était la même que celle de harcèlement, mais avec les items formulés de manière passive (ex. « *How often have others insulted or hurt you in the last 2 months ?* »). Les alphas sont égaux à .81 (au pré-test) et .82 (au post-test).
- **Agression physique** : L'agression physique a été mesurée par une version modifiée de l'échelle de Crick et Grotpeter (1995 ; cité par Yanagida et al., 2016) afin qu'elle corresponde à une mesure auto-rapportée. L'échelle comprend trois items (ex. « *How often have you hit one or more classmates during the last 2 months ?* »). Les alphas sont égaux à .79 (au pré comme au post-test).
- **Victimisation physique** : L'échelle de victimisation physique était la même que celle d'agression physique, mais avec les items formulés de manière passive (ex. « *How often have you been hit by one or more classmates during the last 2 months ?* »). Les alphas sont égaux à .74 (au pré-test) et .76 (au post-test).
- **Agression relationnelle** : L'agression relationnelle a été mesurée via 5 items développés par Crick et Grotpeter (1995 ; cité par Yanagida et al., 2016) adaptés afin de correspondre à une mesure auto-rapportée (ex. « *Some kids leave other kids out on purpose when it's time to play or do an activity. How often have you done that during the last 2 months?* »). Les alphas sont égaux à .83 (au pré-test) et .87 (au post-test).
- **Victimisation relationnelle** : L'échelle de victimisation relationnelle était la même que celle d'agression relationnelle, mais avec les items formulés de manière passive (ex. « *How often during the last 2 months have you been excluded from play or another activity by one or more classmates ?* »). Les alphas sont égaux à .82 (au pré-test) et .81 (au post-test).
- **Cyberharcèlement** : Le cyberharcèlement a été mesuré avec une échelle auto-rapportée de sept items inspirés de Smith et al. (2008 ; cité par Gradinger et al., 2015). Les items portent sur différentes formes de cyberharcèlement via appels, SMS, chats, forums de discussion, messages instantanés, photos et vidéos (ex. « *How often have you insulted or hurt students by mean calls during the last two months?* »). Les alphas sont égaux à .95 (au pré-test) et .96 (au post-test).
- **Cybervictimisation** : L'échelle de cybervictimisation était la même que celle de cyberharcèlement, mais avec les items formulés de manière passive (ex. « *How often have you been insulted or hurt by receiving mean calls during the last two months from other students ?* »). Les alphas sont égaux à .92 (au pré-test) et .95 (au post-test).
- **Usage d'internet** : L'usage d'internet a été mesuré via l'item « *How often do you use the internet ? (E-mails, chat, discussion board, instant message...)* ». Les élèves ont répondu sur une échelle à 6 points (0=*Jamais*, 1=*Moins d'une fois par semaine*, 2=*Une fois par semaine*, 3=*Tous les deux ou trois jours*, 4=*Une fois par jour* et 5=*Plusieurs fois par jour*).
- **Climat de classe** : Le climat de classe a été mesuré via trois items développés par Eder et Mayr (2000 ; cité par Gradinger et al., 2016) du type « *In our class all students work together well and help each other* » auxquels les élèves ont répondu sur une échelle à 4 points (1=*Pas du tout vrai*, 2=*En*

Annexe 4 – Evaluation de ViSC

quelque sorte vrai, 3=*Vrai*, et 4=*Certainement vrai*). Les alphas sont égaux .85 (pour le pré-test) et .86 (pour le post-test).

- **Diversité ethnique** : La diversité ethnique a été calculée en utilisant la formule développée par Simpson (1949; cité par Gradinger et al., 2016). Le calcul donne un score compris entre 0 (tous les élèves viennent du même groupe culturel) et 1 (aucun·e élève ne vient du même groupe culturel).

Les différentes mesures ont été intégrées dans des modèles représentant la relation hiérarchique entre les dimensions investiguées. Selon ces modèles, l'agression (qui reprend le harcèlement, l'agression physique et l'agression relationnelle) et la victimisation (qui reprend la victimisation due au harcèlement, la victimisation physique et la victimisation relationnelle) sont des facteurs de plus haut ordre. Ces modèles ont été testés et les résultats n'ont montré aucune diminution significative d'adéquation entre les modèles hiérarchiques imbriqués pour l'agression et la victimisation. Ils ont donc été utilisés. Seuls le cyberharcèlement et la cybervictimisation ont été traités à part.

Les données concernant la fidélité d'implantation et la réceptivité au niveau de l'école ont été collectées auprès des coachs ViSC après chaque session de formation. Les mesures prises sont les suivantes :

- **Fidélité d'implantation** : La fidélité d'implantation a été mesurée en demandant aux coachs le nombre d'unités de formation dispensées à toute l'école et aux enseignant·e·s qui implantent ViSC dans leur classe. Les scores des écoles varient entre 0,4 (40 % des unités prescrites ont été données) et 2,0 (deux fois plus d'unités que prescrit ont été données).
- **Réceptivité au niveau de l'école** : La réceptivité des enseignant·e·s au niveau de l'école a été mesurée en comparant le nombre d'enseignant·e·s présent·e·s au nombre d'enseignant·e·s supposé·e·s être présent·e·s. Les taux de participation varient entre 30 et 100 %.

Les données concernant les effets du programme au niveau des enseignant·e·s et leur réceptivité au niveau individuel ont été récoltées par questionnaire au pré et au post-test. Mis à part les données démographiques (genre, âge, expérience d'enseignement, nom de l'école), les mesures prises au niveau des enseignant·e·s sont les suivantes :

- **Efficacité personnelle à stopper la violence** : L'efficacité personnelle à stopper la violence a été mesurée via un item unique. Les enseignant·e·s ont dû évaluer leur capacité à stopper la violence entre élèves sur une échelle à 5 points allant de *Très bonne* à *Insuffisante*. La validité de cet item avait été vérifiée dans une étude pilote.
- **Comportement dans les situations de harcèlement** : Le comportement dans les situations de harcèlement a été mesuré via une traduction du *Handling Bullying Questionnaire* (Bauman et al., 2008; cité par Schultes et al., 2014). Les enseignant·e·s devaient lire une situation décrivant un cas de harcèlement puis répondre à 22 items concernant les stratégies utilisées pour gérer la situation. Le questionnaire comprend 5 sous-échelles illustrant 5 réactions possibles : travailler avec l'élève victime de harcèlement, travailler avec l'élève auteur·e de harcèlement, ignorer la situation, enrôler

Annexe 4 – Evaluation de ViSC

d'autres enseignant·e·s et punir l'élève auteur·e de harcèlement. Comme la sous-échelle « *travailler avec la victime* » (ex. « *I would tell the victim to stand up to the bully.* ») correspond le mieux au programme ViSC, c'est elle qui a été utilisée comme indicateur du comportement de l'enseignant·e dans les situations de harcèlement.

- **Réceptivité au niveau individuel :** La réceptivité des enseignant·e·s au niveau individuel a été mesurée en leur demandant s'ils ou elles ont fait partie de l'équipe ViSC qui gère les situations de harcèlement et s'ils ou elles ont implanté le projet de classe ViSC dans leur classe.

Effets de ViSC selon l'évaluation de 2009-2010

1. Des effets ont-ils été observés ?

Pour rappel, mis à part pour le cyberharcèlement et la cybervictimisation, les analyses ont été réalisées sur les facteurs de second ordre que sont la victimisation et l'agression. Le programme a un effet significatif sur la victimisation, mais pas sur l'agression. Selon les résultats portant sur l'ensemble des élèves, entre le pré et le post-test, le niveau d'agression ne change ni dans le groupe contrôle ni dans le groupe expérimental. Pour les filles, les scores d'agression augmentent dans le groupe contrôle et restent stables dans le groupe expérimental alors que pour les garçons ils diminuent dans le groupe expérimental et restent stables dans le groupe contrôle, mais, dans les deux cas, la différence entre les groupes n'est pas significative. Par contre, pour les filles comme pour les garçons, la victimisation diminue dans le groupe expérimental alors qu'elle reste constante dans le groupe contrôle. La différence entre les deux groupes est significative (voir tableau 1).

Concernant les changements dynamiques de l'agression et de la victimisation, les résultats montrent que l'évolution de l'agression est reliée au score d'agression pré-test (la diminution d'agression est plus importante pour les élèves les plus agressif·ve·s au départ), mais pas au score de victimisation pré-test. De même, l'évolution de la victimisation est reliée au score de victimisation pré-test (la diminution de victimisation est plus importante pour les élèves les plus victimisé·e·s au départ), mais pas au score d'agression pré-test. Au pré-test, l'agression et la victimisation sont corrélées. Les changements des niveaux d'agression et de victimisation le sont également.

Le programme a également un effet significatif sur le cyberharcèlement et la cybervictimisation (Grading et al., 2015). En prenant en compte l'ensemble des élèves, le cyberharcèlement augmente dans le groupe contrôle et diminue dans le groupe expérimental tandis que la cybervictimisation reste stable dans le groupe contrôle et diminue dans le groupe expérimental. Dans les deux cas, la différence entre les groupes est significative. Lorsque l'âge est pris en compte, les effets sont les mêmes (seule leur taille varie légèrement). Pour les filles prises séparément, les résultats sont également les mêmes. Par contre, pour les garçons, le cyberharcèlement n'augmente pas dans le groupe contrôle (mais la différence entre les deux groupes reste significative) et la différence des scores de cybervictimisation entre le groupe contrôle et le groupe expérimental n'est que marginalement significative (voir tableau 1).

Tableau 1. Taille des effets (d latent)

Filles	
Facteur d'agression	0.442 (n.s.)
Facteur de victimisation	0,775
Cyberharcèlement	0,301
Cybervictimisation	0,294
Garçons	
Facteur d'agression	0.128 (n.s.)
Facteur de victimisation	0,765
Cyberharcèlement	0,467
Cybervictimisation	0.288 (marginalelement significatif)
Total	
Facteur d'agression	0.185 (n.s.)
Facteur de victimisation	0,725
Cyberharcèlement	0,389
Cybervictimisation	0,295

Inspiré de Yanagida et al. (2016, p. 9) et de Gradinger et al., 2015 (p.100). Les effets significatifs sont en gras.

Gradinger et al. (2016) ont également analysé les effets du programme sur le cyberharcèlement et la cybervictimisation à plus long terme et en contrôlant l'effet de différentes covariables au niveau individuel (niveaux pré-test de cyberharcèlement, de cybervictimisation, d'agression traditionnelle et de victimisation traditionnelle, âge, genre et usage d'internet), mais également au niveau de la classe (climat de classe et diversité ethnique). Selon leur analyse, le programme est efficace : le cyberharcèlement diminue et la cybervictimisation reste stable dans le groupe expérimental alors qu'ils augmentent dans le groupe contrôle. Les changements moyens de cyberharcèlement et cybervictimisation sont associés à un certain nombre de facteurs (scores pré-test d'agression et de victimisation traditionnelle, de cyberharcèlement et de cybervictimisation, genre et âge), mais ces facteurs n'interagissent pas avec la condition, mis à part pour le genre (les changements de cyberharcèlement et de cybervictimisation dépendent du genre dans la condition contrôle, mais pas dans la condition expérimentale).

2. Les effets sont-ils différenciés ?

Le programme semble efficace pour réduire la victimisation, quels que soient l'âge et le genre des élèves. Par contre le niveau de victimisation pré-test a de l'importance : la diminution de la victimisation est plus importante pour les élèves qui présentaient des scores élevés au pré-test. Concernant le cyberharcèlement et la cybervictimisation, contrairement aux hypothèses de Gradinger et al. (2016), le climat de classe et la diversité ethnique ne modèrent pas l'effet du programme. Par contre le niveau de victimisation traditionnelle modère l'effet du programme sur le cyberharcèlement : dans le groupe expérimental, plus le niveau pré-test de victimisation traditionnelle est élevé, moins le cyberharcèlement diminue. Mais c'est la

Annexe 4 – Evaluation de ViSC

seule variable qui est modératrice au niveau individuel. Il semble donc que le programme soit efficace indépendamment des caractéristiques des élèves et des classes.

3. Des mesures de suivi ont-elles été prises ?

Des mesures de suivi sont uniquement rapportées dans l'étude de Gradinger et al. (2016). Elles ont été prises en novembre/décembre 2010 et concernent la cybervictimisation et le cyberharcèlement. Les résultats montrent une durabilité des effets du programme : dans les écoles ayant eu l'intervention, le cyberharcèlement et la cybervictimisation continuent de diminuer 6 mois plus tard (alors que la cybervictimisation reste stable et que le cyberharcèlement augmente dans le groupe contrôle).

4. L'implantation a-t-elle été mesurée ?

L'article de Schultes et al. (2014) est entièrement consacrée à l'implantation du programme, plus particulièrement à l'influence de l'implantation du programme au niveau des enseignant·e·s. Les résultats des analyses montrent que l'efficacité personnelle des enseignant·e·s a augmenté suite au programme, et ce d'autant plus dans les écoles où un nombre important d'enseignant·e·s ont assisté aux formations leur étant destinées. Le programme a également eu un impact sur les comportements des enseignant·e·s dans les situations de harcèlement : au post-test, les enseignant·e·s encouragent davantage les victimes, mais uniquement pour ceux et celles qui ont implanté le projet de classe ViSC dans leur classe (et davantage pour ceux et celles qui le faisaient peu avant l'implantation du programme).

Implantation de ViSC en Allemagne

Le programme a été implanté pour la première fois en Allemagne au début des années 2000. Les résultats concernant l'efficacité de cette implantation sont rapportés dans un article en allemand (Gollwitzer, 2005 ; cité par Gollwitzer, Eisenbach, Atria, Strohmeier, & Banse, 2006). Selon les résultats de cette étude à très petite échelle, les niveaux d'agression ont diminué de manière significative entre le pré et le post-test, mais uniquement dans une des trois classes ayant implantées le programme (Gollwitzer, 2005 ; cité par Gollwitzer et al., 2006).

Une évaluation plus contrôlée du programme a été réalisée en 2004 (Gollwitzer, Banse, Eisenbach, & Naumann, 2007 ; Gollwitzer et al., 2006). Le programme n'avait donc pas encore été modifié pour être intégré à la stratégie nationale autrichienne. Chaque leçon a été dispensée par un duo de formateur·trice·s (un duo s'occupant des classes de l'école A et un autre de celle de l'école B). Les formateur·trice·s ont participé à un workshop et ont été supervisés par les conceptrices de ViSC à l'University of Vienna ainsi que par le premier auteur de l'étude.

Qualité de l'évaluation en Allemagne

Article principal : Gollwitzer, M., Eisenbach, K., Atria, M., Strohmeier, D., & Banse, R. (2006). Evaluation of aggression-reducing effects of the "Viennese Social Competence Training." *Swiss Journal of Psychology*, 65(2), 125-135. doi:10.1024/1421-0185.65.2.125

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une procédure quasi-expérimentale. Les classes n'ont pas été réparties aléatoirement dans les conditions, ce sont les membres de la direction et les enseignant·e·s qui ont choisi les classes dans lesquelles le programme serait implanté et qui ont désigné des classes qui pourraient servir de condition contrôle. Les classes de la condition contrôle ont été désignées sur base de leur ressemblance maximale avec celles de la condition expérimentale en termes de climat de classe et de niveau d'agression. Le pré-test a eu lieu trois semaines avant le début de l'implantation du programme (en mars 2004) et le post-test juste après (en juillet 2004). Un follow-up a eu lieu en novembre 2004 (environ quatre mois après la fin du programme).

Afin d'analyser les données, des ANOVAS univariées ou multivariées tenant compte de la condition et du temps de mesure ont été réalisées. Pour les données ne se distribuant pas normalement, des tests non paramétriques ont été utilisés. Dans l'étude de Gollwitzer et al. (2007), des modèles linéaires hiérarchiques ont été utilisés pour tester l'effet du programme sur la tâche IAT.

2. Quelles étaient les caractéristiques de l'échantillon ?

Le programme a été implanté dans deux écoles secondaires allemandes situées dans des zones rurales. L'auteur principal de l'étude a été contacté par les directions des deux écoles qui étaient au courant que l'University of Koblenz-Landau organisait des activités de prévention de la violence. Dans l'école A, le programme a été implanté dans deux classes : A1 (6ème année) et A2 (8ème année). Dans l'école B, il a été

Annexe 4 – Evaluation de ViSC

implanté dans deux classes de 6^{ème} année (B1 et B2). Deux classes ont été sélectionnées comme contrôles pour l'école A (A3, classe de 6^{ème} année et A4, classe de 8^{ème}). Pour l'école B, une seule classe, B3, a servi de classe contrôle étant donné qu'il s'agissait de la seule classe restante de 6^{ème} année.

Les élèves et parents ont donné leur consentement pour participer à l'étude. Le taux de consentement était de 100 %. Le nombre total d'élèves ayant participé au pré-test est de 184. Les élèves présentant des données manquantes ou des données aberrantes sur plus de deux variables dépendantes ont été retirés de l'étude. L'échantillon final est composé de 149 élèves. Des analyses ont révélé que les élèves ayant des données manquantes présentaient des niveaux plus élevés d'agression rapportée par les pairs au pré-test que les autres élèves. Les avoir ôté-e-s des analyses réduit donc la généralisabilité des résultats.

L'étude de Gollwitzer, Banse, Eisenbach, & Naumann (2007) porte sur les données de cette étude ainsi que sur celle d'une étude antérieure. Une école supplémentaire est donc incluse dans les analyses. Cette école comporte trois classes de 7^{ème} année ayant implanté le programme. L'échantillon final est composé de 283 élèves (177 dans la condition expérimentale et 106 dans la condition contrôle).

3. Par qui et où l'évaluation a-t-elle été menée ?

L'évaluation est menée en Allemagne par Mario Gollwitzer (University of Koblenz-Landau), Katrin Eisenbach (Uni Trier), Moira Atria et Dagmar Strohmeier (University of Vienna), Rainer Banse (University of York) et Amrei Naumann (Uni Trier). L'équipe est donc partiellement la même que celle ayant réalisé les études en Autriche.

4. Quelles et combien de mesures ont été prises ?

Les données ont été récoltées en mai 2004 (pré-test), en juillet 2004 (premier post-test) et en novembre 2004 (second post-test). Étant donné que certaines classes ont été recomposées entre juillet et novembre 2004, les mesures rapportées par les pairs n'ont pas pu être prises dans toutes les classes au suivi.

Pour remplir les questionnaires, les élèves étaient amené-e-s par groupe de 5 dans un local isolé pendant les heures scolaires. Ils et elles étaient encadré-e-s par les deux formateur-trice-s ViSC de leur école ainsi que par un-e des chercheur-euse-s de l'étude. Les questionnaires étaient présentés dans le même ordre à chaque session et pour chaque élève. Après avoir rempli le questionnaire, les élèves allaient dans une autre salle afin de réaliser la tâche IAT. Les questionnaires comprenant les mesures hétéro-rapportées d'agression ont été remplis par les élèves et enseignant-e-s chez eux-elles. Les données sont restées confidentielles et un code a été attribué à chaque élève afin de pouvoir relier les mesures pré et post-test.

Les mesures qui ont été prises sont les suivantes :

- **Enrichissement du répertoire comportemental** : L'enrichissement du répertoire comportemental a été mesuré via deux tâches (uniquement au post-test). Dans les deux tâches, une situation typique de provocation était décrite. Les élèves devaient générer le plus possible de façon de réagir, en se plaçant dans la peau de la victime pour la première situation et dans celle d'un témoin neutre pour la seconde. Les réactions proposées par les élèves ont été catégorisées en réponses « constructives », « destructives » et « ambiguës » et ont été comptabilisées.

- **Mesure auto-rapportée d'agressivité** : Pour mesurer l'agressivité de manière auto-rapportée, les élèves ont dû répondre à la version allemande du *OBVQ* (Olweus, 1991; cité par Gollwitzer et al., 2006) et à l'*Assessment Sheet for Aggressive Behavior in Concrete Situations* (EAS en allemand, Petermann & Petermann, 1993; cité par Gollwitzer et al., 2006). L'*OBVQ* est composé de 22 items (ex. «*How often do you make one of your teacher angry ?*») auxquels les participant-es ont dû répondre sur une échelle à 5 points allant de *Jamais* à *Presque tous les jours*. Les items 1 à 10 mesurent la fréquence de la victimisation tandis que les items 11 à 20 mesurent la fréquence du harcèlement. Deux items qui posaient problème ont été omis. Après cela, les alphas se situaient entre .86 et .91. L'*EAS* consiste en 22 dessins qui représentent une situation typique de conflit ou de provocation. Pour chaque situation, les élèves ont dû se mettre dans la peau de la personne concernée et choisir de quelle façon ils ou elles réagiraient parmi trois options disponibles. Une des options est non agressive, une moyennement agressive et une très agressive. Les items sont ensuite moyennés pour former un index d'agression (avec des alphas allant de .87 à .89). La différence entre les scores au pré-test et au post-test permet d'avoir une mesure du changement dans l'agressivité.
- **Mesure hétéro-rapportée indirecte de l'évolution de l'agressivité** : Pour mesurer l'évolution de l'agressivité de façon hétéro-rapportée et indirecte, les élèves et enseignant-e-s ont reçu une liste reprenant les noms de tou-te-s les élèves de la classe. Ils et elles devaient indiquer à quel point chaque élève se comportait de façon agressive sur une échelle à 5 points allant de *Jamais* à *Toujours*. Pour les enseignant-e-s, la question était plus précise et portait séparément sur l'agression physique, l'agression verbale et l'agression non verbale. Comme toutes les données rapportées par les élèves étaient fortement corrélées à celles rapportées par les enseignant-e-s, elles ont été moyennées en un score global d'agressivité. La différence entre le score au pré-test et celui au post-test permet d'avoir une mesure du changement dans l'agressivité.
- **Mesure hétéro-rapportée directe de l'évolution de l'agressivité** : Pour mesurer le changement dans l'agressivité de façon hétéro-rapportée et directe, les enseignant-e-s et élèves ont dû indiquer si chaque élève était devenu beaucoup moins agressif-ve (-2), un peu moins agressif-ve (-1), n'avait pas changé (0), était devenu un peu plus agressif-ve (+1) ou beaucoup plus agressif-ve (+2). Cette mesure a été prise au post-test comme au suivi. Comme les réponses des élèves et des enseignant-e-s étaient peu corrélées (entre .09 et .22), seules les données venant des élèves ont été prises en compte dans les analyses.
- **IAT agressivité** : Dans l'étude de Gollwitzer et al. (2007), une tâche d'association implicite a été utilisée pour mesurer l'agressivité. Le temps de réponse moyen dans les blocs compatibles (association des catégories moi/agressivité et les autres/paix) a été soustrait au temps de réponse moyen dans les blocs incompatibles (moi/paix, les autres/agressivité).

Annexe 4 – Evaluation de ViSC

Des mesures ont également été prises afin de surveiller l’implantation du programme. Les mesures qui ont été prises sont les suivantes :

- **Faisabilité du programme** : La faisabilité a été mesurée via une feuille que chaque formateur-trice ViSC devait compléter après chaque leçon. Il lui était demandé (1) d’indiquer tout écart entre la structure de la leçon telle qu’elle avait été donnée et telle qu’elle était prévue dans le manuel (2) d’écrire l’horaire de la leçon comme elle avait effectivement été mise en place (3) d’indiquer quelles activités avaient bien fonctionné et (4) d’indiquer quelles activités n’avaient pas bien fonctionné.
- **Acceptation du programme** : L’acceptation du programme a été mesurée en demandant aux élèves de nommer les trois éléments du programme qu’ils et elles avaient le plus et le moins appréciés.
- **Accomplissement des principes du programme** : L’accomplissement des principes du programme a été mesuré en demandant aux élèves, lors de la leçon de réflexion, ce qu’ils et elles avaient appris durant les leçons. Il leur était également demandé quel sujet était le plus important pour eux-elles et pourquoi.

Effets de ViSC selon l’évaluation en Allemagne

1. Des effets ont-ils été observés ?

L’intervention a eu un effet au niveau de l’enrichissement du répertoire comportemental. Pour la première situation (perspective de la victime), les élèves de la condition expérimentale ont rapporté plus de réponses constructives et, de façon marginalement significative, moins de réponses ambiguës que ceux et celles de la condition contrôle. Pour la deuxième situation (perspective d’un témoin), les élèves de la condition expérimentale ont généré moins de réponses destructives, mais, de façon marginalement significative, plus de réponses ambiguës que ceux et celles de la condition contrôle.

Concernant les mesures auto-rapportées d’agressivité, bien que les moyennes vont dans le sens attendu, aucun effet du programme n’est statistiquement significatif via les MANOVA. Des tests non paramétriques ont également été utilisés afin de tenir compte de la non-normalité de la distribution des données. Dans les classes expérimentales, une diminution du harcèlement comme de la victimisation est observée entre le pré et le post-test tandis que dans les classes contrôles seule la victimisation diminue.

Concernant les mesures hétéro-rapportées d’agressivité, il n’y a pas d’effet significatif de l’intervention pour les mesures indirectes. Par contre, au niveau des mesures directes, les élèves de la condition expérimentale perçoivent un déclin plus important de l’agressivité que ceux et celles de la condition contrôle. Toutefois, lors du suivi, cet effet n’est plus significatif.

Au niveau de la tâche IAT (investiguée par Gollwitzer et al., 2007), il n’y a pas d’effet du programme au post-test. Par contre, alors qu’entre le post-test et le follow-up les scores d’agressivité augmentent dans la condition contrôle, ils restent stables dans la condition expérimentale. Il semble donc que le programme ait un effet sur une mesure implicite d’agressivité, mais uniquement plusieurs mois après qu’il ait pris fin. Selon les auteur-e-s, il s’agirait d’un effet de transfert, ce qui expliquerait qu’il n’apparaisse pas tout de suite.

2. Les effets sont-ils différenciés ?

Les auteur·e·s ne rapportent aucune analyse permettant de savoir si les effets sont différenciés. Il est toutefois intéressant de noter que les changements dans le niveau d'agressivité varient fortement d'une classe expérimentale à l'autre. Ces différences sont significatives au niveau du changement d'agressivité mesuré de façon indirecte au post-test ainsi qu'au niveau des changements d'agressivité mesurés de façon directe, au post-test comme au suivi. Cela peut s'expliquer par le fait que l'implantation du programme peut fortement varier d'une classe à l'autre. Mais comme aucune analyse statistique n'a été réalisée sur la qualité de l'implantation et que les caractéristiques de classes n'ont pas été investiguées, il n'est pas possible de savoir à quoi ces différences sont dues.

3. Des mesures de suivi ont-elles été prises ?

Une mesure de suivi a été prise environ quatre mois après le premier post-test, mais uniquement pour une partie des variables. L'effet de l'intervention sur la mesure directe du changement dans l'agressivité n'est plus significatif alors que l'effet de l'intervention sur la tâche IAT le devient. Pour les autres variables, il n'y a pas de différence entre les deux post-tests.

4. L'implantation a-t-elle été mesurée ?

La faisabilité du programme, son acceptation et l'accomplissement de ses principes ont été mesurés, mais l'impact de l'implantation sur les effets du programme n'a pas été investigué.

Concernant la faisabilité, environ la moitié des leçons ont été structurées exactement comme prévu. Dans l'autre moitié des cas, il y avait des écarts, mais qui n'allaient pas à l'encontre des principes du programme (sauf pour une leçon dans une des classes). Concernant l'acceptation du programme, les éléments les plus souvent mentionnés comme étant les plus appréciés par les élèves sont la phase d'action ainsi que les jeux et jeux de rôles joués lors de la phase d'impulsion. Les élèves préfèrent donc les parties interactives. 25 % des élèves n'ont mentionné aucun élément moins apprécié. Les éléments mentionnés par les autres élèves sont le bruit dans la classe, les devoirs, la leçon de réflexion et les parties théoriques du programme. Concernant l'accomplissement des principes du programme, les élèves ont principalement rapporté avoir appris des alternatives comportementales dans les situations critiques, avoir découvert plus de choses sur leur classe et avoir appris les règles de la classe. Selon eux, les sujets les plus importants étaient l'apprentissage de façon de réagir dans les situations de conflit, les vignettes et les jeux de rôles. Ces éléments étaient choisis, car ils avaient un intérêt pratique ou étaient importants pour toute la classe.

Implantation de ViSC à Chypre

Comme dans beaucoup d'autres pays, le harcèlement est un problème sérieux dans les écoles chypriotes. Selon Stavrinides, Paradeisiotou, Tziogouros et Lazarou (2010; cité par Solomontos-Kountouri, Gradinger, Yanagida, & Strohmeier, 2016), 17 % des élèves sont concerné·e·s par le harcèlement, que ce soit en tant que victime, auteur·e ou les deux à la fois. Un processus national, similaire à la stratégie nationale contre la violence en Autriche, a été initié à Chypre et le programme ViSC a été implanté dans plusieurs écoles (Solomontos-Kountouri et al., 2016).

Selon Solomontos-Kountouri et al. (2016), le programme a été implanté avec la plus haute fidélité possible. Seules quelques adaptations ont été nécessaires pour répondre aux particularités culturelles et éducationnelles chypriotes. Le processus d'implantation a été assez complexe. Des négociations ont d'abord eu lieu avec le ministère de l'Éducation pour pouvoir implanter le programme. Il a ensuite fallu former les neuf coachs ViSC et traduire le matériel en grec. Les coachs ViSC ont reçu une formation de 20 h dispensée par des chercheur·euse·s autrichien·ne·s et chypriotes. Ensuite, les coachs ont donné cinq unités de formation à tou·te·s les enseignant·e·s lors des heures scolaires. Et les enseignant·e·s qui implantent le projet de classe ont donné 6 unités de 90 min à leurs élèves.

Par rapport au programme tel qu'il a été implanté en Autriche, deux nouveaux éléments ont été ajoutés : des membres du ministère de l'Éducation ont été impliqués dans l'implantation du programme et ont participé à une réunion officielle de lancement et une présentation de 2 h du programme a été donnée par les chercheur·euse·s aux parents dans chaque école implantant ViSC. Les parents ont reçu des informations sur la façon dont ils pouvaient aider à lutter contre le harcèlement.

Le programme a été implanté dans les écoles secondaires lors de l'année scolaire 2012-2013. L'année suivante, trois formations supplémentaires ont été dispensées à un petit groupe d'enseignant·e·s afin qu'ils et elles puissent former le reste du personnel qui ne l'était pas encore, organiser une présentation pour les parents et implanter le programme auprès des nouveaux élèves. Cette seconde année d'implantation a été ajoutée parce qu'à Chypre certains enseignant·e·s changent d'école chaque année.

Qualité de l'évaluation de ViSC à Chypre

Article principal : Solomontos-Kountouri, O., Gradinger, P., Yanagida, T., & Strohmeier, D. (2016). The implementation and evaluation of the ViSC program in Cyprus: challenges of cross-national dissemination and evaluation results. *European Journal of Developmental Psychology*, 13(6), 737-755. doi:10.1080/17405629.2015.1136618

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude longitudinale quasi-expérimentale. Étant donné le petit nombre d'écoles participant à l'implantation du programme, il n'était pas possible d'adopter un design RCT. Les écoles ont été choisies par la *Secondary School Directory*. Le choix s'est fait dans l'optique d'avoir une implantation la plus élevée possible. Les écoles ont été choisies sur base de la commodité et de la demande (les écoles choisies sont connues pour collaborer efficacement et ont besoin d'un programme d'intervention). Les écoles contrôles

Annexe 4 – Evaluation de ViSC

ont été choisies sur base d'un critère d'équivalence (des écoles similaires ont été appariées aux écoles d'intervention).

Au niveau des analyses statistiques, des modèles linéaires hiérarchiques ont été utilisés afin d'évaluer les effets du programme en tenant compte du fait que le temps est niché dans les élèves et les élèves dans les écoles. La procédure d'estimation utilisée était une procédure d'estimation bayésienne (avec le *prior* réglé sur les paramètres par défaut de façon à ne favoriser aucune hypothèse).

2. Quelles étaient les caractéristiques de l'échantillon ?

Les six écoles participantes sont localisées dans trois provinces différentes. 88,6 % des parents ont donné leur accord pour que leur enfant participe à l'étude. Au total, 1652 élèves de 7^{ème} et 8^{ème} année (602 dans la groupe expérimental et 1050 dans le groupe contrôle) venant de 82 classes (30 dans le groupe expérimental et 52 dans le groupe contrôle) ont participé à au moins un moment de mesure et ont donc été inclus-es dans l'analyse.

Au pré-test, l'échantillon était composé de 1568 élèves (48,9 % de filles, avec un âge moyen de 12,6 ans). Les élèves du groupe contrôle et du groupe expérimental ne se distinguent sur aucune caractéristique démographique mis à part la langue maternelle (moins d'élèves ont le grec pour langue maternelle dans le groupe contrôle) et l'âge (les élèves du groupe contrôle sont légèrement plus jeunes) pour les élèves de 8^{ème} année.

47,9 % des élèves en 7^{ème} année et 52,1 % en 8^{ème} année présentaient des données manquantes. Les questionnaires de 612 élèves étaient manquants lors d'une ou plusieurs vagues de mesure (parce qu'ils ou elles étaient absent·e·s, n'avaient pas envie de participer ou ont rempli le questionnaire de manière invalide) et les questionnaires de 317 élèves comprenaient des données manquantes pour des items isolés. Les élèves présentant des données manquantes rapportent des valeurs plus élevées sur toutes les échelles que les élèves sans données manquantes. Pour tenir compte de cela, une procédure FIML (*full information maximum likelihood*) sous le postulat MAR (*missing at random*) a été utilisée.

3. Par qui et où l'évaluation a-t-elle été menée ?

L'évaluation est menée à Chypre par Olga Solomontos-Kountouri (Naopolis University of Pafos), Petra Gradinger, Takuya Yanagida et Dagmar Strohmeier (University of Applied Sciences Upper Austria). Les auteur·e·s sont donc essentiellement les mêmes que ceux-celles qui ont réalisé les études en Autriche.

4. Quelles et combien de mesures ont été prises ?

Les données ont été prises à trois reprises : en octobre-novembre 2012 (pré-test), en avril-mai 2013 (post-test) et en mars-avril 2014 (follow-up). Avant la récolte des données, il était assuré aux élèves que leur participation était volontaire et que leurs données resteraient confidentielles. Les données ont été récoltées via un questionnaire papier/crayon durant les heures scolaires par Olga Solomontos-Kountouri et un groupe d'assistant·e·s de recherche, avec l'aide des enseignant·e·s.

Annexe 4 – Evaluation de ViSC

Pour toutes les mesures, les élèves ont répondu sur une échelle de 5 points (0=*Jamais*, 1=*Une ou deux fois*, 2=*Deux ou trois fois par mois*, 3=*Une fois par semaine*, 4=*Presque tous les jours*). Les échelles utilisées sont les mêmes que pour les études en Autriche. Les mesures qui ont été prises sont les suivantes :

- **Harcèlement** : Le harcèlement a été mesuré via une échelle auto-rapportée qui comprend un item global et trois items couvrant des formes spécifiques de harcèlement (physique, relationnel et verbal). Le terme « harcèlement » n'est ni défini, ni utilisé dans les items, mais des comportements très spécifiques sont décrits. Les alphas sont égaux à .74 (au pré-test), .77 (au post-test) et .76 (au follow-up).
- **Victimisation** : La victimisation a été mesurée de la même manière que le harcèlement, mais avec les items formulés de manière passive. Les alphas sont égaux à .75 (au pré-test), .78 (au post-test) et .78 (au follow-up).
- **Cyberharcèlement** : Le cyberharcèlement a été mesuré via une échelle auto-rapportée contenant un item global et sept items spécifiques (Smith et al., 2008; cité par Solomontos-Kountouri et al., 2016). Les alphas sont égaux à .84 (au pré-test), .93 (au post-test) et .92 (au follow-up).
- **Cybervictimisation** : La cybervictimisation a été mesurée de la même manière que le cyberharcèlement, mais avec les items formulés de manière passive. Les alphas sont égaux à .89 (au pré-test), .93 (au post-test) et .92 (au follow-up).
- **Agression physique** : L'agression physique a été mesurée via une version modifiée (afin de correspondre à une mesure auto-rapportée) du questionnaire de Crick et Grotpeter (1995; cité par Solomontos-Kountouri et al., 2016). Les alphas sont égaux à .79 (au pré-test), .80 (au post-test) et .84 (au follow-up).
- **Victimisation physique** : La victimisation physique a été mesurée de la même manière que l'agression physique, mais avec les items formulés de manière passive. Les alphas sont égaux à .80 (au pré-test), .80 (au post-test) et .78 (au follow-up).
- **Agression relationnelle** : L'agression relationnelle a été mesurée via une version modifiée (pour correspondre à une mesure auto-rapportée) du questionnaire de Crick et Grotpeter (1995; cité par Solomontos-Kountouri et al., 2016). Les alphas sont égaux à .84 (au pré-test), .88 (au post-test) et .89 (au follow-up).
- **Victimisation relationnelle** : La victimisation relationnelle a été mesurée de la même manière que l'agression relationnelle, mais avec les items formulés de manière passive. Les alphas sont égaux à .85 (au pré-test), .89 (au post-test) et .86 (au follow-up).
- **Agression verbale** : L'agression verbale a été mesurée via trois items spécifiquement conçus pour couvrir des formes actives et passives d'agression verbale (Strohmeier, Aoyama, Gardinger & Toda, 2013; cité par Solomontos-Kountouri et al., 2016). Les alphas sont égaux à .77 (au pré-test), .88 (au post-test) et .79 (au follow-up).

- **Victimisation verbale** : La victimisation verbale a été mesurée de la même manière que l'agression verbale, mais avec les items formulés de manière passive. Les alphas sont égaux à .80 (au pré-test), .84 (au post-test) et .83 (au follow-up).

Effets de ViSC selon l'évaluation à Chypre

1. Des effets ont-ils été observés ?

L'évolution de l'agression n'est pas la même dans le groupe expérimental et dans le groupe contrôle. En 7^{ème} année, l'agression physique diminue davantage dans le groupe expérimental que dans le groupe contrôle. En 7^{ème} année comme en 8^{ème} année, l'agression relationnelle augmente davantage dans le groupe expérimental, mais avec un effet d'interaction quadratique significatif (elle augmente entre le pré et le post-test et diminue entre le post-test et le follow-up). En 8^{ème} année, le harcèlement et le cyberharcèlement augmentent davantage dans le groupe expérimental, mais avec un effet d'interaction quadratique significatif (ils augmentent entre le pré et le post-test et diminuent entre le post-test et le follow-up). Par contre, aucune interaction entre le temps et la condition n'est retrouvée pour le harcèlement et le cyberharcèlement en 7^{ème} année, pour l'agression physique en 8^{ème} année, ainsi que pour l'agression verbale en 7^{ème} comme en 8^{ème} année.

Des différences entre conditions sont également observées au niveau de la victimisation. En 7^{ème} année, la victimisation physique diminue davantage dans le groupe expérimental que dans le groupe contrôle tandis que la cybervictimisation augmente davantage dans le groupe expérimental. Dans les deux groupes, il y a un effet du temps, avec un effet quadratique significatif sur la victimisation due au harcèlement, elle augmente puis diminue de la même façon dans les deux groupes. La victimisation relationnelle augmente également dans les deux groupes, avec un effet quadratique dans les deux groupes, mais l'augmentation et la diminution sont plus fortes dans le groupe expérimental. Et la victimisation verbale augmente davantage dans le groupe expérimental, mais avec un effet d'interaction quadratique significatif (elle augmente entre le pré et le post-test et diminue entre le post-test et le follow-up). En 8^{ème} année, la victimisation, la cybervictimisation et la victimisation relationnelle augmentent davantage dans le groupe expérimental, mais avec un effet d'interaction quadratique significatif (elles augmentent entre le pré et le post-test et diminuent entre le post-test et le follow-up). La victimisation physique diminue de la même manière dans les deux groupes et la victimisation verbale augmente davantage dans le groupe expérimental que dans le groupe contrôle.

En résumé, l'agression et la victimisation physique en 7^{ème} année diminuent davantage dans le groupe expérimental que dans le groupe contrôle. Dans tous les autres cas, soit il n'y a pas d'effet de l'intervention soit les valeurs augmentent davantage dans le groupe expérimental que dans le groupe contrôle, mais une relation quadratique est généralement retrouvée : à la fin de l'année d'implantation du programme, les scores ont davantage augmenté dans le groupe expérimental, mais, lors de la seconde année d'implantation, les scores diminuent davantage dans le groupe expérimental. Les auteur·e·s interprètent ces résultats comme un effet de sensibilisation : les élèves des écoles d'intervention rapporteraient davantage d'agression ou de victimisation parce qu'ils et elles y seraient plus attentif·ve·s. Lors de la

Annexe 4 – Evaluation de ViSC

seconde année, les élèves seraient toujours sensibles au cas d'agression et de victimisation, mais ceux-ci auraient diminué. Si cette hypothèse est vérifiée, cela voudrait dire que le programme est efficace.

2. Les effets sont-ils différenciés ?

Des différences sont observées entre les élèves de 7^{ème} et de 8^{ème} (avec, en général, des différences plus importantes entre les deux groupes pour les élèves de 7^{ème}). Mis à part cela, seul l'effet de l'âge a été testé et, selon les résultats, il ne modère pas l'effet du programme.

3. Des mesures de suivi ont-elles été prises ?

Des mesures de suivi ont été prises un an après la fin de l'implantation du programme. Les résultats sont généralement différents entre le pré et le post-test (augmentation globale des scores dans le groupe expérimental) et entre le post-test et le follow-up (diminution globale des scores dans le groupe expérimental).

4. L'implantation a-t-elle été mesurée ?

Aucune mesure de l'implantation n'est rapportée dans l'article.

Implantation de ViSC en Roumanie

En Roumanie, le programme ViSC a été implanté en combinaison avec le programme REBE (Trip & Bora, 2010, cité par Trip et al., 2015). REBE est un programme basé sur la théorie de la *Rational Emotive Behavioral Therapy* (Ellis, 1962; cité par Trip et al., 2015). Ce programme vise un changement au niveau des mécanismes cognitifs (tandis que ViSC vise les mécanismes comportementaux). Les activités du programme ciblent les déclencheurs de la colère, l'expérience personnelle de la colère, les conséquences de la colère et la tolérance faible à la frustration.

L'implantation du programme ViSC a uniquement été réalisée au niveau des classes, aucune mesure n'a été prise au niveau des écoles. De plus, ce ne sont pas les enseignant·e·s de l'école qui ont dispensé les unités ViSC, mais des intervenant·e·s externes. C'est formateur·trice·s possédaient déjà des connaissances sur la théorie de la *Rational Emotive Behavioral Therapy* et ont participé à une séance de 4 h présentant le programme REBE ainsi qu'à une session de 8 h sur le programme ViSC dispensée par les développeuses autrichiennes du programme. Les formateur·trice·s ont aussi eu des supervisions hebdomadaires avec les chercheur·euse·s roumains. Tout le matériel, traduit en roumain et en anglais, leur a été donné.

Qualité de l'évaluation de ViSC en Roumanie

Article principal : Trip, S., Bora, C., Sipos-Gug, S., Tocai, I., Gradinger, P., Yanagida, T., & Strohmeier, D. (2015). Bullying prevention in schools by targeting cognitions, emotions, and behavior: Evaluating the effectiveness of the REBE-ViSC program. *Journal of Counseling Psychology*, 62(4), 732-740. doi:10.1037/cou0000084

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude longitudinale quasi-expérimentale. Les écoles qui se sont portées volontaires ont été aléatoirement réparties dans la condition expérimentale REBE-ViSC (le programme REBE est implanté d'octobre à février et ViSC de mars à juin) ou dans la condition expérimentale ViSC-REBE (le programme ViSC est implanté d'octobre à février et REBE de mars à juin). Les écoles qui n'ont pas voulu participer à la l'implantation des programmes ont servi d'écoles contrôles.

Au niveau des analyses statistiques, un modèle linéaire hiérarchique à trois niveaux a été utilisé pour tester l'efficacité du programme. Le modèle tient compte du fait que le temps est niché dans les élèves et les élèves sont nichés dans les classes.

2. Quelles étaient les caractéristiques de l'échantillon ?

Toutes les écoles d'Oradea (une ville située au nord-ouest de la Roumanie) ayant des classes de 6^{ème} année ont été contactées pour participer à l'étude. Au total, 11 écoles ont participé : 3 dans le groupe contrôle (315 élèves), cinq dans le groupe REBE-ViSC (385 élèves) et trois dans le groupe ViSC-REBE (270 élèves). Le programme a été implanté dans toutes les classes de 6^{ème} année. Au total, 970 élèves venant de 35 classes ont participé à au moins un temps de mesure et ont donc été inclus dans l'analyse.

Annexe 4 – Evaluation de ViSC

Sur les 970 élèves, 806 élèves ont participé au pré-test, 729 au premier post-test et 545 au second. Au pré-test, l'échantillon est composé de 806 élèves avec un âge moyen de 11,82 ans, 53 % sont des garçons, moins de 3 % n'ont pas le roumain comme langue maternelle, 91,4 % ont un père qui travaille et 83,64 % une mère qui travaille. Seul 1 % des élèves ont leurs deux parents au chômage. Les élèves qui n'ont pas participé à un des post-test ont des scores plus faibles de victimisation que les autres. Pour gérer la présence de données manquantes, une procédure FIML (*full information maximum likelihood*) sous le postulat MAR (*missing at random*) a été utilisée. Afin d'investiguer l'effet des données manquantes, les analyses principales ont été refaites en enlevant les 224 élèves qui n'ont participé qu'à un moment de mesure. Les résultats ne sont pas différents de ceux portant sur l'échantillon total.

3. Par qui et où l'évaluation a-t-elle été menée ?

L'évaluation est menée en Roumanie par Simona Trip, Carmen Bora, Sebastian Sipos-Gug et Ioana Tocaï (University of Oradea) ainsi que Takuya Yanagida, Petra Gradingner et Dagmar Strohmeier (University of Applied Sciences Upper Austria). Les auteur·e·s sont donc en partie les mêmes que ceux·celles qui ont réalisé les études en Autriche.

4. Quelles et combien de mesures ont été prises ?

Les mesures ont été prises en octobre 2011 (avant l'implantation des programmes), en mars 2012 (après l'implantation du premier programme et avant celle du second) et en juin 2012 (après l'implantation des deux programmes). Les données ont été récoltées via des questionnaires en ligne auxquels les élèves ont répondu durant les heures scolaires, sous la supervision de deux assistant·e·s de recherche. Il était bien précisé aux élèves que leurs réponses resteraient confidentielles.

Les mesures qui ont été prises sont les suivantes :

- **Harcèlement** : Le harcèlement a été mesuré via une échelle auto-rapportée qui comprend un item global et trois items couvrant des formes spécifiques de harcèlement (physique, relationnel et verbal). Le terme « harcèlement » n'est ni défini ni utilisé dans les items, mais des comportements très spécifiques sont décrits. Les alphas se situent entre .825 et .879 suivant le moment de mesure.
- **Victimisation** : La victimisation a été mesurée de la même manière que le harcèlement, mais avec les items formulés de manière passive. Les alphas se situent entre .799 et .892 suivant le moment de mesure.
- **Régulation et expression de la colère** : La régulation et l'expression de la colère ont été mesurées avec l'*Anger Regulation and Expression Scale* (DiGiuseppe & Tafrate, 2011 ; cité par Trip et al., 2015). Il s'agit d'un instrument créé spécialement pour les adolescents entre 10 et 17 ans. Les sous-échelles *Internalizing Anger* (colère internalisée) et *Overt Anger* (colère manifeste) ont été utilisées ici. Les participant·e·s ont répondu aux différents items sur une échelle à 5 points (1=*Jamais/presque jamais*, 2=*Presque jamais/environ une fois par mois*, 3=*Parfois/environ une fois par semaine*, 4=*Souvent/plusieurs fois par jour* et 5=*Toujours/presque tous les jours*). Les alphas se situent entre .805 et .906 suivant le moment de mesure.

- **Tolérance à la frustration** : La tolérance à la frustration a été mesurée via deux sous-échelles (*LFT learning* et *LFT Entitlement*) de la *Scale of Low Frustration Tolerance for Students* (Trip & Bora, 2011 cité par Trip et al., 2015). Ces deux sous-échelles mesurent la tolérance à la frustration des élèves et seraient celles correspondant le mieux aux buts de l'intervention. Les élèves ont dû répondre à toutes les questions sur une échelle à 5 points allant de *Désaccord total* à *Accord total* (un score élevé indique donc une faible tolérance à la frustration). Les alphas se situent entre .864 et .932 suivant le moment de mesure.

Effets de ViSC selon l'évaluation en Roumanie

1. Des effets ont-ils été observés ?

Des effets du programme sont observés sur les cognitions : les scores de tolérance faible à la frustration diminuent davantage dans les groupes expérimentaux que dans le groupe contrôle. Il y a aussi une différence entre le groupe *REBE-ViSC* et le groupe contrôle concernant une émotion : les scores de colère manifeste des élèves dans le groupe contrôle augmentent de façon plus importante que ceux des élèves du groupe *REBE-ViSC* (mais cet effet est de tellement petite taille qu'il n'a pas d'importance dans la pratique). Il n'y a aucun effet de l'intervention sur le harcèlement ou sur la victimisation.

2. Les effets sont-ils différenciés ?

Les auteur·e·s ne rapportent aucune analyse permettant de savoir si les effets sont différenciés.

3. Des mesures de suivi ont-elles été prises ?

Aucune mesure de suivi n'a été prise.

4. L'implantation a-t-elle été mesurée ?

Après avoir dispensé chaque unité de *REBE* ou de *ViSC*, les formateur·trice·s devaient compléter un formulaire sur l'implantation. Ils et elles devaient rapporter les méthodes utilisées, la façon dont ils et elles avaient organisé l'activité dans son entièreté, les différentes parties de leçons dispensées ou non, les forces et limites des activités, le point auquel les élèves étaient engagé·e·s dans l'activité et leur appréciation générale de l'unité. Selon les auteur·e·s, la fidélité d'implantation était élevée, mais aucune donnée n'est reprise dans l'article.

Implantation de ViSC en Turquie

Comme dans les autres pays, le harcèlement est un problème en Turquie (Doğan et al, 2017). Entre 12 et 42 % des élèves sont impliqué·e·s dans des situations de harcèlement en tant que victimes, auteur·e·s ou les deux (Burnukara & Ucanok, 2012; Dogan-Ates & Yagmurlu, 2010; Erdur-Baker, 2009; Soydas & Ucanok, 2014; cités par Doğan et al., 2017). L'implantation et l'évaluation de programme anti-harcèlement y sont donc nécessaires. Le programme ViSC a été le premier programme anti-harcèlement avec une politique au niveau de l'école entière à être adapté, implanté et évalué en Turquie. Ce programme a été choisi parce qu'il avait déjà été implanté avec succès dans d'autres pays et qu'il semblait faisable, malgré des ressources limitées en termes de personnel et de moyens financiers.

L'implantation du programme en Turquie s'est déroulée en plusieurs étapes. Tout d'abord, la permission d'implanter le programme dans des écoles secondaires a été obtenue auprès du comité éthique de l'université et de la branche locale du ministre de l'Éducation. Ensuite, un consortium comprenant Aysun Doğan (première auteure de l'étude) et des représentant·e·s de la branche locale du ministère de l'Éducation a été créé. Ce consortium a commencé à se réunir 9 mois avant l'implantation du programme et était en charge de tous les détails organisationnels. C'est lui qui a sélectionné les 10 coachs ViSC (des psychologues) ainsi que les écoles qui participeraient à la condition contrôle et à la condition expérimentale. Troisièmement, tout le matériel a été traduit (par des psychologues bilingues allemand-turc). Quelques ajustements mineurs ont été réalisés afin d'adapter le matériel à la culture turque. Quatrièmement, les coachs ViSC ont été formé·e·s par les développeuses autrichiennes du programme ainsi que par Aysun Doğan. Trois workshops pour les coachs ont été organisés à l'Ege University. Le premier workshop a été organisé en avril 2015 et durait 14 h. La formation était focalisée sur la définition et la reconnaissance du harcèlement, la façon de traiter les cas de harcèlements et les mesures préventives au niveau de l'école. Le deuxième workshop a été organisé en mai 2015 et durait également 14 h. Il se concentrait sur la façon d'implanter le projet de classe ViSC et les activités de classe. Le dernier workshop a eu lieu en octobre 2015 et durait 4 h. Il était focalisé sur la nature du cyberharcèlement et sur les feedbacks des coachs concernant les succès et challenges de l'implantation du programme dans les écoles. Cinquièmement, les coachs ViSC ont organisé les formations dans les écoles. La formation pour tou·te·s les enseignant·e·s durait 3 h et avait lieu en octobre ou novembre 2015. Au même moment, une réunion d'une heure était organisée pour expliquer le projet aux parents et obtenir leur consentement pour que leur enfant participe à l'étude. Sixièmement, des enseignant·e·s volontaires ont été sélectionné·e·s pour implanter le projet de classe entre décembre 2015 et mai 2016. Les coachs ViSC ont formé chaque enseignant·e individuellement puis les enseignant·e·s ont implanté le projet de classe (de la même façon qu'en Autriche). Pour que le programme soit implanté avec une fidélité importante, des assistant·e·s de recherche ont aidé les enseignant·e·s à préparer les différentes séances.

Qualité de l'évaluation de ViSC en Turquie

Article principal : Doğan, A., Keser, E., Şen, Z., Yanagida, T., Gradinger, P., & Strohmeier, D. (2017). Evidence Based Bullying Prevention in Turkey: Implementation of the ViSC Social Competence Program. *International Journal of Developmental Science*, 11, 93–108. doi:10.3233/DEV-170223

1. Quel était le plan de recherche de l'évaluation ?

Un design longitudinal quasi-expérimental a été adopté. Les écoles n'ont pas été réparties aléatoirement dans les conditions, mais, pour les écoles de la condition expérimentale, les classes ont été aléatoirement réparties entre deux conditions expérimentales. Ces deux conditions expérimentales ont été créées afin d'investiguer les effets du dosage : dans la condition « niveau de l'école », les classes se trouvaient dans des écoles où le programme était implanté, mais ne participaient pas aux activités de classe, tandis que dans la condition « niveau de l'école et de la classe », les classes, en plus d'être dans une école implantant le programme, réalisaient le projet de classe ViSC. Le programme a été implanté lors de l'année scolaire 2015-2016, le projet de classe étant implanté de décembre à mai.

Au niveau des analyses statistiques, des modèles linéaires hiérarchiques ont été utilisés afin d'évaluer les effets du programme. Un premier modèle a été utilisé pour investiguer l'effet des conditions expérimentales par rapport à la condition contrôle et un second pour investiguer les éventuelles différences entre les deux conditions expérimentales.

2. Quelles étaient les caractéristiques de l'échantillon ?

Les mesures pré-test ont été prises en novembre 2015, les mesures post-test en juin 2016 et les mesures de suivi en novembre 2016. Le programme a été implanté chez des élèves de 5^{ème} année (la première année de secondaire en Turquie). La condition « niveau de l'école et de la classe » a été implantée dans 9 classes de 3 écoles et la condition « niveau de l'école » dans 8 autres classes des trois mêmes écoles. 9 classes de trois autres écoles ont servi de groupe contrôle. Au total, 642 élèves (227 dans la condition « niveau de l'école et de la classe », 201 dans la condition « niveau de l'école » et 214 dans la condition contrôle) ont participé·e·s à au moins un moment de mesure et ont donc été inclus·es dans les analyses.

Au pré-test, l'échantillon comprenait 602 élèves (52 % de filles, âge moyen de 10,06 ans). Les différents groupes ont été comparés sur les mesures sociodémographiques et aucune différence significative n'a été trouvée.

Au niveau des données manquantes, 232 dossiers au total étaient incomplets. 195 élèves n'ont pas répondu à toutes les mesures parce qu'ils ou elles étaient absent·e·s, ont rempli le questionnaire de manière invalide ou ont changé d'école durant l'étude. 37 élèves ont omis des items individuels. Des analyses ont montré que les élèves ayant complété tous les questionnaires et ceux ayant des données manquantes ne différaient pas sur les variables dépendantes. La technique FIML (*full information maximum likelihood*) sous le postulat MAR (*missing at random*) a tout de même été utilisé pour tenir compte des données manquantes.

3. Par qui et où l'évaluation a-t-elle été menée ?

L'évaluation est menée en Turquie par Aysun Doğan, Eda Keser et Zeynep Şen (Ege University), Takuya Yanagida (University of Vienna), Petra Gradinger et Dagmar Strohmeier (University of Applied Sciences Upper Austria). Les auteur·e·s sont donc en partie les mêmes que ceux·celles qui ont réalisé les études en Autriche.

4. Quelles et combien de mesures ont été prises ?

Les mesures ont été prises à trois reprises : en novembre 2015 (pré-test), juin 2016 (post-test) et novembre 2016 (suivi). Au niveau des mesures démographiques, les élèves ont dû indiquer leur genre, âge, situation familiale (parents en couple ou non), situation financière perçue, ainsi que le niveau d'étude et l'emploi de leurs parents. Pour toutes les mesures, les élèves ont répondu sur une échelle de 5 points (0=*Jamais*, 1=*Une ou deux fois*, 2=*Deux ou trois fois par mois*, 3=*Une fois par semaine*, 4=*Presque tous les jours*). Les mesures qui ont été prises sont les mêmes que pour les études réalisées en Autriche et à Chypre :

- **Harcèlement** : Le harcèlement a été mesuré via une échelle auto-rapportée qui comprend un item global et trois items couvrant des formes spécifiques de harcèlement (physique, relationnel et verbal). Le terme « harcèlement » n'est ni défini ni utilisé dans les items, mais des comportements très spécifiques sont décrits. Les alphas se situent entre .66 et .74 suivant les moments de mesure.
- **Victimisation** : La victimisation a été mesurée de la même manière que le harcèlement, mais avec les items formulés de manière passive. Les alphas se situent entre .77 et .80 suivant les moments de mesure.
- **Cyberharcèlement** : Le cyberharcèlement a été mesuré via une échelle auto-rapportée contenant un item global et sept items spécifiques (Smith et al., 2008; cité par Doğan et al., 2017). Les alphas se situent entre .76 et .84 suivant les moments de mesure.
- **Cybervictimisation** : La cybervictimisation a été mesurée de la même manière que le cyberharcèlement, mais avec les items formulés de manière passive. Les alphas se situent entre .81 et .83 suivant les moments de mesure.
- **Agression physique** : L'agression physique a été mesurée via une version modifiée (pour correspondre à une mesure auto-rapportée) du questionnaire de Crick et Grotpeter (1995; cité par Doğan et al., 2017). Les alphas se situent entre .63 et .75 suivant les moments de mesure.
- **Victimisation physique** : La victimisation physique a été mesurée de la même manière que l'agression physique, mais avec les items formulés de manière passive. Les alphas se situent entre .76 et .77 suivant les moments de mesure.
- **Agression relationnelle** : L'agression relationnelle a été mesurée via une version modifiée (pour correspondre à une mesure auto-rapportée) du questionnaire de Crick et Grotpeter (1995; cité par Doğan et al., 2017). Les alphas se situent entre .62 et .84 suivant les moments de mesure.

Annexe 4 – Evaluation de ViSC

- **Victimisation relationnelle** : La victimisation relationnelle a été mesurée de la même manière que l'agression relationnelle, mais avec les items formulés de manière passive. Les alphas se situent entre .81 et .88 suivant les moments de mesure.
- **Aggression verbale** : L'agression verbale a été mesurée via trois items spécifiquement conçus pour couvrir des formes actives et passives d'agression verbale (Strohmeier, Aoyama, Gardinger & Toda, 2013 ; cité par Doğan et al., 2017). Les alphas se situent entre .66 et .78 suivant les moments de mesure.
- **Victimisation verbale** : La victimisation verbale a été mesurée de la même manière que l'agression verbale, mais avec les items formulés de manière passive. Les alphas se situent entre .78 et .82 suivant les moments de mesure.

Une série d'analyses ont été conduites pour établir des modèles de mesure pour les cinq types d'agression et de victimisation. Les résultats ont montré une adéquation des modèles acceptable pour la plupart des variables (mais pas pour le harcèlement ainsi que pour l'agression et la victimisation relationnelle). Le cyberharcèlement et la cybervictimisation ont été supprimés des analyses étant donné leur fréquence très faible.

Effets de ViSC selon l'évaluation en Turquie

5. Des effets ont-ils été observés ?

L'évolution de l'agression n'est pas la même dans les groupes expérimentaux et dans le groupe contrôle. Le harcèlement et l'agression relationnelle augmentent davantage dans les groupes expérimentaux que dans le groupe contrôle, mais avec un effet d'interaction quadratique significatif (ils augmentent entre le pré et le post-test et diminuent entre le post-test et le suivi). L'agression physique augmente davantage dans le groupe « niveau-école » que dans le groupe contrôle, mais avec un effet d'interaction quadratique significatif (elle augmente entre le pré et le post-test et diminue entre le post-test et le suivi). Il n'y a pas d'effet de l'intervention sur l'agression verbale et aucune différence significative n'est retrouvée entre les deux groupes expérimentaux.

Au niveau de la victimisation, des différences entre groupes sont également présentes. La victimisation due au harcèlement et la victimisation relationnelle augmentent davantage dans les groupes expérimentaux que dans le groupe contrôle, mais avec un effet d'interaction quadratique significatif (elles augmentent entre le pré et le post-test et diminuent entre le post-test et le suivi). Par contre, il n'y a pas d'effet de l'intervention sur la victimisation physique ou sur la victimisation verbale. Aucune différence significative n'est retrouvée entre les deux groupes expérimentaux.

En résumé, il n'y a pas d'effet de l'intervention sur l'agression verbale, sur la victimisation physique ou sur la victimisation verbale. Sur les autres variables, il y a un effet de l'intervention, mais contraire au sens des hypothèses : entre le pré et le post-test, les niveaux d'agression et de victimisation augmentent davantage dans les groupes expérimentaux que dans le groupe contrôle. Toutefois un effet quadratique est

Annexe 4 – Evaluation de ViSC

systematiquement trouvé, avec une diminution plus importante dans les groupes expérimentaux que dans les groupes contrôles entre le post-test et le follow-up. Selon les auteurs, ces résultats peuvent être interprétés comme étant dus à un problème statistique (étant donné qu'au pré-test les niveaux d'agression et de victimisation sont très bas dans les écoles expérimentales -plus que dans les écoles contrôles- un effet plafond est présent) ou étant dus à un effet de sensibilisation. Si cette dernière hypothèse est vérifiée, cela voudrait dire que le programme est efficace.

6. Les effets sont-ils différenciés ?

Aucune différence n'a été observée entre la condition « niveau-école » et la condition « niveau-école et classe », il n'y a donc pas d'effet de dosage du programme. Les auteurs ne rapportent aucune autre analyse permettant de savoir si les effets sont différenciés.

7. Des mesures de suivi ont-elles été prises ?

Une mesure de suivi a été prise un an après le pré-test, c'est-à-dire 5 mois après la fin de l'année d'implantation du programme. Les effets sont différents entre le pré et le post-test (augmentation globale de l'agression et de la victimisation dans les groupes expérimentaux) et entre le post-test et le follow-up (diminution globale de la victimisation et de l'agression dans les groupes expérimentaux).

8. L'implantation a-t-elle été mesurée ?

Aucune mesure de l'implantation n'est rapportée dans l'article.

Résumé des effets de ViSC

Les effets observés sont-ils constants d'une étude à l'autre ?

En prenant en compte les résultats des six études ayant évalué le programme ViSC dans divers contextes, il est possible de tirer plusieurs conclusions.

Tout d'abord, il semble que lorsque le programme est implanté uniquement au niveau de la classe et par des intervenant·e·s externes, il n'a aucun effet sur les niveaux d'agression, de harcèlement et de victimisation. Ces résultats ont été à la fois observés en Autriche (Atria & Spiel, 2007) et en Roumanie (Trip et al., 2015). Par contre, il ne semble pas qu'il y ait un effet du dosage du programme : lorsque celui-ci est implanté au niveau de l'école, le fait qu'il soit également implanté au niveau de la classe n'influence pas les résultats observés (Doğan et al., 2017).

Lorsque ViSC est implanté au niveau de l'école dans son entièreté, les résultats sont plus encourageants. En Autriche (Grading et al., 2015; Yanagida et al., 2016), des effets positifs du programme ont été observés sur la victimisation, le cyberharcèlement et la cybervictimisation (avec un maintien des effets 6 mois plus tard). Toutefois, le programme n'a pas eu d'effet sur le harcèlement. En Allemagne (Gollwitzer et al., 2006), l'implantation du programme a eu des effets positifs sur le harcèlement auto-rapporté ainsi que sur l'agressivité hétéro-rapportée (mais uniquement lorsque l'évolution de l'agressivité est mesurée de manière directe et tout de suite après l'implantation du programme).

À Chypre (Solomontos-Kountouri et al., 2016) et en Turquie (Doğan et al., 2017), un même schéma de résultats a été observé pour un nombre important de variables mesurant l'agression, le harcèlement ou la victimisation : entre le pré-test et le post-test, les scores moyens augmentent davantage dans le groupe expérimental que dans le groupe contrôle puis, entre le post-test et le follow-up, ils diminuent davantage dans le groupe expérimental que dans le groupe contrôle. Les auteurs interprètent ces résultats comme un effet de sensibilisation : les élèves des écoles d'intervention rapporteraient davantage d'agression ou de victimisation suite à la première année d'implantation du programme parce qu'ils et elles y seraient plus attentif·ve·s. Lors de la seconde année, les élèves seraient toujours sensibles au cas d'agression et de victimisation, mais ceux-ci auraient diminué. Si cette hypothèse est vérifiée, cela voudrait dire que le programme est efficace.

Le programme semble-t-il plus efficace pour une certaine population ?

Il semble que l'âge, le genre et le niveau scolaire n'influencent pas de manière importante les effets du programme ViSC. Par contre, les niveaux initiaux d'agression, de harcèlement et de victimisation modèrent les effets du programme : les élèves ayant des scores plus élevés dans ces dimensions au pré-test présentant des diminutions plus importantes au post-test que les autres élèves (Grading et al., 2016; Yanagida et al., 2016).

Il semble également que le programme est efficace indépendamment des caractéristiques des classes telles que le climat de classe ou la diversité ethnique (Gradinger et al., 2016). Toutefois, en Allemagne (Gollwitzer et al., 2006), l'intensité des changements dans le niveau d'agressivité variait fortement d'une classe à l'autre. Dans les autres pays, les différences d'effet d'une classe à l'autre n'ont pas été investiguées. Il semble donc probable que la façon dont l'enseignant·e implante le programme ou que d'autres caractéristiques de la classe modèrent les effets du programme.

Annexe 5 : Evaluation de *Second Step*

Implantation de *Steps to Respect*

Steps to Respect a été implanté de nombreuses fois aux États-Unis. Il a été évalué à deux reprises, en 2000-2001 et en 2008-2009. Il est d'ailleurs répertorié comme ayant fait ces preuves sur deux sites américains cataloguant des programmes de prévention et/ou d'intervention. Il est considéré comme « efficace » par Crime Solution (Crime Solution, 2012), un site développé par l'institut national de justice, et comme « prometteur » par Blueprints (Blueprints, nd), un site développé par le *Center for the Study and Prevention of Violence* de l'University of Colorado Boulder.

Qualité de l'évaluation de 2000-2001

Article principal : Frey, K. S., Hirschstein, M. K., Snell, J. L., Edstrom, L.V., MacKenzie, E. P., & Broderick, C. J. (2005). Reducing playground bullying and supporting beliefs: An experimental trial of the steps to respect program. *Developmental Psychology*, 41(3), 479-491. doi:10.1037/0012-1649.41.3.479

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude longitudinale avec un design RCT. Les écoles participantes ont été appariées puis réparties aléatoirement entre la condition expérimentale et la condition contrôle. Les leçons ont été implantées de décembre à mai. Les mesures pré-test ont été prises en novembre et les mesures post-test en avril-mai, elles n'ont donc pas été prises à la même période de l'année ce qui peut biaiser les résultats obtenus.

Pour la partie longitudinale de l'étude (Frey, Hirschstein, Edstrom, & Snell, 2009), les écoles du groupe expérimental ont été suivies deux ans. Toutefois ce n'était pas le cas des écoles du groupe contrôle qui ont commencé à implanter le programme après avoir été placé sur liste d'attente pendant un an. En l'absence d'un groupe contrôle, les données des élèves au temps 3 et 4 (12 et 18 mois après le pré-test) ont été comparées aux données du groupe expérimental de même niveau scolaire au temps 1 et 2 (au pré-test et 6 mois après).

Au niveau des analyses statistiques, des modèles linéaires hiérarchiques ont été utilisés afin de tester l'impact du programme sur les variables dépendantes. Les modèles, à deux niveaux, tiennent compte du fait que le temps est niché dans les élèves qui sont nichés dans les classes et contrôlent l'effet du genre, du niveau scolaire et des scores obtenus au pré-test.

2. Quelles étaient les caractéristiques de l'échantillon ?

Six écoles primaires du Nord-Ouest pacifique ont participé à l'étude. Deux écoles ont participé à partir de novembre 2000 et quatre autres à partir de novembre 2001. Pour qu'une école puisse participer, il fallait qu'au moins 80% des membres du personnel aient voté en faveur de l'implantation du programme, que l'école accepte d'être assignée aléatoirement à une des conditions et qu'elle accepte de ne pas introduire

Annexe 5 – Evaluation de *Second Step*

d'autre intervention similaire pendant la durée de l'étude. Les écoles ont été appariées sur base de leur taille, composition ethnique, et du pourcentage d'enfants recevant des repas de midi gratuit ou à prix réduit. Au pré-test, il n'y a aucune différence entre les deux conditions sur les variables démographiques ou d'intérêt.

Le programme a été dispensé à tou·te·s les élèves de la 3^{ème} à la 6^{ème} année. 64% d'entre eux et elles ont reçu un consentement parental pour participer à l'étude. Lors du pré-test, il y avait 1126 élèves, 590 en 3^{ème} ou 4^{ème} année (recevant le 1^{er} niveau du programme), 277 en 5^{ème} année (recevant le 2^{ème} niveau) et 259 en 6^{ème} année (recevant le 3^{ème} niveau). Il y avait autant de filles que de garçons dans l'échantillon. 620 élèves ont été aléatoirement sélectionnés pour faire partie d'un sous-échantillon d'élèves pour lequel·le·s des observations en cours de récréation ont également été réalisées. Cela correspond environ à 10 élèves dans chaque classe.

Le traitement des données manquantes s'est fait en ôtant des analyses les élèves pour lequel·le·s toutes les données nécessaires n'étaient pas disponible. Que ce soit dans le groupe contrôle ou dans le groupe expérimental, le taux de participation au post-test était de 91,3%. L'échantillon final est composé de ces 1023 élèves, mais l'échantillon varie entre 907 et 919 suivant les variables étant donné que certain·e·s élèves présentaient des données manquantes sur certaines échelles uniquement. Concernant les données observationnelles, les élèves ayant été observé·e·s pendant moins de 40 min au total n'ont pas été inclus·es dans les analyses. L'échantillon final est composé de 544 élèves (taux de participation de 92,4% dans les deux groupes).

Tou·te·s les enseignant·e·s (36 dans chaque condition) ont accepté de compléter les mesures ainsi que, pour ceux et celles de la condition expérimentale, d'être observé·e·s en classe deux fois par mois. En échange de leur participation, ils et elles recevaient \$100 au pré comme au post-test. 84,9% des enseignant·e·s étaient des femmes.

Pour l'étude de Hirschstein et al. (2007) portant sur l'implantation du programme, seules les trois écoles de la condition expérimentale ont été prises en compte. L'échantillon total est donc composé de 859 élèves dont 549 ont obtenu le consentement parental et l'échantillon pour les observations en cour de récréation est composé de 296 élèves. Les enseignant·e·s participant à l'étude sont les 36 enseignant·e·s des classes expérimentales. Lors de l'étude de suivi (Frey et al., 2009), les 225 élèves du groupe expérimental qui, au post-test, était en 3^{ème} (95 élèves) ou en 4^{ème} année (130 élèves) ont été suivis pendant deux ans. Les caractéristiques sociodémographiques des élèves n'ont pas changé d'une année à l'autre. 22 enseignant·e·s ont également participé à l'étude longitudinale.

3. Par qui et où l'évaluation a-t-elle été menée ?

L'évaluation a été menée aux États-Unis par Karin Frey, Miriam Hirschstein, Jennie Snell, Leihua Van Schoiack Edstrom, Elisabeth MacKenzie et Carole Broderick, toutes membres du *Committee for Children*, l'organisme ayant développé le programme.

4. Quelles et combien de mesures ont été prises ?

Les enseignant·e·s et les élèves ont répondu aux questionnaires en novembre 2000 (ou 2001) pour le pré-test et en avril-mai 2001 (ou 2002) pour le post-test. Les évaluations par les enseignant·e·s des compétences interpersonnelles de leurs élèves ont été complétées mi-octobre 2000 (ou 2001) et mai-juin 2001 (ou 2002). Pour l'étude longitudinale, les questionnaires ont également été remplis en novembre 2001 (ou 2002) et mai-juin 2002 (ou 2003).

Les questionnaires ont été complétés en classe. Les élèves étaient encadré·e·s par des assistant·e·s de recherche qui lisaient les items à voix haute et répondaient individuellement à leurs questions éventuelles. Mis à part les compétences d'interaction avec les pairs, les différentes variables ont été identifiées via une analyse factorielle réalisée sur les 60 items de l'échelle *The Student Experience Survey : What School is Like for Me* dans une étude précédente (Edstrom, Bruschi & MacKenzie, 2004 ; cité par Frey et al., 2005a). Mis à part pour l'assertivité, les données ont été transformées afin d'améliorer la normalité des distributions. Les mesures qui ont été prises sont les suivantes :

- **Compétences d'interaction avec les pairs** : Les compétences d'interaction des élèves avec leurs pairs ont été évaluées par leur enseignant·e via la sous-échelle *Peer-Preferred Social Behavior* de la *Walker-McConnell Scale of Social Competence and School Adjustment Elementary Version* (Walker & McConnell, 1995 ; cité par Frey et al., 2005a). Cette sous-échelle est composée de 17 items (ex. « *Voluntary provides assistance to peers who require it* ») auxquels il faut répondre sur une échelle à 5 points allant de *Jamais* à *Fréquemment*.
- **Victimisation** : La victimisation a été mesurée via 8 items auto-rapportés (ex. « *A group of kids at school called me mean names.* »). L'alpha de l'échelle est égal à .84.
- **Harcèlement direct** : Le harcèlement direct a été mesuré via 8 items auto-rapportés (ex. « *I called kids names at school* »). L'alpha de l'échelle est égal à .87.
- **Harcèlement indirect** : Le harcèlement indirect a été mesuré via 8 items auto-rapportés (ex. « *I told my friends to ignore kids that I was mad at.* »). L'alpha de l'échelle est égal à .76.
- **Acceptation du harcèlement** : L'acceptation du harcèlement a été mesurée via 7 items auto-rapportés (ex. « *It's okay to say something mean to a kid who really makes you angry.* »). L'alpha de l'échelle est égal à .86.
- **Difficulté à répondre de façon assertive** : La difficulté de répondre de manière assertive a été mesurée via 5 items auto-rapportés (ex. « *Kids at school are teasing you. How hard would it be to calmly tell them to stop ?.* »). L'alpha de l'échelle est égal à .81.
- **Responsabilité en tant que témoin** : La responsabilité en tant que témoin a été mesurée via 5 items auto-rapportés (ex. « *If my friends were telling lies about another kid I would tell them to stop.* »). L'alpha de l'échelle est égal à .88.

Annexe 5 – Evaluation de *Second Step*

- **Intérêt en tant que spectateur** : L'intérêt en tant que spectateur a été mesuré via 6 items auto-rapportés (ex. « *It's fun to watch other kids being ganged up on.* »). L'alpha de l'échelle est égal à .88. Étant donné le manque de variabilité des scores, l'échelle a été ôtée de l'analyse.
- **Perception de la réactivité des adultes** : La perception de la réactivité des adultes a été mesurée via 4 items auto-rapportés (ex. « *Adults at my school stop kids from being bullied.* »). L'alpha de l'échelle est égal à .59.
- **Ami·e·s solidaires** : Le fait d'avoir des ami·e·s solidaires a été mesuré via 3 items auto-rapportés (ex. « *At school, I have friends who really care about me.* »). L'alpha de l'échelle est égal à .76.

Les observations ont été réalisées sur deux mois et demi, au pré-test comme au post-test. Elles ont été effectuées d'octobre à décembre et d'avril à juin la première année, ainsi que d'avril à juin la deuxième année pour l'étude de suivi. Chaque élève a été observé·e durant des sessions de 5 min environ une fois par semaine pendant 10 semaines. Les observations ont été réalisées dans un ordre aléatoire. Les personnes codant les comportements des élèves ont suivi un entraînement de 200h et ont dû atteindre un certain critère ($k=.70$) avant de pouvoir réaliser les observations. Ils et elles n'étaient au courant ni des hypothèses de l'étude ni des conditions expérimentales. Une période d'habitation a été mise en place avant le commencement réel des observations. Les données ont été collectées en utilisant un programme sur un mini-ordinateur portable. Chaque comportement a été codé selon un système de catégories mutuellement exclusives. Les données ont été agrégées à travers les 10 sessions et des sous-catégories ont été combinées pour les analyses afin d'obtenir des estimations plus fidèles et une distribution normale des données. Les cinq codes composites sont les suivants :

- **Harcèlement** : Ce code inclut les comportements d'agression physique, verbale et indirecte soit impliquant un déséquilibre détectable de pouvoir entre l'agresseur et sa cible, soit répétés durant la même session.
- **Encouragement du harcèlement** : Ce code inclut les comportements de rire ou d'encouragements verbaux face à une situation de harcèlement ainsi que l'observation passive et soutenue d'une situation de harcèlement
- **Aggression ne relevant pas du harcèlement** : Ce code inclut les comportements d'agression physique ou indirecte qui ne sont pas répétés et n'ont pas lieu dans le cadre d'une relation de pouvoir déséquilibrée.
- **Comportement prosocial** : Ce code inclut les comportements ou mots neutres ou positifs envers un autre enfant (ex. initier une conversation).
- **Comportement social argumentatif** : Ce code inclut les comportements ou mots négatifs, mais non agressifs envers un autre enfant (ex. ignorer quelqu'un, décider de tout).
- **Commérages diffamatoires** : Ce code inclut les discussions désobligeantes sur une troisième personne (ex. « *Is the cootie girl in your class ?* », « *Did you hear Dan cheated on the exam ?* »).

Annexe 5 – Evaluation de *Second Step*

Les conversations positives ou neutres concernant une troisième personne ne sont pas incluses dans cette catégorie.

Les interventions des adultes et élèves visant à mettre fin à une situation de harcèlement ont également été codées, mais n'étaient pas assez fréquentes pour être analysées. Des tests de fiabilité ont été effectués sur les autres variables et les kappas obtenus sont consistants avec ceux obtenus dans des recherches précédentes sur l'agressivité. Pour le harcèlement, l'encouragement du harcèlement, l'agression ne relevant pas du harcèlement et les commérages diffamatoires, les données sont rapportées en termes de nombre d'événements par heure. Pour le comportement prosocial et le comportement social argumentatif, elles sont rapportées en termes de pourcentage de temps. Les analyses ont été réalisées sur le score de différence entre le pré et le post-test (afin d'avoir davantage une distribution normale).

Des consultant·e·s du programme sont régulièrement venu·e·s observer les enseignant·e·s lors des leçons de compétences. Plusieurs personnes ont parfois observé les mêmes leçons, afin d'augmenter fiabilité des mesures, et chaque enseignant·e a été observé lors de cinq leçons (les scores à chaque leçon étant moyenné pour former un score global). Le but de ces observations était d'évaluer ce que Hirschstein et al. (2007) appellent les composantes « talk » du programme (c'est-à-dire l'enseignement formel). Les composantes « walk » du programme (c'est-à-dire le transfert de ce qui est appris à la vie quotidienne) ont été évaluées via le *Social-Emotional Learning Checklist-Bullying Report* (Hirschstein, Edstrom & Frey, 2000 ; cité par Hirschstein et al., 2007) que les enseignant·e·s ont dû remplir une fois par mois. La réponse à chaque item se fait via une échelle à quatre points allant de *Jamais* à *Quatre fois ou plus la semaine passée*. Une analyse factorielle a permis d'identifier trois dimensions : le soutien à la généralisation des compétences socioémotionnelles, le soutien à la généralisation des compétences en lien avec la prévention du harcèlement et le coaching des élèves impliqué·e·s dans des situations de harcèlement. Les mesures d'implantation qui ont été prises sont donc les suivantes :

- **Adhérence (talk)** : L'adhérence a été mesurée lors de l'observation des leçons via une checklist des activités effectuées (ex. « *Teacher defines bullying* »). La fiabilité inter juge est acceptable ($\kappa=.81$).
- **Qualité des leçons (talk)** : La qualité de différents aspects des leçons (clarté des explications, réactivité des participant·e·s, gestion de la classe et ton émotionnel) a été évaluée via une échelle à 3 points (pauvre, bonne ou excellente). Un score global a ensuite été calculé. La fiabilité inter juge est acceptable ($\kappa=.62$).
- **Soutien à la généralisation des compétences socioémotionnelles (walk)** : Le soutien à la généralisation des compétences socioémotionnelles a été mesuré via 5 items (ex. « *I intervene in a student conflict by prompting students to manage their emotions.* »). L'alpha est égal à .79.
- **Soutien à la généralisation des compétences en lien avec la prévention du harcèlement (walk)** : Le soutien à la généralisation des compétences en lien avec la prévention du harcèlement a été mesuré via 5 items (ex. « *I prompted students to stand up for someone being picked on.* »). L'alpha est égal à .78.

- **Coaching des élèves impliqué-e-s dans des situations de harcèlement (walk)** : Le coaching des élèves impliqué-e-s dans des situations de harcèlement a été mesuré via 4 items (ex. « *I checked in with a student who had been bullied to see how things were going* »). L'alpha est égal à .85.

Effets de *Steps to Respect* selon l'évaluation de 2000-2001

1. Des effets ont-ils été observés ?

Le programme a un effet sur les croyances auto-rapportées. Après l'intervention, par rapport aux élèves des écoles contrôles, ceux et celles des écoles expérimentales acceptent moins le harcèlement, se sentent plus responsables d'intervenir lorsque leurs ami-e-s harcèlent d'autres élèves, et rapportent une plus grande réactivité de la part des adultes. Les différences entre groupes sont dues au fait que, dans le groupe expérimental, les croyances n'ont pas changé alors qu'elles se sont dégradées dans le groupe contrôle. L'effet du programme est seulement marginalement significatif concernant la difficulté à répondre de manière assertive. Toutefois l'interaction entre le groupe et le niveau scolaire est significative : les élèves de la condition expérimentale trouvent moins difficile de répondre de manière assertive que ceux de la condition contrôle mais uniquement lorsqu'ils et elles sont en 5^{ème} et 6^{ème} année. Les tailles des effets sont présentées dans le tableau 1.

Le programme n'a ni d'effet sur le comportement auto-rapporté, qu'il s'agisse du harcèlement direct, indirect ou de la victimisation. Il n'a pas non plus d'effet sur les compétences d'interactions sociales des élèves évaluées par l'enseignant-e.

Concernant les analyses réalisées sur les données observationnelles, le programme a eu un effet significatif sur le harcèlement dans le sens attendu. Cet effet interagit de façon marginalement significative avec le harcèlement au pré-test : parmi les élèves auteur-e-s de harcèlement au pré-test, ceux et celles du groupe expérimental harcèlent moins au post-test que ceux et celles du groupe contrôle. Pour l'encouragement du harcèlement, un effet marginalement significatif allant dans le sens des hypothèses est observé. Concernant la victimisation et l'agression ne relevant pas du harcèlement, il n'y a aucun effet de l'intervention. Le programme a par contre un effet sur les interactions sociales. Les comportements prosociaux augmentent et les comportements argumentatifs diminuent dans le groupe expérimental alors qu'ils restent stables dans le groupe contrôle.

Pour ce qui est des commérages diffamatoires (Low et al., 2010), le programme a un effet, mais uniquement pour la perpétuation et uniquement chez les élèves qui comméraient au pré-test : un déclin entre le pré et le post-test est observé chez les élèves du groupe expérimental par rapport à ceux et celles du groupe contrôle.

Annexe 5 – Evaluation de *Second Step*

Tableau 1. Taille des effets (d de Cohen) sur les variables principales

Questionnaires	
Acceptation du harcèlement/de l'agression	d entre .16 et .19
Responsabilité en tant que témoin	
Perception de la réactivité des adultes	
Difficulté à répondre de manière assertive	d=.28 pour les élèves de 5^{ème} et 6^{ème} n.s. pour les élèves de 3 ^{ème} et 4 ^{ème}
Agression/harcèlement direct	n.s.
Agression/harcèlement indirecte	n.s.
Victimisation	n.s. (mais $p < .10$)
Compétences d'interaction	n.s.
Observations	
Harcèlement	d=.31 pour l'intervention et d=.15 pour la prévention
Encouragement du harcèlement	n.s. (mais $p < .10$)
Victimisation	n.s.
Agression ne relevant pas du harcèlement	n.s.
Commérages diffamatoires (perpétuation)	d=0.53
Commérages diffamatoires (victimisation)	n.s.
Comportement prosocial	d=.22
Comportement social argumentatif	d=.13

Inspiré de Frey et al. (2005, pp. 486-487). Les effets significatifs sont en gras.

En résumé, le programme a un effet, quoique faible, sur les croyances auto-rapportées ainsi que sur certains comportements (mais uniquement au niveau des données observationnelles). Les auteures précisent que, étant donné la prévalence très importante du harcèlement (d'après les données observationnelles, 77% des élèves ont au moins à un moment donné harcelé ou encouragé le harcèlement), même un effet modeste peut avoir des conséquences importantes. Selon leurs calculs, lors d'une récréation de 60 min, dans une classe de 25 élèves, il y a en moyenne 30 actes de harcèlement dans une classe contrôle contre 24 dans une classe expérimentale. Sur une période de 72 jours (de mars à mi-juin), cela représenterait 900 actes d'harcèlement de moins dans la classe expérimentale.

2. Les effets sont-ils différenciés ?

Les effets du genre, du niveau scolaire et des scores au pré-test sur l'efficacité du programme ont été investigués. Une interaction significative entre la condition et les scores au pré-test a été observée pour la perpétuation de commérages diffamatoires et une interaction marginalement significative a été observée pour le harcèlement. Il semble donc que le programme soit davantage efficace pour les élèves auteur·e·s de harcèlement avant l'introduction du programme (effet d'intervention plutôt que de prévention). Le niveau scolaire interagit avec la condition concernant la difficulté à répondre de manière assertive. L'intervention permet de diminuer la perception de difficulté, mais seulement chez les élèves les plus âgés (5^{ème} et 6^{ème} année). Selon les auteures, ce serait parce que répondre de manière assertive au harcèlement

Annexe 5 – Evaluation de *Second Step*

est, de manière générale, plus difficile pour les élèves plus jeunes. Bien que l'interaction entre le genre et la condition ne soit pas significative pour les comportements prosociaux, en observant les intervalles de confiances, il semble qu'il n'y ait pas de différences entre les filles de la condition contrôle et celles de la condition expérimentale, mais qu'il y en ait une chez les garçons (ceux de la condition expérimentale présentant davantage de comportements prosociaux que ceux de la condition contrôle).

Low et al. (2010) ont réalisé des analyses afin d'investiguer la contribution des croyances soutenant le harcèlement et des amitiés sur la perpétuation de et victimisation par commérages diffamatoires. Le fait d'accepter le harcèlement est associé à une augmentation de la victimisation, mais uniquement dans le groupe contrôle. Ce résultat est consistant avec l'hypothèse des auteures selon laquelle le fait d'accepter le harcèlement est associé à une réaction agressive face à la victimisation (ce qui empire généralement la situation) sauf si l'enfant a appris, grâce au programme, d'autres façons de réagir. Également comme prévu par les auteures, le fait d'avoir des ami·e·s solidaires est associé à une diminution de la victimisation, mais uniquement dans le groupe expérimental. Il semble que, pour que l'effet des ami·e·s soit bénéfique, il faut que ceux et celles-ci aient appris à réagir efficacement face au harcèlement.

3. Des mesures de suivi ont-elles été prises ?

Des analyses de suivi effectuées 12 et 18 mois après le pré-test sont rapportées dans l'étude de Frey et al. (2009). Le programme était alors toujours implanté dans les écoles. Cette étude porte uniquement sur les élèves qui étaient en 3^{ème} et 4^{ème} année lors du pré-test.

Concernant les croyances auto-rapportées, un effet est uniquement observé sur la difficulté à répondre de manière assertive. L'acceptation du harcèlement reste stable dans les deux groupes et la responsabilité des témoins ainsi que la réactivité perçue des adultes diminuent autant dans le groupe expérimental que dans le groupe contrôle (composé des élèves de 5^{ème} et 6^{ème} du groupe expérimental au pré-test). Par contre, après 12 et 18 mois d'implantation du programme, la difficulté à répondre de manière assertive a diminué dans le groupe expérimental alors qu'elle est restée stable dans le groupe contrôle. Concernant le comportement auto-rapporté et les compétences sociales, il n'y a aucune différence entre les deux groupes.

Les comportements de harcèlement, victimisation et encouragement du harcèlement observés durant les récréations diminuent chez les élèves qui, au pré-test, présentaient parfois ce type de comportements. La diminution du harcèlement est telle que 18 mois après le pré-test, les scores des élèves qui étaient impliqué·e·s dans les situations de harcèlement au pré-test ne diffèrent plus de ceux des élèves qui ne l'étaient pas. Même en considérant l'ensemble des élèves (impliqué·e·s ou non), la différence entre le groupe contrôle et le groupe expérimental est significative. Le même pattern de résultats est observé pour le comportement agressif et les interactions sociales argumentatives. Par contre, le programme n'a pas d'effet sur les comportements prosociaux.

En conclusion, après deux ans, l'effet du programme sur les comportements observés est renforcé. Par contre, il n'y a plus d'effet sur les croyances (sauf pour la difficulté à répondre de manière assertive) et toujours pas d'effet sur les comportements auto-rapportés. Il semble y avoir à la fois un effet d'intervention et de prévention vu que le harcèlement et la victimisation diminuent chez les élèves impliqués au pré-test et n'augmentent pas chez les autres.

4. L'implantation a-t-elle été mesurée ?

L'étude de Hirschstein et al. (2007) est entièrement consacrée à l'implantation. D'après les enseignant·e·s, à la fin de l'année, les politiques et procédures du programme sont bien implantées au niveau de l'école (3.25 en moyenne sur une échelle allant de 1 à 4). Ils et elles rapportent également avoir implanté les leçons de compétences de façon importante (99,2% en moyenne). D'après les consultantes du programme ayant observé des leçons, l'adhérence est élevée (91% en moyenne) et la qualité des leçons est juste au-dessus du milieu de l'échelle (2,24 en moyenne sur une échelle allant de 1 à 4). Les enseignant·e·s rapportent soutenir la généralisation des compétences socioémotionnelles et des compétences en lien avec la prévention du harcèlement un peu plus d'une fois par semaine (respectivement 1.40 et 1.41 en moyenne) et coacher les élèves environ une fois par semaine (1.06 en moyenne). L'adhérence et la qualité des leçons sont légèrement corrélées tandis que le soutien à la généralisation des compétences socioémotionnelles, celui à la généralisation des compétences en lien avec la prévention du harcèlement et le coaching des élèves impliqué·e·s dans des situations de harcèlement sont fortement corrélés.

Les variables d'implantation prédisent certains effets du programme (les tailles d'effet sont présentées dans le tableau 2). L'adhérence prédit uniquement les compétences interpersonnelles évaluées par les enseignant·e·s : les élèves exposé·e·s à tous les concepts clés et compétences essentielles de *Steps to Respect* sont évalué·e·s comme socialement plus compétent·e·s que les autres. La qualité des leçons prédit la difficulté à répondre au harcèlement et la victimisation auto-rapportée : les élèves ayant reçu un enseignement de haute qualité perçoivent plus de difficulté à répondre au harcèlement et rapportent plus de victimisation que les autres élèves. Ce résultat peut sembler contre-intuitif. Selon les auteures, il est possible que dans les classes où des leçons de haute qualité sont données, le stigma associé à la victimisation soit moindre et les élèves sont plus conscient·e·s du courage et des compétences nécessaires pour intervenir dans des situations de harcèlement. Ces élèves ne seraient en fait pas plus victimes de harcèlement que les autres, mais simplement plus enclin·e·s à rapporter dans le questionnaire la victimisation vécue. Cela expliquerait pourquoi la qualité des leçons prédit la victimisation auto-rapportée mais pas la victimisation observée.

Le soutien à la généralisation des compétences (que ce soit les compétences socioémotionnelles ou en lien avec la prévention du harcèlement) prédit uniquement l'agressivité et uniquement pour les élèves plus âgé·e·s : les élèves de 5^{ème} et 6^{ème} année qui reçoivent un soutien à la généralisation de leurs compétences sont moins souvent observé·e·s adopter un comportement agressif. Le coaching des élèves impliqué·e·s dans des situations de harcèlement prédit la perception de la réactivité des adultes ainsi que l'observation de comportements de victimisation, d'encouragement du harcèlement et d'agression. Les élèves dont l'enseignant·e fait beaucoup de coaching perçoivent une réactivité moins élevée des adultes (mais cet effet est tellement petit qu'il n'a pas d'implication pratique). Ils et elles sont également moins fréquemment observé·e·s être victimes de harcèlement et encourager le harcèlement (mais uniquement s'ils et elles étaient victimes ou encourageaient le harcèlement au pré-test). Et finalement, chez les élèves de 5^{ème} et 6^{ème} année, le fait d'avoir un·e enseignant·e qui coache fortement ses élèves est associé à une diminution de l'agressivité observée.

Annexe 5 – Evaluation de *Second Step*

Tableau 2. Taille des effets (d de Cohen) de différents aspects d'implantation

Aspects « talk »	
Adhérence - compétences interpersonnelles	d=0.30
Qualité des leçons - difficulté à répondre au harcèlement	d=0.44
Qualité des leçons - victimisation auto-rapportée	d=0.36
Aspects « walk »	
Soutien à la généralisation des compétences socioémotionnelles - agressivité	d=1.45
Soutien à la généralisation des compétences de prévention -agressivité	d=0.86
Coaching-réactivité des adultes face au harcèlement	d=0.12
Coaching-victimisation observée	d=0.48
Coaching - encouragement du harcèlement	d=0.50
Coaching - agression ne relevant pas du harcèlement	d=1.15

Inspiré de Hirschstein et al. (2007, pp. 11-14). Seuls les effets significatifs sont présentés.

En conclusion, ce sont particulièrement les différences au niveau de l'aspect « walk » de l'implantation qui prédisent les effets du programme, une implantation plus importante étant associée à une amélioration des comportements dans la cour de récréation.

Qualité de l'évaluation de 2008-2009

Article principal : Brown, E. C., Low, S., Smith, B. H., & Haggerty, K. P. (2011). Outcomes from a School-Randomized Controlled Trial of Steps to Respect: A Bullying Prevention Program. *School Psychology Review*, 40(3), 423–433. doi:10.1037/e550932011-004

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude longitudinale avec un design RCT, les écoles participantes ont été appariées puis réparties aléatoirement entre la condition expérimentale et la condition contrôle. Les leçons ont été dispensées de décembre 2008 à mai 2009. Le pré-test a été réalisé en automne 2008 et le post-test au printemps 2009 (donc à des périodes différentes de l'année).

Des modèles linéaires hiérarchiques ont été utilisés pour tester l'impact du programme sur les variables dépendantes. Pour les mesures prises au niveau des élèves, le modèle hiérarchique tient compte du fait que le temps est niché dans les élèves, les élèves dans les classes et les classes dans les écoles. Le modèle tient compte des variables contrôles que sont l'âge, le genre, l'origine ethnique, le niveau scolaire et les scores obtenus au prétest (au niveau des élèves), le nombre d'élèves par classe (au niveau des classes) et l'aire géographique (au niveau des écoles). Pour les données obtenues auprès des membres du personnel, le modèle utilisé n'a que deux niveaux (temps niché dans les membres du personnel et membres du personnel nichés dans les écoles) et tient compte des variables contrôles que sont le genre, l'origine ethnique, l'âge, le nombre d'années dans l'école et la position (enseignante, administrative ou non académique) au niveau individuel et l'aire géographique au niveau de l'école. Pour le personnel, les données pré et post-test n'ont pas pu être reliées.

2. Quelles étaient les caractéristiques de l'échantillon ?

Des e-mails ont été envoyés à des contacts au niveau de certains districts scolaires de trois aires géographiques de Californie. Les districts en question remplissaient certaines conditions (diversité socio-économique et ethnique importante, lien avec le *Committee for Children*, besoin ou désir important de prévention du harcèlement scolaire et absence d'usage actuel d'un programme de prévention du harcèlement scolaire). 96 districts se sont montrés intéressés. Des membres du *Committee for Children* ont été visiter toutes les écoles éligibles pour discuter de l'étude. Pour participer, il fallait qu'au moins 80% des enseignant·e·s veuillent implanter le programme et acceptent de participer à la formation et que l'école s'engage à implanter le programme comme indiqué lors de la formation et accepte d'être aléatoirement assignée à la condition expérimentale ou contrôle. Le programme était fourni gratuitement à toutes les écoles participantes. Suite à une analyse de puissance, les chercheurs ont décidé de garder 34 écoles. Ces écoles ont été appariées sur base des données du *National Center for Education Statistics* concernant le nombre d'élèves, d'enseignant·e·s, les changements dans le nombre d'élèves, le pourcentage d'élèves ayant droit à un repas de midi gratuit ou à prix réduit, les pourcentages d'élèves issues de diverses origines ethniques et le pourcentage d'élèves dont l'anglais n'est pas la langue maternelle. Une école a quitté l'étude avant le prétest sans pouvoir être remplacée. L'échantillon était donc composé de 33 écoles venant de régions aux différents niveaux d'urbanisation, avec en moyenne 24 enseignant·e·s et 479 élèves par école et 40% d'élèves ayant droit à un repas de midi gratuit ou à prix réduit.

Tout le personnel scolaire était invité à participer à l'étude. 1307 personnes ont répondu au pré-test et 1296 au post-test (soit 77 et 76% de l'équipe au complet). 58% des participant·e·s étaient des enseignant·e·s, 10% des assistant·e·s d'enseignement, 3,3% des membres du personnel de cafétéria, 2,8% des membres de la direction, 1,4% des psychologues ou conseiller·ère·s scolaires, 1,4% des concierges, 0,7% des conducteur·trice·s de bus et 7,7% des personnes occupant une autre fonction. 90% des participant·e·s étaient des femmes et 85% se considéraient Espagnol, Latino ou Hispanique. L'âge moyen était de 46 ans et le nombre moyen d'année de travail dans l'école de 3-5 ans.

Dans chaque école participante, trois ou quatre enseignant·e·s de 3^{ème}, 4^{ème} ou 5^{ème} année ont été sélectionné·e·s pour compléter des questionnaires sur leurs élèves. Cela fait 128 enseignant·e·s (52 de 3^{ème} année, 62 de 4^{ème} année, 11 de 5^{ème} année et 3 de classes composées de plusieurs années). En échange de leur participation, ils et elles ont reçu \$5 par élève ainsi qu'un bonus additionnel de \$25 si les questionnaires étaient rendus en moins de deux semaines. Les enseignant·e·s de la condition expérimentale ont également reçu \$75 s'ils et elles avaient rempli les 11 parties du journal de bord d'implantation avant la fin de l'étude.

Le programme a été dispensé à 3119 élèves. 22 n'ont pas reçu le consentement parental, 173 n'ont pas pu répondre au questionnaire à cause d'une faible maîtrise de l'anglais ou d'un trouble développemental et 14 ont refusé de participer. L'échantillon final est donc composé de 2940 élèves (94%). Environ la moitié sont des garçons (49%), 52% sont blancs, 7% afro-américains, 6% d'origine asiatique et 35% d'une autre origine. 43% des élèves sont d'origine hispanique. L'âge moyen des élèves est de 8,9 ans. Pour leur participation, tou-te-s les élèves ont reçu un cadeau d'une valeur de \$5. Dans l'étude de Low, Van Ryzin, Brown, Smith, & Haggerty (2014), portant sur la qualité de l'implantation du programme, seules les 17 écoles de la condition expérimentale ont été prise en compte. L'échantillon comprend donc 1424 élèves. Au pré-test, les élèves

Annexe 5 – Evaluation de *Second Step*

de la condition expérimentale présentaient des niveaux plus élevés de harcèlement que ceux de la condition contrôle. Pour les 23 autres variables dépendantes, il n’y avait pas de différence entre les conditions.

Les scores aux échelles n’ont pas été calculés s’il y avait plus de 67% d’items manquants. La prévalence de données manquantes était assez faible (moins de 10%), mais elles ont tout de même été prises en compte via des analyses d’imputations multiples.

3. Par qui et où l’évaluation a-t-elle été menée ?

L’évaluation est menée en Californie par Eric Brown et Kevin Haggerty (University of Washington), Sabine Low et Mark Van Ryzin (Wichita State University) et Brian Smith (*Committee for Children*). Certain·e·s des auteur·e·s ont donc déjà participé à la première étude d’évaluation et/ou ont des liens avec le *Committee for Children*. Les écoles de l’échantillon sont plus diversifiées (en termes d’origine ethnique des élèves et d’aire géographique) que ne l’étaient celles dans la première étude.

4. Quelles et combien de mesures ont été prises ?

Blabla Les élèves ont répondu à une version révisée du *Colorado Trust’s Bullying Prevention Student Survey* (Csuti, 2008). Ce questionnaire a permis de mesurer les variables suivantes :

- **Soutien entre élèves** : Le soutien entre élèves a été mesuré à l’aide de 6 items (ex. « *Students my age are there for me whenever I need help.* ») auxquels il est possible de répondre via une échelle à 4 points allant de *Pas du tout vrai* à *Complètement vrai*. L’alpha est égal à .78.
- **Attitudes envers le harcèlement** : Les attitudes envers le harcèlement ont été mesurées à l’aide de 7 items (ex. « *How wrong or okay is it when students push, shove, or pick fights with weaker students?* ») auxquels il est possible de répondre via une échelle à 5 points allant d’*Extrêmement mal* à *Tout à fait acceptable*. L’alpha est égal à .87.
- **Attitudes envers l’intervention vis-à-vis du harcèlement** : Les attitudes envers l’intervention vis-à-vis du harcèlement ont été mesurées à l’aide de 4 items (ex. « *How wrong or okay is it when students defend others who are being shoved around by strong students?* ») auxquels il est possible de répondre via une échelle à 5 points allant d’*Extrêmement mal* à *Tout à fait acceptable*. L’alpha est égal à .79.
- **Prévention du harcèlement** : La prévention du harcèlement a été mesurée à l’aide d’un item (« *Teachers and staff at my school are doing the right things to prevent bullying.* ») auxquels il est possible de répondre via une échelle à 4 points allant de *Désaccord complet* à *Accord complet*.
- **Intervention des élèves face au harcèlement** : L’intervention des élèves face au harcèlement a été mesurée à l’aide de 4 items (ex. « *Students in your school would help out if a student or group of students pushed, shoved or tried to pick a fight with a weaker student.* ») auxquels il est possible de répondre via une échelle à 4 points allant de *Jamais* à *Toujours*. L’alpha est égal à .76.

Annexe 5 – Evaluation de *Second Step*

- **Intervention des adultes face au harcèlement** : L'intervention des adultes face au harcèlement a été mesurée à l'aide de 4 items (ex. « *Teachers and staff in your school would help out if a student spread rumors or lies about another student behind their back.* ») auxquels il est possible de répondre via une échelle à 4 points allant de *Jamais* à *Toujours*. L'alpha est égal à .85.
- **Comportement positif des témoins** : Le comportement positif des témoins a été mesuré à l'aide de 5 items (ex. « *I tried to defend students who always get pushed or shoved around.* ») auxquels il est possible de répondre via une échelle à 5 points allant de *Jamais* à *Beaucoup*. L'alpha est égal à .69.
- **Problèmes liés au harcèlement** : Les problèmes liés au harcèlement ont été mesurés à l'aide de 7 items (ex. « *How big of a problem in your school is students picking fights with other students?* ») auxquels il est possible de répondre via une échelle à 4 points allant de *Pas du tout un problème* à *Un problème énorme*. L'alpha est égal à .87.
- **Harcèlement** : Le harcèlement a été mesuré à l'aide de 7 items (ex. « *I teased or said mean things to certain students.* ») auxquels il est possible de répondre via une échelle à 5 points allant de *Jamais* à *Toujours*. L'alpha est égal à .87.
- **Victimisation** : La victimisation a été mesurée à l'aide de 4 items (ex. « *A particular student or group of students spread rumors or told lies about me.* ») auxquels il est possible de répondre via une échelle à 5 points allant de *Jamais* à *Toujours*. L'alpha est égal à .75.
- **Climat parmi les élèves** : Le climat parmi les élèves a été mesuré à l'aide de 4 items (ex. « *Students in my school can be trusted* ») auxquels il est possible de répondre via une échelle à 4 points allant de *Désaccord complet* à *Accord complet*. L'alpha est égal à .68.
- **Liens avec l'école** : Les liens avec l'école ont été mesurés à l'aide de 5 items (ex. « *This is a pretty close-knit school where everyone looks out for each other.* ») auxquels il est possible de répondre via une échelle à 4 points allant de *Désaccord complet* à *Accord complet*. L'alpha est égal à .76.
- **Climat avec le personnel** : Le climat avec le personnel a été mesuré à l'aide de 4 items (ex. « *Teachers and staff in my school usually get along with students.* ») auxquels il est possible de répondre via une échelle à 4 points allant de *Désaccord complet* à *Accord complet*. L'alpha est égal à .82.

Les attitudes des élèves envers le harcèlement et envers les interventions face au harcèlement ainsi que la perception de prévention du harcèlement ont été considérées comme des variables ordinales étant donné leur distribution non normale. Les scores de harcèlement ne se distribuant également pas du tout normalement ont été transformés en variable dichotomique (avec les réponses jamais opposées aux autres).

Les données ont été récoltées auprès du personnel de l'école pendant la formation ou durant des réunions à l'aide du *School Environment Survey*. Il s'agit d'un bref questionnaire papier/crayon adapté de Csuti (2008 ; cité par Brown et al., 2011). Ce questionnaire a permis de mesurer les variables suivantes :

Annexe 5 – Evaluation de *Second Step*

- **Politique et stratégies anti-harcèlement de l'école** : La politique et les stratégies anti-harcèlement ont été évaluées à l'aide de 8 items (ex. « *How much is your school doing to demonstrate administrator commitment and leadership to address bullies, bullied, and bystanders?* ») auxquels il est possible de répondre via une échelle à 4 points (allant de *Rien du tout* à *Beaucoup*). L'alpha est égal à .93.
- **Intervention des élèves face au harcèlement** : L'intervention des élèves face au harcèlement a été évaluée à l'aide de 5 items (ex. « *Students in this school would help out if a student is making fun of or teasing another student who is obviously weaker* ») auxquels il est possible de répondre via une échelle à 7 points (allant de *Jamais* à *Toujours*). L'alpha est égal à .92.
- **Intervention du personnel face au harcèlement** : L'intervention du personnel face au harcèlement a été évaluée à l'aide de 5 items (ex. « *Teachers and staff in this school would help out if a student or group of students is trying to pick a fight with a weaker student* ») auxquels il est possible de répondre via une échelle à 7 points (allant de *Jamais* à *Toujours*). L'alpha est égal à .95.
- **Climat parmi les élèves** : Le climat parmi les élèves a été évalué à l'aide de 4 items (ex. « *Students in this school are willing to help other students* ») auxquels il est possible de répondre via une échelle à 4 points (allant de *Désaccord complet* à *Accord complet*). L'alpha est égal à .82.
- **Climat scolaire** : Le climat scolaire (relation entre les membres du personnel, avec les élèves, avec les parents...) a été évalué à l'aide de 7 items (ex. « *Staff in this school can be trusted* ») auxquels il est possible de répondre via une échelle à 4 points (allant de *Désaccord complet* à *Accord complet*). L'alpha est égal à .91.
- **Problèmes liés au harcèlement** : Les problèmes liés au harcèlement ont été évalués à l'aide de 8 items (ex. « *How big of a problem in your school is students spreading rumors or lies about students they are mad at or don't like?* ») auxquels il est possible de répondre via une échelle à 4 points (allant de *Pas de problème du tout* à *Un énorme problème*). L'alpha est égal à .82.

Les enseignant·e·s ont répondu en ligne au *Teacher Assessment of Student Behavior* pour chacun·e de leurs élèves. Ce questionnaire a permis de mesurer les variables suivantes :

- **Compétences sociales** : Les compétences sociales ont été mesurées à l'aide de 5 items (ex. « *Gets along with classmates* ») auxquels il est possible de répondre via une échelle à 5 points allant de *Jamais* à *Toujours*. L'alpha est égal à .82.
- **Compétences académiques** : Les compétences académiques ont été mesurées à l'aide de 4 items (ex. « *Needs help to stay on task* ») auxquels il est possible de répondre via une échelle à 5 points allant de *Jamais* à *Toujours*. L'alpha est égal à .86.
- **Accomplissement académique** : L'accomplissement académique a été mesuré à l'aide de 3 items (ex. « *Academically, how would you rate this student in terms of reading?* ») auxquels il est possible de répondre via une échelle à 5 points allant de *Doit beaucoup s'améliorer* à *Se situe au-dessus de la moyenne*. L'alpha est égal à .95.

Annexe 5 – Evaluation de *Second Step*

- **Harcèlement physique** : Le harcèlement physique a été mesuré à l'aide de 4 items (ex. « *Pushed, shoved, or tripped a weaker student* ») auxquels il est possible de répondre via une échelle à 5 points allant de *Jamais* à *Toujours*. L'alpha est égal à .91.
- **Harcèlement non physique** : Le harcèlement non physique a été mesuré à l'aide de 4 items (ex. « *Spread rumors about another student.* ») auxquels il est possible de répondre via une échelle à 5 points allant de *Jamais* à *Toujours*. L'alpha est égal à .80.

Les enseignant·e·s de la condition expérimentale ont également complété un journal de bord d'implantation. Ils et elles devaient l'avoir rempli à la fin de chaque semaine durant laquelle une leçon avait été dispensée. Les mesures principales étaient les suivantes :

- **Adhérence** : Pour chaque leçon, les enseignant·e·s devaient cocher sur la liste des activités de la leçon celles qu'ils et elles avaient effectuées. Un pourcentage moyen a ensuite été calculé pour les leçons 4 à 11 (les autres données ont été perdues). C'est ce pourcentage qui représente le degré d'adhérence.
- **Degré d'engagement des élèves** : L'engagement des élèves dans les leçons est évalué via 4 items (« *To what extent were your students engaged by this lesson (e.g., asking questions, volunteering)?* », « *To what extent were students distracting other students during this lesson?* », « *To what extent do you feel your students could demonstrate the objectives of this lesson?* » et « *How difficult was it to manage students during this lesson?* »). Pour chaque item, les enseignant·e·s ont dû répondre sur une échelle à 4 points allant de *Pas du tout* à *Beaucoup*. Les scores entre leçons et entre items ont été moyennés pour former un score global. L'alpha est égal à .86.
- **Implantation des politiques et procédures du programme** : Afin d'évaluer l'implantation des politiques et procédures du programme au niveau de l'école, les enseignant·e·s ont dû juger la qualité de cette implantation sur une échelle à 4 points allant de *Pauvre* à *Excellente*.

Les enseignant·e·s ont également dû répondre à des questions concernant le pourcentage d'élèves présent·e·s à chaque leçon, le temps passé à préparer et dispenser les leçons, les modifications apportées aux leçons et les activités ajoutées pour renforcer les leçons.

Effets de *Steps to Respect* selon l'évaluation de 2008-2009

1. Des effets ont-ils été observés ?

Au niveau des questionnaires remplis par les élèves, des effets significatifs sont observés pour le climat parmi les élèves, la prévention du harcèlement, l'intervention des élèves face au harcèlement, l'intervention des adultes face au harcèlement et les comportements positifs des témoins. Les élèves des écoles expérimentales, par rapport à ceux et celles des écoles contrôles, rapportent un déclin significativement moins important de la prévention du harcèlement ainsi qu'une augmentation plus importante de

Annexe 5 – Evaluation de *Second Step*

l'intervention des élèves et des adultes face au harcèlement et des comportements positifs de la part des témoins. De plus, alors que le climat parmi les élèves est plus positif au pré-test qu'au post-test dans les écoles contrôles, l'inverse est observé dans les écoles expérimentales. Par contre, aucun effet n'est observé sur le soutien entre élèves, les attitudes envers le harcèlement, les attitudes envers l'intervention vis-à-vis du harcèlement, les problèmes liés au harcèlement, le harcèlement, la victimisation, les liens avec l'école et le climat avec le personnel.

Au niveau des questionnaires remplis par les enseignant·e·s, le programme a des effets sur les compétences sociales des élèves et sur le harcèlement physique. Alors qu'un déclin de compétences sociales est observé chez les élèves des écoles contrôles, il n'y a pas de changement dans les écoles expérimentales. Le harcèlement physique augmente dans les deux conditions, mais de façon plus importante dans la condition contrôle. Par contre, aucun effet n'est observé sur les compétences académiques, l'accomplissement académique et le harcèlement non physique.

Au niveau des questionnaires remplis par le personnel, un effet est observé sur toutes les variables sauf l'intervention du personnel face au harcèlement. Il y a, dans les écoles expérimentales par rapport aux écoles contrôles, une plus grande augmentation des politiques et stratégies anti-harcèlement, du climat positif entre élèves ainsi que du climat scolaire. Il y a également une diminution plus faible des interventions de la part des élèves face au harcèlement et une diminution plus importante des problèmes liés au harcèlement. Les tailles des effets significatifs sont rapportées dans le tableau 3.

En conclusion, des effets positifs sont observés sur la moitié des variables dépendantes. Les effets se situent à différents niveaux de l'école et touchent des variables proximales comme des variables distales. Toutefois, la grande majorité des effets sont de petite taille ($>.30$). De plus, le programme n'a pas d'effet sur le harcèlement et la victimisation auto-rapportés.

Annexe 5 – Evaluation de *Second Step*

Tableau 3. Taille des effets (d de Cohen et Adjusted Odd Ratio) sur les variables d'intérêt

Elèves	
Soutien entre élèves	n.s.
Attitudes envers le harcèlement	n.s.
Attitudes envers l'intervention vis-à-vis du harcèlement	n.s.
Prévention du harcèlement	OR=1.27
Intervention des élèves face au harcèlement	d=0.12
Intervention des adultes face au harcèlement	d=0.13
Comportement positif des témoins	d=0.14
Problèmes liés au harcèlement	n.s.
Harcèlement	n.s.
Victimisation	n.s.
Climat parmi les élèves	d=0.19
Lien avec l'école	n.s.
Climat avec le personnel	n.s.
Enseignants	
Compétences sociales	d=0.13
Compétences académiques	n.s.
Accomplissement académique	n.s.
Harcèlement physique	OR=0.61
Harcèlement non physique	n.s.
Membres du personnel	
Politique et stratégies anti-harcèlement de l'école	d=0.38
Intervention des élèves face au harcèlement	d=0.28
Intervention du personnel face au harcèlement	n.s.
Climat parmi les élèves	d=0.21
Climat scolaire	d=0.26
Problèmes liés au harcèlement	d=-0.35

Inspiré de Brown et al. (2011, pp. 436-438). Les effets significatifs sont en gras.

2. Les effets sont-ils différenciés ?

Les auteurs ne rapportent aucune analyse permettant de savoir si les effets sont différenciés.

3. Des mesures de suivi ont-elles été prises ?

Aucune mesure de suivi n'a été prise.

4. L'implantation a-t-elle été mesurée ?

A la fin de l'année, les enseignant·e·s considèrent que les politiques et procédures du programme sont bien implantées au niveau de l'école (3.25 en moyenne sur une échelle allant de 1 à 4). 83% des enseignant·e·s

Annexe 5 – Evaluation de *Second Step*

rapportent avoir enseigné au moins 80% des leçons et 91% au moins 60%. Ils et elles rapportent avoir enseigné 99,2% des leçons de compétence en moyenne. Approximativement 75% des élèves ont été exposé·e·s à au moins 95% des leçons (et seuls 3% des élèves ont été exposé·e·s à moins de 75% du programme). La participation des élèves au programme est évaluée comme haute par les enseignant·e·s. 18% des enseignant·e·s ont rapporté avoir omis un élément ou plus d'une leçon, toutefois, le pourcentage d'enseignant·e·s rapportant avoir complété tous les objectifs est élevé (92%).

L'étude de Low et al. (2014) porte entièrement sur l'implantation du programme. Leurs analyses portent d'abord les prédicteurs de l'implantation au niveau de la classe. Les variables qui prédisent l'engagement des élèves sont le soutien entre élèves, les attitudes anti-harcèlement des élèves, le climat parmi les élèves et les liens avec l'école (relation positive) ainsi que les problèmes liés au harcèlement d'après le personnel et la proportion d'élèves ayant droit à un repas de midi gratuit ou à prix réduit (relation négative). Lorsque tous ces prédicteurs sont rassemblés dans un modèle final, seuls les effets du climat parmi les élèves et du pourcentage d'élèves ayant droit à un repas de midi gratuit ou à prix réduit restent significatifs (ils expliquent à eux deux 20% de la variance dans l'engagement des élèves). Seules les attitudes anti-harcèlement des élèves permettent de prédire l'adhérence (celle-ci est plus élevée lorsque les élèves ont davantage d'attitudes anti-harcèlement).

Au niveau de l'impact de l'implantation sur les effets du programme, l'adhérence n'est associée à aucune variable, peut-être parce qu'elle varie peu d'un·e enseignant·e à l'autre. Par contre, un degré d'engagement élevé des élèves est associé à des niveaux plus élevés d'attitudes anti-harcèlement, d'attitudes positives des élèves envers les interventions face au harcèlement et de climat positif parmi les élèves. Un degré d'engagement élevé est aussi associé à un niveau plus faible de victimisation (et, de façon marginalement significative, à des niveaux plus élevés de liens avec l'école et de climat positif avec le personnel ainsi qu'un niveau plus faible de harcèlement).

En conclusion, l'engagement des élèves est prédit par le climat parmi les élèves ainsi que par un indice du niveau de pauvreté de la communauté dans laquelle se trouvent les écoles et permet de prédire certains effets du programme, un engagement plus important étant associé à des bénéfices plus grands.

Implantation de *Second Step*

Environ 32000 écoles au travers des États-Unis ont implanté *Second Step* depuis 1987 et, depuis 2004, près de 8 millions d'élèves et 2 millions d'adultes ont participé au programme (NRPP, n.d.). Il est également largement utilisé au Canada et a été adapté pour être utilisé en Australie, Allemagne, Autriche, Suisse, Luxembourg, Nouvelle-Zélande, Norvège, Suède, Royaume-Uni, Danemark, Japon, Turquie, Irak, Chili et Venezuela (Frey, Hirschstein, & Guzzo, 2000; Holsen, Iversen, & Smith, 2009, KidsMatter, n.d.; Schick & Cierpka, 2010). Ce n'est toutefois pas parce qu'il est largement diffusé que le programme est forcément efficace. Le site Crime Solution répertorie plusieurs versions du programme. La troisième version de *Second Step* (2002) est considérée comme « efficace », mais la quatrième version (2011) est considérée comme n'ayant pas d'effets (Crime Solution, 2012, 2017).

Qualité de l'évaluation de 2010-2011 dans les écoles secondaires

Articles :

- Espelage, D. L., Low, S., Polanin, J. R., & Brown, E. C. (2013). The impact of a middle school program to reduce aggression, victimization, and sexual violence. *Journal of Adolescent Health, 53*(2), 180-186. doi:10.1016/j.jadohealth.2013.02.021
- Espelage, D. L., Low, S., Polanin, J. R., & Brown, E. C. (2015a). Clinical trial of Second Step© middle-school program: Impact on aggression & victimization. *Journal of Applied Developmental Psychology, 37*(1), 52-63. doi:10.1016/j.appdev.2014.11.007
- Espelage, D. L., Low, S., Ryzin, M. J. Van, & Polanin, J. R. (2015b). Clinical Trial of Second Step Middle School Program: Impact on Bullying, Cyberbullying, Homophobic Teasing, and Sexual Harassment Perpetration. *School Psychology Review, 44*(4), 464–479. doi:10.17105/spr-15-0052.1
- Espelage, D., Rose, C., & Polanin, J. (2015). Social-Emotional Learning Program to Reduce Bullying, Fighting, and Victimization Among Middle School Students With Disabilities. *Remedial and Special Education, 36*(5), 299-311. doi:10.1177/0741932514564564

1. Quel était le plan de recherche de l'évaluation ?

Il s'agit d'une étude longitudinale avec un design RCT. Les écoles ont été appariées et, dans chaque pair, une école a été aléatoirement assignée à la condition expérimentale et l'autre à la condition contrôle. Les écoles de la condition contrôle étaient placées sur une liste d'attente de 3 ans. Elles avaient toutefois la possibilité, durant ces trois ans, d'implanter un autre programme anti-harcèlement, *Stories of Us* (Faull, Jimerson, Swearer & Espelage, 2008 ; cité par Espelage et al., 2015a). Seule une école de la condition contrôle a choisi de l'implanter et elle a arrêté après un an (Espelage et al., 2015a).

Les mesures ont été prises avant l'implantation du programme en automne 2010 (temps 1), à la fin de la première année d'implantation au printemps 2011 (temps 2), à la fin de la deuxième année d'implantation au printemps 2012 (temps 3) et à la fin de la troisième année d'implantation au printemps 2013 (temps 4).

Annexe 5 – Evaluation de *Second Step*

L'étude de Espelage et al. (2013) évalue les effets du programme au temps 2, celle de Espelage et al. (2015a) au temps 3 et celles de Espelage et al. (2015b) et Espelage, Rose et Polanin (2015) au temps 4.

Une analyse *Intent-to-Treat* a été réalisée : tou·te·s les participant·e·s de la condition expérimentale ont été considéré·e·s comme ayant reçu le traitement même si ce n'était pas forcément le cas. Les données ont été analysées via des modèles linéaires hiérarchiques. Au niveau de l'élève, l'âge, le genre, l'origine ethnique et les scores pré-test ont été pris en compte. Au niveau de l'école, la condition, l'état, le pourcentage d'élèves ayant droit à un repas gratuit ou à prix réduit et la taille de l'échantillon dans l'école ont été pris en compte.

Dans l'étude d'Espelage et al. (2015a), les analyses ont été faites sur toutes les écoles puis séparément pour chaque état. Pour l'étude d'Espelage et al. (2015b) les analyses ont été réalisées en deux parties. Tout d'abord, l'effet direct du programme sur les variables dépendantes a été examiné. Ensuite, des analyses ont été réalisées pour investiguer les effets indirects du programme sur les variables clés en fonction des changements dans les trajectoires individuelles de délinquance entre les temps 1, 2 et 3. Comme il n'était pas prévu que le programme ait un effet significatif sur les mesures de harcèlement et de victimisation, il ne s'agit pas vraiment d'une analyse de médiation, la délinquance est plutôt considérée comme une variable d'intervention (les effets du programme sur la délinquance et de la délinquance sur les variables de harcèlement sont supposés être significatifs sans que l'effet du programme sur les variables de harcèlement le soit). Pour l'étude de Espelage, Rose et Polanin (2015), les analyses ont uniquement été réalisées au niveau des élèves étant donné la petite taille de l'échantillon.

2. Quelles étaient les caractéristiques de l'échantillon ?

Des membres de certains districts d'Illinois et du Kansas ont été contactés afin de leur présenter le projet et d'identifier les écoles qui seraient intéressées par l'étude. 39 écoles ont été identifiées dont une a été exclue et deux autres ont refusé de participer. Les écoles participantes ont dû accepter d'être aléatoirement assignées à la condition contrôle ou expérimentale et de ne pas implanter d'autres programmes de prévention du harcèlement durant les trois ans de l'étude. Les 36 écoles ont été appariées avant d'être aléatoirement réparties entre la condition contrôle et la condition expérimentale. Dans chaque état, les écoles ont été appariées sur base du nombre d'élèves, de l'évolution du nombre d'élèves entre 2008 et 2009, du pourcentage d'élèves ayant droit à un repas de midi gratuit ou à prix réduit, des pourcentages d'élèves de différentes origines ethniques et du pourcentage d'élèves dont l'anglais n'est pas la langue maternelle. Ces données étaient issues du *National Center for Educational Statistics*.

La condition contrôle était composée de 12 écoles de l'Illinois et 6 du Kansas comprenant 2074 élèves en 6^{ème} année dont 80% (soit 1676 élèves) ont reçu un consentement parental pour participer à l'étude. La condition expérimentale était également composée de 12 écoles de l'Illinois et 6 du Kansas, comprenant 2341 élèves dont 83% (soit 1940 élèves) ont reçu le consentement parental. Au temps 2, aucune école n'avait abandonné l'étude, mais 228 élèves de la condition contrôle et 222 de la condition expérimentale n'ont plus participé à l'étude. Une école de la condition expérimentale et deux écoles de la condition expérimentale n'ont plus pu être suivies après le temps 2 et deux écoles supplémentaires de la condition contrôle n'ont plus pu être suivies après le temps 3. De plus, une école de la condition expérimentale a mis fin à sa participation à l'étude avant le temps 4. Au total, entre le pré-test et le temps 4, sur les 3616 participant·e·s, 704 élèves de la condition contrôle et 645 de la condition expérimentale ont été perdu·e·s.

Annexe 5 – Evaluation de *Second Step*

Les élèves avaient environ 11 en moyenne. 48% étaient des filles et un peu plus de 70% avaient droit à un repas gratuit ou à prix réduit. Il y avait dans la condition contrôle, un peu plus d'élèves hispaniques (36,6% par rapport à 32% dans la condition expérimentale) et moins de blanc·he·s (20,6 % par rapport à 28,2% dans la condition expérimentale). Mis à part cela, il n'y avait aucune différence significative entre les deux groupes.

Pour l'étude d'Espelage, Rose, & Polanin (2015), les analyses ont été réalisées sur un sous-échantillon d'élèves présentant une déficience. Tou-te-s les élèves présentant une déficience, peu importe de quel type (trouble émotionnel, trouble spécifique de l'apprentissage, trouble de santé...), étaient inclus-es dans l'étude, ce qui fait un total de 123 élèves (47 dans la condition contrôle et 76 dans la condition expérimentale) issu-e-s de 12 écoles.

Les données manquantes ont été prises en compte via une procédure d'imputations multiples.

3. Par qui et où l'évaluation a-t-elle été menée ?

L'évaluation a été menée par Dorothy Espelage (University of Illinois), Sabine Low (Arizona State University), Marc Van Ryzin (University of Oregon) Joshua Polanin (Loyola University Chicago) et Eric Brown (University of Washington). Certain-e-s de ces auteur-e-s avaient participé aux évaluations de *Steps to Respect*.

4. Quelles et combien de mesures ont été prises ?

A chaque temps de mesure, deux chercheur·euse·s ont récolté les données dans les classes pendant les heures scolaires. Si des élèves étaient absent·e-s le jour de la récolte des données, une session de rattrapage était organisée. Les participant·e-s étaient identifié·e-s par un code afin de pouvoir relier leurs réponses aux différents temps de mesure tout en assurant la confidentialité des données. Après avoir été informé·e-s de la nature générale de l'étude, les élèves avaient environ 40 min pour compléter le questionnaire. Ils et elles recevaient un surligneur pour leur participation. Chaque item du questionnaire était lu à voix haute.

Les mesures qui ont été prises sont les suivantes :

- **Harcèlement verbal/relationnel** : Le harcèlement verbal/relationnel a été évalué via les neuf items de l'*University of Illinois Bully Scale* (Espelage & Holt, 2001 ; cité par Espelage et al., 2013) qui mesure la fréquence des moqueries, des insultes, de l'exclusion sociale et des rumeurs. Les élèves doivent indiquer combien de fois, durant les 30 derniers jours, ils ou elles se sont moqué·e-s d'autres élèves, ont ennuyé d'autres élèves pour s'amuser, ont exclu d'autres élèves de leur groupe d'ami·e-s... Les options de réponse sont *Jamais*, *Une ou deux fois*, *Trois ou quatre fois*, *Cinq ou six fois* et *Plus de sept fois*. Suivant les temps de mesure, les alphas se situent entre .76 et .86.
- **Victimisation** : La victimisation par les pairs a été évaluée via les trois items de l'*University of Illinois Victimization Scale* (Espelage & Holt, 2001 ; cité par Espelage et al., 2013). Les élèves ont dû indiquer combien de fois durant les 30 derniers jours ils ou elles avaient été insulté·e-s, embêté·e-s ou frappé·e-s par d'autres élèves. Les options de réponse sont *Jamais*, *Une ou deux fois*, *Trois ou quatre fois*, *Cinq ou six fois* et *Plus de sept fois*. Suivant les temps de mesure, les alphas se situent entre .85 et .93.

Annexe 5 – Evaluation de *Second Step*

- **Agression physique** : L'agression physique a été évaluée via les quatre items de l'*University of Illinois Fighting Scale* (Espelage & Holt, 2001 ; cité par Espelage et al., 2013). Cette échelle mesure les comportements de conflit physique (ex. « *I fought students I could easily beat* ») durant les 30 derniers jours. Les options de réponse sont *Jamais*, *Une ou deux fois*, *Trois ou quatre fois*, *Cinq ou six fois* et *Plus de sept fois*. Suivant les temps de mesure, les alphas se situent entre .70 et .81.
- **Perpétration d'insultes homophobes** : La perpétration d'insultes homophobes a été évaluée via la partie perpétration de l'échelle de 10 items *Homophobic Content Agent Target Scale* (Poteat & Espelage, 2007 ; cité par Espelage et al., 2013). Cette échelle évalue la perpétration et victimisation par insultes homophobes durant les 30 derniers jours. L'échelle est introduite par cette phrase « *Some kids call each other names homo, gay, lesbo, fag, or dyke. How many times in the last 30 days did YOU say these words to...* ». Les élèves ont dû dire combien de fois ils et elles avaient dit ces mots à un·e ami·, quelqu'un qu'ils ou elles n'aimaient pas, quelqu'un qu'ils ou elles ne connaissaient pas bien, quelqu'un qu'ils ou elles pensaient gay et quelqu'un qu'ils ou elles ne pensaient pas gay. Les options de réponse sont *Jamais*, *Une ou deux fois*, *Trois ou quatre fois*, *Cinq ou six fois* et *Plus de sept fois*. Suivant les temps de mesure, les alphas se situent entre .80 et .84.
- **Victimisation par insultes homophobes** : La victimisation par insultes homophobes a été évaluée via la partie victimisation de l'échelle de 10 items *Homophobic Content Agent Target Scale* (Poteat & Espelage, 2007 ; cité par Espelage et al., 2013). Les questions sont exactement les mêmes, mais formulées de façon passive. Suivant les temps de mesure, les alphas se situent entre .80 et .81.
- **Harcèlement sexuel** : Le harcèlement sexuel a été évalué via une version modifiée de l'*American Association of University Women Sexual Harassment Scale* (Espelage et al., 2012 ; cité par Espelage et al., 2013) qui mesure la fréquence de comportements de violence sexuelle effectués dans l'année écoulée. L'échelle de harcèlement contient 10 items mesurant des comportements tels que faire des commentaires sexuels, répandre des rumeurs en lien avec la sexualité, tirer les vêtements de façon sexuelle, embrasser une personne sans son accord... Les options de réponse sont *Jamais*, *Une à trois fois*, *Quatre à neuf fois* et *Plus de dix fois*. Suivant les temps de mesure, les alphas se situent entre .75 et .83.
- **Victimisation sexuelle** : La victimisation sexuelle a été évaluée via une version modifiée de l'*American Association of University Women Sexual Harassment Scale* (Espelage et al., 2012 ; cité par Espelage et al., 2013). L'échelle de victimisation contient les mêmes items que celle de harcèlement, mais formulés de manière passive. Suivant les temps de mesure, les alphas se situent entre .80 et .84.
- **Cyberharcèlement** : Le cyberharcèlement a été évalué via une échelle de quatre items basée sur le travail d'Yberra, Espelage et Mitchell (2007 ; cité par Espelage et al., 2015b). Il est demandé aux élèves combien de fois dans l'année écoulée ils ou elles ont fait des commentaires désobligeants à quelqu'un en ligne, rependu des rumeurs sur quelqu'un en ligne, fait des commentaires agressifs ou menaçants à quelqu'un en ligne et envoyé des messages méchants. Les options de réponse sont

Annexe 5 – Evaluation de *Second Step*

Jamais, Une à trois fois, Quatre à neuf fois et Plus de dix fois. Les alphas sont égaux à .74 au pré-test et .76 au temps 4.

- **Délinquance** : La délinquance non violente a été évaluée via les huit items de la *Problem Behavior Frequency Scale* issue du *Multisite Violence Prevention Project* (Dahlberg, Toal, Swahn & Behrens, 2005 ; cité par Espelage et al., 2015b). Il est demandé aux élèves de rapporter à quelle fréquence ils ou elles s'engagent dans des comportements problématiques tels que se faire renvoyer, voler un·e autre élève ou dans un magasin, entrer dans un endroit sans payer, sécher l'école, tricher à un test, abimer des biens ou faire des graffiti. Les options de réponse sont *Jamais, Une à trois fois, Quatre à neuf fois et Plus de dix fois*. Suivant les temps de mesure, les alphas se situent entre .78 et .84.

Vu leur distribution, les variables *Agression physique, Harcèlement verbal/relationnel, Victimisation, Perpétration d'insultes homophobes* et *Victimisation par insultes homophobes* ont été dichotomisés sur base de cut-point théorique. Les élèves ont été considéré·e·s comme auteur·e ou victime d'harcèlement à partir du moment où ils et elles n'avaient pas répondu *Jamais* à au moins deux items (ou à au moins un item pour la victimisation et le harcèlement sexuel, tous deux très peu fréquent).

Les enseignant·e·s ont également complété un journal de bord d'implantation après chaque leçon. Les mesures principales étaient l'adhérence au programme et l'engagement des élèves, mais ces données n'ont pas été exploitées dans les analyses.

Effets de *Second Step* selon l'évaluation de 2010-2011 dans les écoles secondaires

1. Des effets ont-ils été observés ?

Après un an, le seul effet du programme concerne l'agression physique. Par rapport aux élèves de la condition contrôle, les élèves de la condition expérimentale présentent une diminution de l'agression physique telle qu'ils et elles ont une probabilité 42% fois moindre de rapporter avoir perpétré une agression physique. Par contre, le programme n'a pas d'effet sur la victimisation (due au harcèlement, par insultes homophobes ou sexuelles), sur le harcèlement verbal/relationnel, la perpétration d'insultes homophobes et le harcèlement sexuel.

Après deux ans d'implantation du programme, l'effet sur l'agression physique n'est plus significatif. De plus, il n'y a toujours pas d'effet du programme sur le harcèlement verbal/relationnel, la victimisation, la perpétration d'insultes homophobes ou la victimisation sexuelle. Par contre, pour la victimisation par insultes homophobes et le harcèlement sexuel, bien qu'il n'y ait pas d'effet significatif de l'intervention pour le modèle général, il existe une interaction significative entre la condition et l'état. Dans les deux cas, le programme a un effet significatif en Illinois, mais pas au Kansas. Dans les écoles d'Illinois, suite à l'intervention, les probabilités d'être victime d'insultes homophobes ou d'être auteur·e de harcèlement sexuel sont respectivement 56,3% et 38,8% plus faibles dans les écoles ayant eu l'intervention (voir tableau 4).

Annexe 5 – Evaluation de *Second Step*

Tableau 4. Taille des effets (Unadjusted Odd Ratio) après deux et trois ans d'implantation

Temps 2		
Harcèlement verbal/relationnel		OR=.91 (.79-1.05)
Victimisation		OR= .96 (.84-1.10)
Agression physique		OR=.69 (.60-.79)
Perpétration d'insultes homophobes		OR=1.00 (.86-1.15)
Victimisation par insultes homophobes		OR=1.00 (.86-1.15)
Harcèlement sexuel		OR=.94 (.81-1.11)
Victimisation sexuelle		OR=.96 (.84-1.10)
Temps 3		
Harcèlement verbal/relationnel		OR=.80 (.65-.99)
Victimisation		OR=1.22 (.99-1.47)
Agression physique		OR=.89 (.73-1.08)
Perpétration d'insultes homophobes		OR=1.08 (.88-1.33)
Victimisation par insultes homophobes	Kansas	OR=1.13 (.91-1.43)
	Illinois	OR=0.66 (.54-.81)
Harcèlement sexuel	Kansas	OR=1.05 (.85-1.27)
	Illinois	OR=0.70 (.58-.85)
Victimisation sexuelle		OR=1.11 (.92-1.36)

Inspiré de Espelage et al. (2015a, pp. 59-60) et Espelage et al. (2013, p. 184) Les résultats significatifs sont en gras et les intervalles de confiance (à 95%) sont indiqués entre parenthèses.

Après trois ans d'implantation du programme, la victimisation a été laissée de côté, les analyses ont uniquement été réalisées pour la perpétration de violence. Que ce soit pour le harcèlement verbal/relationnel, la perpétration d'insultes homophobes, le harcèlement sexuel ou le cyberharcèlement, il n'y a pas d'effet du programme. Toutefois, il y a un effet significatif du programme sur la trajectoire de la délinquance. Les élèves de la condition expérimentale montrent des niveaux plus bas de délinquance au cours du temps que ceux et celles de la condition contrôle. Et, en passant par l'évolution de la délinquance, le programme a un effet indirect significatif sur le harcèlement verbal/relationnel, la perpétration d'insultes homophobes et le cyberharcèlement (mais pas sur le harcèlement sexuel).

En conclusion, après les trois ans d'implantation du programme, des réductions dans différents comportements de harcèlement sont trouvées via des réductions dans l'augmentation des comportements délinquants. Ces données soutiennent l'idée d'un modèle en cascade où une suppression de l'augmentation de la délinquance est associée à une progression plus faible des comportements de harcèlement, sur lesquels il est plus compliqué d'agir directement (Espelage et al., 2015b).

2. Les effets sont-ils différenciés ?

Chez les élèves présentant une déficience, après trois ans d'implantation du programme, il y a un effet significatif de la condition sur le harcèlement, mais pas sur la victimisation ou l'agression physique. Par rapport aux élèves de la condition contrôle, les scores de harcèlement des élèves de la condition

Annexe 5 – Evaluation de *Second Step*

expérimentale diminuent significativement entre les quatre temps de mesure. Il y a donc un effet direct du programme chez ces élèves, alors que ce n'était pas le cas chez les autres. Cela pourrait s'expliquer par le fait que le programme ciblerait bien leurs déficits possibles au niveau des compétences sociales et communicationnelles, déficits qui seraient à l'origine de l'implication dans des situations de harcèlement.

Les effets étaient également différenciés en fonction de l'état dans lequel se situent les écoles avec, pour certaines variables, des effets observés en Illinois, mais pas au Kansas.

3. Des mesures de suivi ont-elles été prises ?

Les mesures ont été prises à 5 moments distincts. L'efficacité du programme est uniquement montrée après trois ans d'implantation du programme, mais, comme la manière d'analyser les données qui permet de montrer des résultats n'avait pas été utilisée dans les deux premières études, il n'est pas possible de savoir si l'efficacité du programme varie avec le temps.

4. L'implantation a-t-elle été mesurée ?

Aucune analyse n'est rapportée concernant les données d'implantation du programme.

Autres études portant sur *Second Step*

Le programme général *Second Step* (sans l'unité de prévention du harcèlement) a été évalué dans un nombre assez important d'études. Au niveau préscolaire, McMahon, Washburn, Felix, Yakin et Childrey (2000) ont montré, suite à l'implantation de *Second Step*, une amélioration des connaissances des enfants ainsi que de leurs comportements d'après des observations réalisées dans les écoles. Il n'y avait par contre pas d'effet au niveau des évaluations du comportement des élèves par leurs enseignant-e-s et il s'agissait d'une étude à petite échelle sans groupe contrôle. Upshur et al. (2017) ont montré dans une étude RCT que la dernière version du programme (*Second Step Early Learning*, 2011) permettait une amélioration des fonctions exécutives. Par contre, au niveau des compétences socioémotionnelles, l'effet de *Second Step* était seulement marginalement significatif (d'après les auteures un effet est plus difficile à observer au niveau des compétences socioémotionnelles parce qu'elles sont travaillées dans toutes les classes préscolaires, contrairement aux fonctions exécutives).

Concernant les écoles primaires et secondaires, plusieurs évaluations de *Second Step* ont également été réalisées. La version de 1990 du programme (destinée uniquement aux écoles du secondaire inférieur) a été évaluée par Orpinas, Parcel, Mcalister et Frankowski (1995) ainsi que par Orpinas et al. (2000). Orpinas et al. (1995) ont remarqué des effets positifs concernant les connaissances, attitudes et comportements des élèves. Mais ces effets étaient uniquement présents dans certaines classes et ne se maintenaient pas dans le temps. Orpinas et al. (2000) ont investigué les effets de *Second Step* en combinaison avec d'autres composantes (création d'un Conseil d'école de promotion de la santé, formation de pairs médiateurs, formation des enseignant-e-s à la gestion de conflit et newsletters envoyés aux parents) dans le cadre du projet *Students for Peace*. Aucun effet significatif de l'intervention n'a été retrouvé (il y a même une

Annexe 5 – Evaluation de *Second Step*

tendance allant dans le sens contraire des hypothèses). Les auteur·e·s expliquent ce résultat par la difficulté liée à l'implantation effective du programme.

La version de 1992 du programme (destinée uniquement aux écoles primaires) a été évaluée par Grossman et al. (1997 ; cité par Frey et al., 2000), Taub (2001) et Edwards, Hunt, Meyers, Grogg et Jarrett (2005). Grossman et al. (1997 ; cité par Frey et al., 2000) ont observé dans leur étude RCT à grande échelle un effet positif du programme au niveau des observations comportementales sans que cela se traduise par un effet positif dans les évaluations des enfants par leurs enseignant·e·s et par leurs parents. Taub (2001) a trouvé un effet positif du programme, au niveau des observations comportementales comme des évaluations des enseignant·e·s, mais uniquement concernant les comportements prosociaux (pas d'effet sur les comportements antisociaux). La version du programme testée par Edwards et al. (2005) était légèrement modifiée afin de cibler le harcèlement et pas uniquement l'agression en générale. Des effets ont été observés sur les connaissances des élèves et certaines mesures auto-rapportées (stratégies de coping vis-à-vis du harcèlement et autodétermination) ainsi que sur les évaluations des enseignant·e·s du comportement prosocial des élèves. Toutefois les effets étaient de très petite taille et aucun groupe contrôle n'était inclus dans l'étude.

La version de 1997 du programme (primaire et secondaire) a été évaluée par Sprague et al. (2001), par Edstrom, Frey et Baland (2002), par McMahon et Washburn (2003) et par Frey, Nolen, Edstrom et Hirschstein (2005b). Sprague et al. (2001) ont évalué les effets de *Second Step* en combinaison avec *Effective Behavioral Support* (un système de formation, d'assistance technique et d'évaluation du climat et de la discipline scolaire) dans des écoles primaires et secondaires. Un effet positif du programme a été trouvé concernant le nombre d'envois au bureau de discipline, mais pas concernant la sureté de l'école d'après le conseil de management (composé d'administrateur·trice·s, d'enseignant·e·s et de parents). Edstrom et al. (2002) ont retrouvé un effet positif du programme uniquement au niveau de l'exclusion sociale chez les élèves exposé·e·s à la première année du programme. Chez les élèves exposé·e·s à la deuxième année du programme, et dont les enseignant·e·s rapportent avoir dispensé davantage de leçons, il y a un effet positif au niveau de l'agressivité (que ce soit pour l'exclusion sociale, l'agression physique ou l'agression verbale) comme au niveau des difficultés liées aux compétences sociales. McMahon et Washburn (2003) ont évalué l'effet du programme sur des élèves issu·e·s d'une communauté très défavorisée présentant un taux important d'actes criminels violents. Un effet positif du programme a été observé sur les connaissances des élèves, mais, concernant les niveaux d'agression et les comportements prosociaux (rapportés par les pairs et les enseignant·e·s) ainsi que le sentiment d'appartenance à l'école (auto-rapporté), des effets positifs du programme sont uniquement présents dans une des deux écoles (il n'y avait pas de groupe contrôle dans l'étude). L'étude RCT de Frey, Nolen, et al. (2005b) a montré un effet positif de *Second Step* au niveau des comportements anti et prosociaux évalués par les enseignant·e·s ainsi que certains effets positifs concernant les buts et la satisfaction des élèves dans un jeu du Dilemme du prisonnier et concernant l'agressivité, la satisfaction et la justification de cette satisfaction dans une situation de négociation. Il n'y avait par contre pas d'effets sur les réactions des élèves face à des scénarios de provocations ambiguës, sur les issus du jeu du Dilemme du prisonnier et sur la coopération dans la tâche de négociation.

La version de 2002 de *Second Step* a été évaluée par Cooke et al. (2007) à l'échelle d'une ville dans son entièreté (le programme a été implanté dans toutes les écoles de la ville, il n'y avait donc pas de groupe

Annexe 5 – Evaluation de *Second Step*

contrôle dans l'étude). Des changements positifs ont été observés concernant certaines mesures auto-rapportées mais pas toutes (sans qu'un pattern clair de résultats ne se dégage). Une amélioration était également présente au niveau des observations comportementales pour les comportements antisociaux (mais pas pour les comportements prosociaux). Aucun effet n'était retrouvé concernant le nombre d'envois au bureau de discipline.

La dernière version du programme pour les écoles primaires (*Second Step for Elementary School*, sorti en 2012) a été évaluée par Low, Cook, Smolkowski et Buntain-Ricklefs (2015). Le programme n'a d'effet sur aucune des nombreuses variables prises en compte (que ce soit par mesure auto-rapportée ou par observations comportementales), mis à part un effet positif sur l'hyperactivité.

Second Step a également été évalué en Allemagne et en Norvège. La version allemande du programme se nomme *Faustlos* (ce qui signifie « sans poing ») et a été évaluée par Bowi, Ott, & Tress (2008) et Schick & Cierpka (2006, 2005). Dans l'étude de Schick et Cierpka (2005) de mesures auto et hétéro-rapportées (par les parents) ont été prises. Le programme avait un effet positif sur l'anxiété, la dépression, la crainte de perdre le contrôle et l'acceptation par les pairs, mais aucun effet sur les nombreuses autres variables (dont les problèmes sociaux, les comportements agressifs, les compétences sociales, l'empathie...). Dans l'étude de Schick et Cierpka (2006), qui ciblait les enfants de maternelles, des changements positifs du programme ont été observés au niveau des mesures auto-rapportées et au niveau des observations comportementales, mais pas au niveau des mesures hétéro-rapportées (que ce soit par les enseignant·e·s ou par les parents). L'étude de Bowi et al. (2008) a montré un effet positif de *Faustlos* au niveau des réactions choisies par les enfants face à des scénarios imaginaires. Suite au programme, les élèves de la condition expérimentale choisissent moins de réactions agressives et font davantage preuve d'empathie. Holsen, Smith et Frey (2008) ont examiné l'efficacité de *Steg for Steg* (« Pas à pas »), la version norvégienne du programme. Les compétences sociales des élèves ont augmenté suite au programme, mais il n'a eu aucun effet sur les comportements internalisés et externalisés. Sur base des mêmes données, Holsen et al. (2009) ont investigué les effets du programme en fonction du statut socio-économique des élèves. Les effets étaient plus importants pour les élèves ayant un statut socio-économique moins élevé. Larsen et Samdal (2012) ont évalué les effets de *Steg for Steg* sur les perceptions d'enseignant·e·s de primaire et de maternelle qui implantaient le programme depuis plusieurs années. Il est ressorti des entretiens que les enseignant·e·s avaient une attitude positive envers le programme dont ils et elles voyaient les intérêts. De plus, leurs propos suggéraient que *Steg for Steg* avait une influence positive sur des aspects variés de leur pratique, telle que la résolution de conflits, la collaboration avec les parents, les interactions avec les élèves, la façon dont ils et elles voyaient leur rôle d'enseignant·e et la démocratie dans la classe. Toutefois, l'échantillon était biaisé, les enseignant·e·s n'ayant pas voulu utiliser le programme ou l'ayant abandonné ne pouvaient pas être inclus·es dans l'étude. De plus, il n'y a aucune donnée observationnelle qui viendrait soutenir les dires des enseignant·e·s.

Résumé des effets de *Second Step*

L'efficacité de l'unité de prévention du harcèlement n'a pas encore été testée. Il est indiqué sur le site *Second Step* qu'une étude sur cette unité est actuellement menée par Espelage et Little. Il faudra attendre les résultats, qui ne sont pas encore publiés, avant de pouvoir conclure quant à son efficacité. Cependant, comme elle reprend toutes les composantes de *Steps to Respect*, on peut supposer que ces effets seront semblables à ceux de ce programme. Des effets positifs de *Steps to Respect* ont été retrouvés dans les deux études évaluant son efficacité. Cependant, le programme ne semble pas avoir d'effet sur la victimisation et le harcèlement auto-rapportés et la majorité des effets observés sont de petites tailles. Par contre des effets sont bien présents sur les comportements observés et sont plus importants après une deuxième année d'implantation du programme. Il ne faut en effet pas oublier que les effets du programme ont été évalués après un ou deux ans d'implantation alors que les leçons *Steps to Respect* sont prévues pour être dispensées sur trois ans. Il est donc possible que les effets d'une implantation à long terme soient plus importants. A priori, le programme semble efficace indépendamment des caractéristiques des élèves. Cela n'a toutefois pas été testé en détail. Il semble également que le programme agisse tant au niveau de la prévention qu'au niveau de l'intervention. Une implantation de qualité semble être nécessaire pour que le programme soit efficace. Les aspects identifiés comme particulièrement prédictifs des effets du programme sont l'engagement des élèves dans les leçons ainsi que les pratiques adoptées par l'enseignant·e en dehors des moments formels de leçons.

Pour les écoles secondaires, la toute dernière version du programme vient de sortir et n'a donc pas encore été évaluée. Il faut donc se baser sur la version de 2008 (qui comprenait déjà une partie spécifique sur le harcèlement) pour extrapoler sur son efficacité. Des effets du programme ont été observés par Espelage et al. (2013), Espelage et al. (2015a), Espelage et al. (2015b) et Espelage, Rose et Polanin (2015). Toutefois, ces effets ne se maintiennent pas d'une année à l'autre (un effet est présent sur l'agression physique après un an, mais ne se maintient pas par la suite), ne sont présents que sur certaines mesures (par exemple seulement concernant les insultes homophobes, mais pas sur la victimisation verbale en général), que dans certaines écoles (des effets sont observés dans les écoles en Illinois, mais pas au Kansas) ou de façon indirecte (il y a un effet indirect du programme sur la délinquance et de la délinquance sur le harcèlement sans aucun effet direct du programme sur le harcèlement) et sont de petite taille.

En plus des résultats mitigés qui en ressortent, les évaluations de *Second Step* ne sont pas toujours réalisées de manière idéale. Des reviewers externes ont réalisé une évaluation de la qualité des recherches portant sur *Second Step* pour le *National Registry of Evidence-based Programs and Practices* de la *Substance Abuse and Mental Health Services Administration* américaine (NREPP, n.d.). Ils et elles ont indépendamment évalué les études en se basant sur 6 critères : la fiabilité des mesures, la validité des mesures, la fidélité d'intervention, les données manquantes et taux d'abandon, les variables confondantes potentielles et la pertinence des analyses. Chaque critère était évalué sur une échelle de 0 à 4 (4 reflétant une qualité élevée et 0 une qualité très faible). Les notes attribuées sont 2.3 pour la fiabilité des mesures, 2.5 pour la validité des mesures, 2.0 pour la fidélité d'intervention et les données manquantes, 2.3 pour les variables confondantes et 3.3 pour la pertinence des analyses ce qui donne une moyenne générale de 2.4. Certaines études ont toutefois l'avantage

d'utiliser des méthodes observationnelles, qui sont généralement peu utilisées étant donné l'investissement important qu'elles demandent. Dans plusieurs études (Grossman et al., 1997; cité par Frey et al., 2000; Low et al., 2010; Schick & Cierpka, 2006) des effets de l'intervention étaient uniquement visibles au niveau des données observationnelles. Il semble donc que ce type de mesure est plus sensible et permet d'observer des effets qu'il serait possible de manquer autrement.

En conclusion, bien qu'il y ait une longue histoire d'évaluation derrière le programme *Second Step*, les résultats de ces évaluations sont plutôt mitigés. L'efficacité du programme est variable d'une étude à l'autre et d'une variable à l'autre au sein d'une même étude, sans qu'aucun pattern clair de résultats ne se dégage (les effets sont parfois observés uniquement au niveau des mesures auto-rapportées, parfois uniquement au niveau des mesures hétéro-rapportées, ils sont observés sur des variables différentes d'une étude à l'autre, ils sont parfois plus importants pour les filles, parfois pour les garçons, au même âge des effets très différents sont observés...). Il est donc difficile de conclure que le programme est efficace, surtout en ce qui concerne le harcèlement et la victimisation qui ne sont pas ces cibles principales.

Annexe 6 : Evaluation de *Zippy's Friends*

Zippy's Friends a été évalué à plusieurs reprises dans différents endroits du monde. Ces évaluations et leurs résultats sont présentés chronologiquement dans la suite de ce chapitre. Elles ne sont pas aussi détaillées que dans les autres chapitres étant donné que des mesures liées au harcèlement sont rarement prises.

Implantation de *Zippy's Friends* en 1998 au Danemark

La version initiale du programme, qui s'appelait alors *Reaching Young Europe*, a été évaluée au Danemark lors de l'année scolaire 1998-1999 par Mishara & Ystgaard (2000 ; cité par Mishara & Ystgaard, 2006). L'évaluation, comparant 214 élèves participant au programme à un groupe contrôle de 109 élèves, a permis de montrer que le programme pouvait être implanté avec succès. Il était apprécié par les enseignant·e·s comme par les élèves, qui y participaient avec plaisir. De plus, par rapport au groupe contrôle, une amélioration plus importante de certaines compétences sociales a été observée dans le groupe expérimental. Par contre, aucun changement n'a été observé au niveau des stratégies d'adaptation, qu'elles soient rapportées par les élèves eux et elles-mêmes ou par les enseignant·e·s (Mishara & Ystgaard, 2000 ; cité par Mishara & Ystgaard, 2006). Sur base de ces résultats, une version révisée du programme a été développée. Cette version met davantage l'accent sur les stratégies d'adaptation (Mishara & Ystgaard, 2006).

Implantation de *Zippy's Friends* en 2000 au Danemark et en Lituanie

La version actuelle de *Zippy's Friends* a été évaluée pour la première fois par Mishara & Ystgaard (2006). Le programme a été implanté parallèlement au Danemark et en Lituanie lors de l'année scolaire 2000-2001. Le but de l'étude était d'évaluer l'efficacité de *Zippy's Friends* dans deux contextes culturels fort différents. 418 élèves lituanien·ne·s de 3^{ème} maternelles (6 ans en moyenne) et 322 élèves danois·es de 1^{ère} primaire (7 ans en moyenne) ont participé à l'étude. En Lituanie, les élèves étaient réparti·e·s entre le groupe contrôle et le groupe expérimental tandis que les résultats des élèves danois·es du groupe expérimental ont été comparés au groupe contrôle de l'étude de 1998. Étant donné leur jeune âge, les élèves ont répondu oralement aux questionnaires standardisés durant des entretiens individuels. Les mesures ont été prises juste avant le début de l'implantation du programme, puis juste après la fin de celui-ci afin de tester son efficacité. De plus, les enseignant·e·s ont rempli des rapports après chaque séance et ont été rencontrés lors d'entretiens en fin d'année afin de suivre l'implantation du programme (Mishara & Ystgaard, 2006).

Les 24 séances du programme ont généralement été complétées comme prévu. Lorsqu'une modification y était apportée, il s'agissait généralement d'une adaptation au niveau du timing (ex. séance divisée en deux parties) ou d'un ajout d'activités (ex. exercices de gymnastiques durant la séance, organisation d'un festin pour clôturer le programme). La majorité des enseignant·e·s estiment que les séances étaient utiles et expliquées de façon suffisamment claire et complète. Les enseignant·e·s rapportent également que les

Annexe 6 – Evaluation de *Zippy's Friends*

élèves ont beaucoup aimé le programme et que, bien qu'ils et elles n'aient pas tou-te-s activement participé à toutes les séances, tou-te-s ont activement participé à une séance au moins (Mishara & Ystgaard, 2006).

Les compétences sociales ont été mesurées à la fois auprès des enseignant-e-s (coopération, affirmation de soi et contrôle de soi) et des élèves (coopération, affirmation de soi, contrôle de soi et empathie). En Lituanie, il y a un effet positif du programme concernant les trois compétences sociales mesurées auprès des enseignant-e-s et concernant le contrôle de soi et l'empathie (mais pas la coopération et l'affirmation de soi) mesurés auprès des élèves. Au Danemark, un effet positif est uniquement observé pour le contrôle de soi mesuré auprès des enseignant-e-s (aucun effet pour les autres compétences ou pour les mesures auto-rapportées). Les comportements problématiques (comportements internalisés, externalisés et d'hyperactivité) ont uniquement été mesurés auprès des enseignant-e-s en Lituanie (car ils n'avaient pas été évalués dans le groupe servant de contrôle au Danemark). L'effet attendu est bien observé, mais uniquement pour les comportements externalisés et d'hyperactivité, pas pour les comportements internalisés. Les stratégies d'adaptation ont également été mesurées auprès des enseignant-e-s et des élèves (qui devaient à la fois dire à quelle fréquence ils et elles utilisaient chaque stratégie et à quel point ils ou elles la perçoivent comme utile, les deux scores étant multipliés en un score global). Un effet positif du programme est observé au niveau des scores provenant des élèves dans les deux pays ainsi qu'au niveau des scores provenant des enseignant-e-s en Lituanie (absence d'effet significatif au Danemark). Les effets du programme sont identiques chez les filles et les garçons. Des analyses ont été réalisées classe par classe et montrent que, bien que des améliorations à plusieurs niveaux soient observées dans toutes les classes, excepté une dans chaque pays, l'étendue de ces améliorations varie d'une classe à l'autre (Mishara & Ystgaard, 2006).

Ces résultats sont encourageants. Toutefois, les effets observés sont de petite taille pour la majorité (mis à part quelques-uns au niveau des mesures hétéro-rapportées). De plus, de nombreuses limites méthodologiques sont présentes dans l'étude de Mishara & Ystgaard (2006), dont les principales sont l'absence d'assignation aléatoire des écoles en Lituanie et de réel groupe contrôle au Danemark et le fait que les modèles statistiques utilisés ne tiennent pas compte de la structure hiérarchique des données.

En Lituanie, des mesures à long terme ont été prises un et neuf ans après l'implantation du programme. La première étude de suivi a été réalisée par Monkevicienė, Mishara, & Dufour (2006). En Lituanie, les écoles primaires et maternelles sont séparées, ce qui fait que tou-te-s les élèves arrivent dans une nouvelle école lors de leur entrée en primaire. C'est l'adaptation à ce changement qui a été investiguée dans l'étude en question. Les enseignant-e-s et parents de 140 élèves ayant participé au programme et à l'étude l'année précédente ainsi que de 106 élèves fréquentant les mêmes classes, mais n'ayant jamais été exposé-e-s à *Zippy's Friends* ont répondu à des questionnaires trois mois après le début de l'année scolaire. Ils et elles ne savaient pas que l'étude portait sur *Zippy's Friends* et les enseignant-e-s n'étaient, à priori, pas au courant de quel-le-s élèves avaient ou non participé au programme l'année précédente. Le premier questionnaire auquel les parents et enseignant-e-s ont répondu concerne l'adaptation comportementale et émotionnelle à l'école. Les élèves du groupe expérimental y ont obtenu des scores plus élevés que ceux et celles du groupe contrôle, que ce soit d'après les réponses des parents ou des enseignant-e-s. Les enseignant-e-s ont également complété un questionnaire concernant les réactions observées chez l'élève dans le nouvel environnement scolaire (réactions aux situations nouvelles, aux routines, aux échecs, aux étrangers, aux

Annexe 6 – Evaluation de *Zippy's Friends*

pairs...). Les scores sont plus positifs pour les élèves de la condition expérimentale que pour ceux et celles de la condition contrôle. Et finalement, les parents comme les enseignant-e-s ont répondu à un questionnaire portant sur les problèmes rencontrés par l'élève (question fermée) et les méthodes utilisées pour y faire face (question ouverte). Ils et elles rapportent presque deux fois plus de problèmes pour les élèves du groupe contrôle que pour les élèves du groupe expérimental. Par rapport à ceux du groupe contrôle, les parents des élèves du groupe expérimental rapportent moins de problèmes concernant les relations et l'égoïsme tandis que les enseignant-e-s rapportent moins de problèmes concernant les relations, l'autorégulation, l'apprentissage, la discipline et l'égoïsme (c'est-à-dire concernant tous les problèmes mesurés sauf la communication et l'estime de soi). De plus, il semble que les élèves du groupe expérimental utilisent davantage de méthodes pour faire face aux problèmes, et plus spécifiquement davantage de méthodes jugées comme appropriées, que ceux et celles du groupe contrôle. Monkeviciene, Mishara, & Dufour (2006) concluent de leur étude que l'exposition à *Zippy's Friends* est liée à une meilleure adaptation lors de la transition entre l'école primaire et maternelle. Toutefois, l'absence de mesure pré-test limite l'interprétation des résultats.

Une seconde étude de suivi a été réalisée en 2009 par Gudauskien (2010). Neuf ans après leur participation à *Zippy's Friends*, 73 élèves, maintenant en 9^{ème} année (15 ans) ont été retrouvés, ont reçu le questionnaire et ont accepté de le compléter. Un questionnaire de 10 questions a été développé. Il tournait autour de trois questions principales: *Do you remember the program Reaching Young Europe, in which the main character was the stick insect named Zippy, which took place in your kindergarten?*, *Has this program helped you to overcome difficulties in everyday life?* et *What do you think about the need for such a program for children?*. 81% des participant-e-s se souviennent du programme, 68% rapportent que *Zippy's Friends* les a aidé-e-s à surmonter des difficultés de la vie de tous les jours et 79% considèrent que les enfants ont besoin d'un tel programme. Les participant-e-s rapportent que *Zippy's Friends* permet d'avoir confiance en soi, de communiquer facilement, d'améliorer ses compétences de contrôle de soi et qu'il est utile pour les jeunes enfants parce qu'il les aide à se comprendre et à comprendre les autres ainsi qu'à développer des compétences comportementales. Le but principal de l'étude était de voir si les élèves se souvenaient du programme neuf ans plus tard, ce qui est effectivement le cas pour la majorité des élèves. Il semble également que la plupart des élèves ont gardé un bon souvenir de *Zippy's Friends* et ont l'impression qu'il leur a été utile, ce qui ne dit toutefois rien sur l'efficacité réelle du programme.

Implantation de *Zippy's Friends* en 2004 au Royaume-Uni (Southampton)

Comme expliqué par Holmes & Faupel (n.d.), *Zippy's Friends* a été implanté à Southampton en plusieurs étapes. Une première étude pilote a été menée de 2003 à 2004 et n'a pas montré de différences significatives entre le groupe contrôle et le groupe expérimental. Le programme a ensuite été implanté dans d'autres classes lors de l'année 2004-2005 puis de l'année 2005-2006. Holmes & Faupel (n.d.) rapportent les résultats de la deuxième année d'implantation. Le programme a été implanté dans 22 classes de 1^{ère}, 2^{ème} ou 3^{ème} année de primaire. Trois autres écoles ont été placées sur liste d'attente et servaient de groupe contrôle. Dans chaque classe, il a été demandé à l'enseignant-e de choisir deux élèves se situant dans la moyenne au niveau du développement et des habilités émotionnels. Les mesures ont été prises chez ces

Annexe 6 – Evaluation de *Zippy's Friends*

élèves (36 dans le groupe expérimental et 24 dans le groupe contrôle), leurs parents et enseignant·e·s juste avant et après l'implantation du programme. De plus, les enseignant·e·s ont dû donner, via 15 items, leur appréciation du programme. Sur une échelle de 1 à 5, les scores moyens varient entre 2.69 et 4.08 suivant les items (avec un score moyen de 3.81 à la question « *Overall, how would you rate Zippy's Friends on a scale of 1-5 with 1= very poor to 5 = excellent ?* »), ce qui reflète une vision plutôt positive du programme.

Les enseignant·e·s ont rempli un questionnaire investiguant cinq dimensions de la littératie émotionnelle : la conscience de soi, l'autorégulation, la motivation, l'empathie et les compétences sociales. Que ce soit pour le score total, pour la conscience de soi ou pour la motivation, alors qu'au pré-test les élèves de la condition contrôle obtenaient des scores plus élevés, il n'y a plus de différence entre les deux groupes au post-test. Pour les autres sous-dimensions, il n'y a aucune différence entre les deux groupes, que ce soit au pré ou au post-test. Les parents ont rempli un questionnaire portant sur les comportements positifs et négatifs de leur enfant. Le pattern de résultats observé est le même : alors que les élèves du groupe contrôle présentaient moins de problèmes au pré-test, il n'y a plus de différences entre les groupes au post-test. Afin de mesurer leur vocabulaire émotionnel, il était demandé aux élèves de décrire comment se sentaient les personnages d'une courte histoire qui leur était lue lors d'un entretien individuel. Une augmentation du nombre de mots simples utilisés est présente dans les deux groupes tandis qu'une augmentation du nombre de mots complexes utilisés est uniquement présente dans le groupe expérimental. Les élèves de 2^{ème} et 3^{ème} année ont également dû, via des jeux de rôles, mimer les réactions qu'ils et elles adopteraient face à 15 situations (ex. entrée dans un groupe de pair, réponse à l'échec ou au succès, réponse à la provocation...). Aucune différence entre les deux groupes n'a été retrouvée.

Holmes & Faupel (n.d.) concluent de leur étude que *Zippy's Friends* a un effet positif sur les élèves, mais d'un point de vue méthodologique et statistique les résultats sont peu convaincants (les auteur·e·s ont apparemment juste comparé les deux groupes au pré puis au post-test et la sélection des élèves est problématique).

Implantation de *Zippy's Friends* en 2004 en Chine (Hong Kong)

Wong (2008) a évalué l'efficacité de *Zippy's Friends* à Hong Kong. Le programme a été implanté dans 10 écoles maternelles auprès de 139 élèves. 128 élèves issu·e·s de classes des mêmes écoles, mais dans lesquelles le programme n'était pas implanté ont servi de groupe contrôle. Juste avant et après l'implantation du programme, les enseignant·e·s et élèves ont répondu à un questionnaire portant sur les stratégies d'adaptation des élèves ainsi que sur l'efficacité de ces stratégies. Des tests de Wilcoxon ont été utilisés pour investiguer les effets du programme. Pour les mesures auto-rapportées, il n'y a aucune différence entre les élèves du groupe expérimental et ceux et celles du groupe contrôle. Par contre, pour les mesures hétéro-rapportées, une différence est observée. Certaines stratégies d'adaptation sont plus utilisées et jugées plus utiles par les élèves du groupe expérimental au post-test qu'au pré-test alors que dans le groupe contrôle l'utilisation ou l'utilité perçue de ces stratégies reste identique ou diminue. Les stratégies d'adaptation qui sont davantage utilisées et/ou jugées utiles sont en grande partie celles apprises dans les leçons Zippy (ex. rester calme, en parler à quelqu'un, réfléchir, dessiner...). Les enseignant·e·s ont également dû donner leurs impressions du programme. Ils et elles rapportent avoir tiré des bénéfices de

Annexe 6 – Evaluation de *Zippy's Friends*

Zippy's Friends, tout comme les élèves qui auraient par ailleurs apprécié le programme et beaucoup participé. Toujours d'après les enseignant·e·s le programme est bien reçu auprès des parents qui rapportent avoir remarqué des changements positifs chez leur enfant.

Implantation de *Zippy's Friends* en 2006 au Canada (Québec)

Zippy's Friends a été implanté et évalué au Québec lors de l'année scolaire 2006-2007 (Denoncourt, 2012). Le programme a été implanté auprès de 310 élèves de 1^{ère}-2^{ème} année issu·e·s de 16 classes dont les enseignant·e·s s'étaient porté·e·s volontaires pour implanter *Zippy's Friends* et participer à l'étude d'évaluation. Les enseignant·e·s de 19 autres classes, comprenant 303 élèves, se sont porté·e·s volontaires pour servir de condition contrôle. Seulement la moitié des élèves environ (190 dans la condition expérimentale et 121 dans la condition contrôle) ont reçu un consentement parental pour participer à l'étude. Les données ont été récoltées en novembre 2006, avant le début d'implantation du programme et en mai 2007 après la clôture de celui-ci. Selon Dufour, Denoncourt & Mishara (2011 ; cité par Denoncourt, 2012), l'implantation s'est déroulée comme prévu, avec peu de variation de fidélité d'une classe à l'autre, et le programme a été apprécié par les élèves comme par les enseignant·e·s.

Dans sa thèse, Denoncourt (2012) rapporte les résultats d'analyses quantitatives (ANOVA) et qualitatives investiguant les effets du programme. Les enseignant·e·s ont complété pour chaque élève un questionnaire évaluant leurs compétences sociales et leurs difficultés d'adaptation. Le questionnaire comporte huit sous-dimensions, trois portant sur l'expression des émotions (déprimé-joyeux, anxieux-confiant et irritable-tolérant), trois portant sur les relations avec les pairs (isolé-intégré, agressif-contrôlé et égoïste-prosocial) et deux portants sur les relations avec l'enseignant·e (résistant coopératif et dépendant autonome). Une évolution différente entre les deux groupes est uniquement observée pour les sous-dimensions liées aux relations avec l'enseignant·e : les élèves du groupe expérimental sont plus coopératif·ve·s et autonomes au post-test alors que ceux et celles du groupe contrôle deviennent plus résistant·e·s et ne changent pas en autonomie. Lors d'entretiens individuels, les élèves ont répondu oralement à un questionnaire mesurant leur perception du soutien social reçu de leurs pairs, leurs ami·e·s proches, leurs parents et leur enseignant·e. La perception de soutien social reçu de la part de l'enseignant·e augmente significativement dans le groupe expérimental alors qu'elle reste stable dans le groupe contrôle. À l'inverse, la perception de soutien social perçu des pairs augmente significativement dans le groupe contrôle alors qu'elle reste stable dans le groupe expérimental. Il n'y a aucune différence entre les groupes concernant le soutien social de la part des parents sauf lorsque seuls les enfants présentant des problèmes socioaffectifs sont pris en compte. Dans ce cas-là, le soutien diminue dans le groupe contrôle et reste stable dans le groupe expérimental. Il n'y a aucun effet du programme sur le soutien social de la part des ami·e·s. Les effets sont dans tous les cas de petite taille (η^2 partiel entre .01 et .02).

Sur base d'un critère de facilité, dix classes ont été choisies pour participer à la partie qualitative de l'étude. 20 élèves de la condition expérimentale et 19 de la condition contrôle ont participé à un entretien individuel avec des questions semi-ouvertes. Les premières questions portaient sur leur réseau social. Les personnes cité·e·s sont sensiblement les mêmes et aussi nombreuses dans les deux groupes. Par contre, les élèves du groupe expérimental expliquent avec plus de précision et nomment davantage de raisons expliquant

Annexe 6 – Evaluation de *Zippy's Friends*

pourquoi ces personnes sont importantes pour eux-elles. Il était ensuite demandé aux élèves de se souvenir d'un moment où une personne de leur entourage s'était sentie en colère/triste/nerveuse/jalouse et d'expliquer comment ils et elles l'avaient compris, puis de se souvenir d'un moment où eux et elles-mêmes s'étaient senti-e-s comme cela et d'expliquer l'aide qu'ils et elles voudraient recevoir et leurs réactions dans ce type de situation. Des différences entre les groupes sont visibles dans la description et l'explication des situations. Certains thèmes sont uniquement ou plus fréquemment abordés par les élèves du groupe expérimental qui donnent souvent des descriptions et explications plus détaillées. Concernant les réactions adoptées face aux situations, les élèves du groupe expérimental ont un discours plus nuancé et plus précis. De plus, ils et elles mentionnent davantage de stratégies et sont plus nombreux à en même temps demander de l'aider et utiliser des stratégies d'eux ou d'elles-mêmes. Finalement, le soutien offert par les élèves a été investigué via deux petits jeux de rôle. Il était demandé aux élèves de réagir face un enfant qui disait être victime de moquerie puis face à un enfant attristé par la mort de sa grand-mère. Par rapport à ceux et celles du groupe contrôle, les élèves du groupe expérimental proposent des stratégies plus actives visant à régler le problème, s'impliquent davantage dans l'aide qu'ils et elles proposent et proposent une plus grande diversité de stratégies.

Implantation de *Zippy's Friends* en 2007 en Norvège

Zippy's Friends a été évalué en 2007-2008 en Norvège via une évaluation au design RCT (Holen, Waaktaar, Lervåg, & Ystgaard, 2012). Les 77 écoles primaires de trois régions de Norvège ont été invitées à y participer. 35 écoles ont accepté et le programme a alors été donné dans toutes leurs classes de 2^{ème} année. Les écoles ont été appariées avant d'être aléatoirement réparties entre la condition contrôle et la condition expérimentale. Les données ont été récoltées auprès de 1315 élèves, 1278 parents et 1210 enseignant-e-s juste avant et juste après l'implantation du programme. L'analyse des données tient compte de leur structure hiérarchique ainsi que de la présence de données manquantes (peu nombreuses). Comme les élèves ne savaient pas tou-te-s lire, les questionnaires leur étant destinés ne contenaient que des dessins (de pouce vers le haut ou vers le bas), signes et nombres. Les questionnaires étaient accompagnés d'un manuel expliquant aux enseignant-e-s les comment faire passer à leur classe. Les élèves et leurs parents ont répondu à un questionnaire sur les stratégies d'adaptation. L'analyse factorielle effectuée sur ce questionnaire a révélé trois facteurs pour la version complétée par les élèves (stratégie active/ régulation émotionnelle, désengagement et opposition) et trois autres facteurs pour celle des parents (recherche d'aide, stratégie active et désengagement). Une amélioration est présente dans les deux groupes et est uniquement plus importante dans le groupe expérimental que dans le groupe contrôle concernant la réduction des stratégies d'opposition d'après les élèves et l'augmentation des stratégies actives d'après les parents. Les enseignant-e-s et les parents ont également rempli un questionnaire évaluant le fonctionnement émotionnel et comportemental des enfants (avec comme sous-dimensions les symptômes émotionnels, les troubles des conduites, l'hyperactivité/inattention, les problèmes relationnels et le comportement prosocial). Ce questionnaire comprend également un supplément sur l'impact de ces problèmes éventuels. Le programme n'a d'effet significatif sur aucune des sous-dimensions. Le seul effet présent concerne l'impact des problèmes d'après les enseignant-e-s (avec une diminution significative dans le groupe expérimental). Des analyses ont ensuite été réalisées en tenant compte du genre et du statut

Annexe 6 – Evaluation de *Zippy's Friends*

socio-économique. Il s'avère alors que les effets positifs sur les stratégies d'adaptation sont uniquement présents chez les filles et chez les élèves avec un statut socio-économique moins élevé et que l'effet positif sur l'impact des problèmes est uniquement présent chez les garçons. Les tailles d'effets varient entre .146 et .380 pour le groupe total et entre .224 et .551 pour les sous-groupes.

Les enseignant·e·s ont également rempli des questionnaires concernant l'implantation du programme et leur satisfaction vis-à-vis de celui-ci. Près de 85% des enseignant·e·s ont rapporté avoir complété toutes les séances (les autres en ayant généralement laissé tomber une ou deux). 13,4% rapportent des déviations mineures avec le manuel. Sur une échelle allant de 1 à 5, la note médiane des enseignant·e·s est de 4 à la fois pour leur satisfaction envers le programme et leur confiance dans le fait qu'il a rempli ces objectifs.

Holen, Waaktaar, Lervåg et Ystgaard (2013) ont également investigué les effets de *Zippy's Friends* sur les performances académiques des élèves et le climat de classe. Les élèves et enseignant·e·s ont rempli trois questionnaires portant sur le climat de la classe ainsi que sur les compétences académiques et l'intégration sociale de l'élève. Une analyse factorielle exploratoire a révélé deux facteurs pour le climat de la classe : le climat en tant que tel et le harcèlement. Il n'y a aucun effet du programme au niveau des mesures auto-rapportées. Par contre, en se basant sur les questionnaires remplis par les enseignant·e·s, le programme est associé à une amélioration du climat de classe, une diminution du harcèlement et une augmentation des compétences académiques, les tailles d'effet étant respectivement de .614, .550 et .415. Le programme n'a pas d'effet sur l'intégration sociale des élèves. Il n'y a aucun effet du genre des élèves ou de leur statut socio-économique.

Implantation de *Zippy's Friends* en 2008 en Irlande

Zippy's Friends a été évalué en Irlande lors de son implantation en 2008-2009 et 2009-2010 (Clarke & Barry, 2010). Le programme a été dispensé sur deux années scolaires (avec 12 séances chaque année) dans le cadre du cours *Social, Personal and Health Education* qui est obligatoire dans toutes les écoles primaires irlandaises (Clarke, Bunting, & Barry, 2014). Il s'agit d'une étude avec un design RCT. 48 écoles défavorisées ont été invitées à participer et les 44 écoles ayant accepté ont aléatoirement été réparties entre une des conditions expérimentales et la condition contrôle (18 classes dans chaque condition, la plupart issues d'écoles différentes). 766 élèves de 1^{ère} puis 2^{ème} année (7-8 ans) ont participé à l'étude. Dans la première condition expérimentale, il était demandé aux enseignant·e·s d'implanter le programme le plus fidèlement possible tandis que dans la seconde il leur était dit d'utiliser le programme comme une ressource pour dispenser le cours de *Social, Personal and Health Education*. Comme il n'y a eu aucune différence d'adhérence aux leçons entre les deux conditions expérimentales, elles ont été fusionnées. Cela suggère que les enseignant·e·s s'investissent dans le programme même s'ils et elles ne sont pas obligé·e·s de le faire (Clarke et al., 2014).

Les données ont été récoltées avant l'implantation de *Zippy's Friends*, juste après et 12 mois plus tard. Les analyses statistiques réalisées tiennent compte de la structure hiérarchique des données et prennent en compte la présence de données manquantes. Les enseignant·e·s ont rempli deux questionnaires, le premier évaluant la littératie émotionnelle (avec comme sous-dimensions la conscience de soi, l'autorégulation, la

Annexe 6 – Evaluation de *Zippy's Friends*

motivation, l'empathie et les compétences sociales) et le second évaluant le fonctionnement émotionnel et comportemental des enfants (avec comme sous-dimensions les symptômes émotionnels, les troubles des conduites, l'hyperactivité/inattention, les problèmes relationnels et le comportement prosocial). L'effet du programme est significatif concernant la littératie émotionnelle avec une amélioration plus importante dans le groupe expérimental que dans le groupe contrôle concernant toutes les sous-dimensions excepté l'empathie. Cet effet positif se maintient 12 mois plus tard. Concernant le fonctionnement émotionnel et comportemental, que ce soit tout de suite après ou 12 mois plus tard, aucun effet positif du programme n'est observé. Le seul effet présent concerne le score de troubles des conduites et va dans le sens opposé des hypothèses (il diminue davantage dans le groupe contrôle). Mis à part l'effet sur la conscience de soi ($\beta=0.35$), les effets significatifs observés sont tous de petite taille (<0.15). Par ailleurs, les effets sont plus importants dans les classes où l'adhérence est plus élevée (Clarke et al., 2014).

Des données ont également été récoltées auprès des élèves (Clarke, Sixsmith, & Barry, 2015). Un sous-échantillon de 9 classes (161 élèves) a été aléatoirement sélectionné pour participer à un workshop au pré et au post-test. Pour investiguer leurs stratégies d'adaptation, les enfants devaient dessiner et décrire en une ligne une situation au cours de laquelle ils et elles s'étaient senti·e·s tristes (1^{er} dessin) ainsi que ce qu'ils et elles pouvaient faire pour se sentir mieux (2^{ème} dessin). Les situations et stratégies d'adaptation ont été classées en plusieurs thèmes afin d'être analysées. Au pré-test, il n'y a pas de différence apparente entre les stratégies d'adaptation utilisées par les élèves du groupe expérimental et par ceux et celles du groupe contrôle. Par contre, au post-test, il semble y avoir des différences entre groupes, les élèves du groupe expérimental adoptant plus souvent des stratégies orientées vers la recherche de solution ou de soutien social que ceux et celles de la condition contrôle. Une activité portant sur les sentiments a également été réalisée à un moment intermédiaire ainsi qu'au post-test. 6 vignettes décrivant une situation étaient lues aux élèves qui devaient expliquer comment l'enfant dans la situation pouvait se sentir. Aux deux temps de mesure, les élèves de la condition expérimentale utilisaient une variété plus importante de mots pour désigner le fait d'être inquiet ou fâché. De plus, au post-test, ils et elles utilisaient un vocabulaire plus large et élaboré pour articuler les sentiments décrits et avaient davantage tendance à donner des causes et solutions possibles aux sentiments que ceux et celles de la condition expérimentale (Clarke et al., 2015).

Clarke, O'Sullivan, & Barry (2010) ont réalisé une étude de cas dans deux des écoles ayant implanté le programme et s'étant portées volontaires. Les deux écoles ont une attitude positive envers le programme, accordent une grande importance à la promotion de la santé mentale à l'école et ont implanté le programme avec une adhérence importante. Des interviews individuelles et discussions de groupes ont été menées auprès de membres du personnel des écoles, de parents et de membres de la communauté locale. De plus, les enseignant·e·s ayant implanté le programme (quatre dans l'école A et un·e dans l'école B) ont complété des questionnaires portant sur le climat de l'école et sur leurs impressions du programme. Bien que les écoles soient toutes les deux considérées comme défavorisées, elles présentent des profils très différents. L'école A est une grande école urbaine catholique. 41% des élèves sont d'origine internationale et un quart des élèves sont issu·e·s de la communauté des gens du voyage et ont des parents souvent peu éduqués, au chômage et qui subissent des discriminations. Il y a également beaucoup de jeunes mères célibataires parmi les parents d'élèves. Les parents sont peu impliqués dans la vie de l'école, il n'y a par exemple pas d'associations de parents et seulement environ un tiers d'entre eux viennent aux réunions qui

Annexe 6 – Evaluation de *Zippy's Friends*

leur sont destinées. Comme l'école est classée comme ayant le statut le plus désavantagé possible, elle reçoit beaucoup de support supplémentaire, dont beaucoup de programmes, ce qui donne un sentiment d'occupation important. De plus, les élèves sont souvent vu·e·s individuellement pendant les heures de cours ce qui fait que leur présence pendant les séances *Zippy* était moins importante que dans l'école B. L'école B est une petite école rurale protestante située dans une région désavantagée au niveau économique, social et éducatif. Un sens de la communauté très important y existe : les familles sont fortement impliquées dans l'école, il y a une appréciation et un soutien mutuel entre les enseignant·e·s et les parents et l'école a beaucoup de contact avec les autres écoles de la région ainsi qu'avec la communauté locale. D'après les retours des enseignant·e·s des deux écoles, après son implantation, le programme était vu de façon beaucoup plus positive dans l'école B que dans l'école A. Bien qu'il ne soit pas possible de savoir si le programme a réellement été plus efficace dans l'école B, il semble que, par certaines de ces caractéristiques (le lien important avec les familles et la communauté par exemple), elle était davantage prête à planter *Zippy's Friends* que l'école A qui aurait eu besoin d'un support plus important pour dépasser certaines barrières (Clarke, O'Sullivan, & Barry, 2010).

Implantation de *Zippy's Friends* en 2009 au Royaume-Uni (Northamptonshire)

Lors de l'année scolaire 2009-2010, plusieurs écoles du Northamptonshire ont implanté *Zippy's Friends* dans le cadre d'un projet ciblant la santé mentale dans les écoles. L'efficacité de cette implantation a été évaluée par Davidson (n.d.). Trois écoles ayant une classe où le programme a été implanté ont récolté des données auprès des enseignant·e·s dans la classe en question ainsi que dans une autre classe où le cours de *Personal, social, health and economic education* était donné comme à son habitude (afin de servir de groupe contrôle). De plus dans deux des trois écoles, des données ont également été récoltées auprès de quelques parents d'élèves. Les enseignant·e·s ont répondu à un questionnaire mesurant l'estime de soi et les parents à un questionnaire mesurant les comportements positifs et négatifs. Dans tous les cas, les données n'ont été récoltées que pour les élèves considéré·e·s comme étant les plus vulnérables. L'échantillon est donc trop petit (23 élèves pour les questionnaires des enseignant·e·s, 7 pour ceux des parents) pour conclure quoique ce soit des données obtenues (qui vont plutôt dans le sens d'une amélioration de l'estime de soi et d'une absence d'effet sur les comportements adoptés). Des données qualitatives ont également été prises (entretiens avec deux enseignant·e·s, questionnaire complété par un·e enseignant·e, observation d'une séance, rencontre avec une classe...). Il ressort de ces données que les élèves comme les enseignant·e·s semblent apprécier le programme et le trouver utile.

Implantation de *Zippy's Friends* en 2011 en Belgique

En 2011-2012 et 2012-2013, l'ASBL Educa Santé a réalisé une « recherche-action » dont le but était d'implanter *Les amis de Zippy* dans certaines écoles primaires de la Fédération Wallonie-Bruxelles et d'évaluer la faisabilité de cette implantation (Educa Santé, 2014). 17 écoles primaires de différentes régions de Wallonie ont participé à l'étude, soit en 2011-2012, soit en 2012-2013. 28 classes ont commencé à implanter le programme et 22 ont été jusqu'au bout. Cela fait un total de 453 élèves, dont 318 en 1^{ère} année, 91 en 2^{ème} et 44 dans l'enseignement spécialisé.

Les données ont été récoltées via des questionnaires, des entretiens individuels, des focus-groupe et des séances plénières. Les enseignant·e·s ont identifié des apports du programme pour eux et elles-mêmes tant sur le plan personnel que professionnel (avec les élèves comme avec les collègues). Ils et elles ont également identifié des apports positifs auprès des élèves. Le programme aurait des bénéfices au niveau des apprentissages scolaires, des apprentissages sociaux, des habilités à exprimer les émotions, à mieux communiquer, à trouver plusieurs solutions face à un problème, à résoudre les conflits et à vaincre le harcèlement. Pour ce dernier point toutefois beaucoup d'enseignant·e·s considèrent qu'ils et elles ne peuvent pas le savoir parce qu'il n'y aurait pas de harcèlement dans leur classe. Des focus-groupes ont été réalisés avec les élèves qui ont dû répondre à quatre questions : « *Que se passe-t-il dans mon corps quand je suis nerveux ?* » « *Comment faire pour se relaxer quand je suis nerveux ?* » « *Mon ami a perdu son chat, il est triste, comment faire pour l'aider ?* » « *À la récréation il y a un grand qui me fait peur, il me dit des choses méchantes et me menace, comment faire pour que cela s'arrête ?* ». Mis à part pour la première question, toutes les classes ont été capables de proposer plusieurs réponses, ce qui serait un élément en faveur de l'efficacité du programme (mais il n'y a eu ni mesure pré-test ni groupe contrôle). La moitié des parents rapportent avoir observé un changement de comportement chez leur enfant suite au programme (dans l'expression des sentiments, la gestion de la colère, la gestion des conflits, le fait de trouver des solutions aux problèmes, les habilités relationnelles ou la gestion des pertes). Les enseignant·e·s et parents semblent donc trouver le programme efficace. Toutefois, aucune mesure valide qui pourrait corroborer ces impressions n'a été prise, ni auprès des élèves, ni auprès des enseignant·e·s ou des parents (il leur a simplement été demandé leur avis sur l'efficacité du programme, ce qui est une mesure très biaisée). Il est impossible de savoir si *Zippy's Friends* a été efficace.

Les enseignant·e·s ont rencontré plusieurs difficultés dans l'implantation du programme. Ces difficultés concernent principalement la gestion du temps (ex. difficile de réaliser toutes les activités prévues en 50 min), la gestion de l'espace physique (ex. impossibilité de créer un coin Zippy), la relation avec les collègues (ex. manque de soutien ou curiosité déplacée), le maintien de la neutralité, la gestion des réactions des enfants et les thèmes abordés (ex. sentiments, harcèlement, mort...). Il semble que l'utilisation du matériel ne posait pas de problème pour la majorité des enseignant·e·s même s'ils et elles auraient aimé avoir davantage de matériel avec des réalisations concrètes à proposer aux élèves et éviter les longs moments d'écoute. Les difficultés rencontrées seraient dues à un manque de maturité et de compétences des élèves, à un manque de pratique et de maîtrise de l'animation et à une difficulté à gérer des situations fortement porteuses de charge émotionnelle. Les conditions qui ont été identifiées par au moins deux enseignant·e·s comme nécessaires à l'implantation de *Zippy's Friends* sont : la possibilité d'organiser un coin Zippy, un groupe d'élèves de petite taille, un niveau intellectuel et de maturité suffisant chez les élèves, une possibilité

Annexe 6 – Evaluation de *Zippy's Friends*

d'adapter le moment et la durée des séances, l'implication des collègues, la présence de dispositifs de soutien et de suivi, le soutien de la direction et l'intégration de *Zippy's Friends* au projet d'école. La majorité des enseignant·e·s sont favorables à la poursuite du programme l'année suivante. Ils et elles estiment qu'il devrait être étendu à l'ensemble des écoles, mais à condition que la participation des enseignant·e·s soit volontaire, que le programme soit plutôt donné en 2^{ème} année, qu'il concerne l'ensemble de l'école et pas juste une année, que les enseignant·e·s puissent bénéficier d'un réseau d'échanges de pratiques, de support de communication et que les directeur·trice·s soient bien informés.

Implantation de *Zippy's Friends* en 2012 aux Pays-Bas

Une évaluation de *Zippy's Friends* avec un design RCT a été réalisée par le Trimbos Institut aux Pays-Bas. Il est expliqué dans leur rapport (The Trimbos Institute, 2016) que les 29 écoles ayant participé à l'étude sont caractérisées par un pourcentage élevé d'élèves issu·e·s de familles au statut socio-économique bas. Environ 40% des élèves ont au moins un de leurs parents qui est né hors Europe. Les classes de 1^{ère} et 2^{ème} année de ces écoles ont été aléatoirement assignées à la condition contrôle (591 élèves) ou expérimentale (586 élèves). L'évaluation a été menée sur deux ans, de septembre 2012 à juillet 2014. Les élèves en 1^{ère} année puis 2^{ème} année ont reçu *Zippy's Friends* deux fois tandis que les élèves en 2^{ème} au début de l'étude ont reçu *Zippy's Friends* puis, en 3^{ème} année, *Apple's Friends*. Les mesures ont été prises avant l'implantation du programme puis après un et deux d'implantation. Selon les journaux de bord d'implantation remplis par les enseignant·e·s, alors que les trois premiers modules ont généralement été bien implantés c'était de moins en moins le cas pour les modules suivants (uniquement 30% des enseignant·e·s ont complété au moins trois des quatre séances du dernier module).

Afin d'évaluer leurs compétences sociales et émotionnelles, les élèves ont été pris·es individuellement pour répondre à des questions portant sur différents scénarios. Toutes les analyses réalisées sont des analyses multiniveaux qui prennent en compte le fait que les données des élèves sont nichées dans les classes. D'après leurs résultats aux questionnaires, une amélioration plus importante des compétences de reconnaissance émotionnelle et des stratégies d'adaptation est présente chez les élèves du groupe expérimental par rapport à ceux et celles du groupe contrôle. Les mesures prises auprès des parents montrent également une amélioration plus importante de la littératie émotionnelle (score global et sous-dimension motivation) dans le groupe expérimental tandis que les mesures prises auprès des enseignant·e·s ne révèlent aucune différence entre les deux groupes. Les parents et enseignant·e·s ont également complété un questionnaire portant sur les comportements internalisés et externalisés. À nouveau, au niveau des comportements rapportés par les enseignant·e·s, il n'y a aucune différence significative entre le groupe contrôle et le groupe expérimental. Par contre, les mesures prises auprès des parents révèlent une amélioration plus importante dans le groupe expérimental au niveau des comportements externalisés d'hyperactivité/inattention et antisociaux/agressifs. Aucun effet n'est observé pour les comportements externalisés égocentriques/explosifs ou les comportements internalisés. L'effet du programme sur la réussite académique a également été investigué, mais il n'y a aucune différence entre les deux groupes aux résultats de tests d'orthographe et de mathématique que les élèves passent deux fois par an aux Pays-Bas. Il n'y a aucune interaction entre les effets du programme et le genre, le niveau éducatif des parents, le

Annexe 6 – Evaluation de *Zippy's Friends*

climat de classe, la composition des classes ou l'adhérence au programme. Par contre, le programme a plus d'effet sur les enfants étant en 2^{ème} année au début de son implantation que sur ceux et celles étant en 1^{ère} année. Cela veut soit dire que *Zippy's Friends* est plus efficace pour les élèves plus âgé·e·s, soit qu'il est plus efficace en combinaison avec *Apple's Friends* que seul.

Implantation de *Zippy's Friends* en 2014 au Royaume-Uni (pour les enfants à besoins spécifiques)

La version de *Zippy's Friends* adaptée pour les élèves à besoins spécifiques a été évaluée au Royaume-Uni lors de l'année scolaire 2014-2015. Selon le rapport final (Lawson, Stenfort Kroese & Unwin, n.d.), le programme est faisable dans des écoles spécialisées et il est bien accepté tant par les élèves que par leurs parents et enseignant·e·s. Le rapport porte sur l'implantation des trois premiers modules du programme auprès de 53 élèves venant d'école de diverses régions d'Angleterre. Les instruments de mesure et les analyses réalisées ne sont pas détaillés dans le rapport de Lawson et al. (n.d.) qui expliquent que des données ont été récoltées auprès des élèves (tâche de reconnaissance émotionnelle), des enseignant·e·s et des parents (questionnaires). Il est difficile de conclure quoi que ce soit des données vu qu'apparemment l'étude ne comportait aucun groupe contrôle, mais il semble que les capacités de reconnaissance émotionnelle des élèves aient augmentées. D'après les mesures prises auprès des enseignant·e·s, le programme est associé à une amélioration de différentes habilités (coopération, communication, assertivité, responsabilité, conscience de soi, gestion des émotions et compétences relationnelles, mais pas motivation et empathie). Toutefois, d'après les mesures prises auprès des parents, aucun changement n'est observé. Les enseignant·e·s ont donné des retours positifs concernant la formation, le soutien reçu et le programme général tout en pointant certains problèmes (certaines parties du programme sont trop complexes, les élèves ont du mal à maîtriser les concepts de jalousie et nervosité et le programme n'est pas adapté pour certains enfants qui présentent par exemple de sévères troubles de la communication ou des troubles autistiques).

Implantation de *Zippy's Friends* en 2015 en République Tchèque

Lors de l'année 2015-2016, *Zippy's Friends* a été implanté et évalué en République Tchèque. Les résultats sont rapportés dans un rapport rédigé par Žufnίček et al. (n.d.). L'évaluation a été réalisée avec un design RCT, 14 écoles ont été aléatoirement réparties entre le groupe contrôle (341 élèves) et le groupe expérimental (466 élèves). Les élèves participant à l'étude étaient en classe préscolaire (uniquement dans le groupe expérimental), 1^{ère}, 2^{ème} ou 3^{ème} année. Les enseignant·e·s ont complété un questionnaire conçu spécialement pour l'étude au pré et au post-test. Ce questionnaire comprend 31 items décrivant des comportements liés à l'autogestion et aux compétences sociales. Les parents des élèves l'ont également complété au post-test. Il n'est pas clairement précisé dans le rapport comment les données ont été analysées, mais, d'après Žufnίček et al. (n.d.), une amélioration plus importante dans le groupe expérimental que dans le groupe contrôle est observée pour un bon nombre des comportements mesurés auprès des enseignant·e·s (il n'y a par contre pas d'effet du groupe pour les mesures prises auprès des parents). Les

Annexe 6 – Evaluation de *Zippy's Friends*

enseignant·e·s ont également jugé le niveau académique des élèves (en les classant parmi quatre catégories : niveau excellent, très bon, dans la moyenne et sous la moyenne). Une amélioration du niveau académique est observée chez les élèves de la condition expérimentale alors que ce n'est pas le cas chez ceux et celles de la condition contrôle. Il semble par ailleurs que le programme soit aussi efficace pour les filles que les garçons et profite également aux élèves à besoins éducatifs spécifiques. Par contre, il semble moins efficace pour les élèves de 3^{ème} année que pour ceux et celles de 1^{ère} et 2^{ème} année. Les élèves de 3^{ème} année ont 8-9 ans, or *Zippy's Friends* est plutôt prévu pour les 5-7 ans.

Résumé des effets de *Zippy's Friends*

Zippy's Friends a été évalué à de nombreuses reprises, et ce dans des contextes variés. Toutefois, la majorité de ces études d'évaluation comportent d'importantes limites méthodologiques ce qui rend difficile toute conclusion sur l'efficacité du programme.

Les études réalisées au Royaume-Uni (Davidson, n.d.; Holmes & Faupel, n.d.; Lawson et al., n.d.) présentent trop de défauts méthodologiques (échantillons très petits, manque de groupe contrôle et/ou sélection par les enseignant·e·s des élèves participant à l'étude) pour que les résultats obtenus soient interprétables. Les autres études portent sur davantage d'élèves (min. 250) et incluent toutes un groupe contrôle, mais les analyses statistiques effectuées ne sont pas les plus appropriées et l'assignation des classes aux conditions n'a pas été faite aléatoirement. Les résultats de ces études sont généralement plutôt positifs avec une augmentation plus importante des stratégies d'adaptation (Denoncourt, 2012; Wong, 2008), des compétences sociales (Žufníček et al., n.d.) ou des deux (Mishara & Ystgaard, 2006) dans le groupe contrôle par rapport au groupe expérimental. Trois évaluations se distinguent par leur qualité méthodologique plus importante : l'évaluation menée en Norvège (Holen et al., 2012, 2013), celle effectuée en Irlande (Clarke & Barry, 2010; Clarke et al., 2014, 2015) et celle effectuée aux Pays-Bas (The Trimbos Institute, 2016). Selon ces études, *Zippy's Friends* a un effet bénéfique concernant les stratégies d'adaptation des élèves et leur compétence en littératie émotionnelle, mais très peu d'effet concernant leur fonctionnement émotionnel et comportemental et aucun concernant leurs performances académiques.

Il semble donc que *Zippy's Friends* puisse avoir un effet bénéfique sur les stratégies d'adaptation des enfants, la cible principale du programme. De plus, lorsque des mesures ont été prises, les effets se maintenaient un an plus tard (Clarke et al., 2014; Monkevicienė et al., 2006). Par contre, les effets obtenus sont dans la grande majorité des cas de petite, voire très petite taille et, dans plusieurs études, ils sont uniquement présents pour les mesures prises auprès des enseignant·e·s, mais pas pour celles prises auprès des élèves ou de leurs parents (Holen et al., 2013; Mishara & Ystgaard, 2006; Wong, 2008; Žufníček et al., n.d.). Seule l'étude de Holen et al. (2013) comporte une mesure du harcèlement. Selon cette étude le programme est associé à une diminution du harcèlement scolaire lorsque celui-ci est mesuré auprès des enseignant·e·s, mais les élèves de la condition expérimentale ne rapportent aucune amélioration. Pour l'instant, *Zippy's Friends* ne peut donc pas être considéré comme un programme efficace pour réduire le harcèlement.