



HAL
open science

La méthode de la "classe puzzle" est-elle efficace pour améliorer l'apprentissage ?

Arnaud Stanczak

► **To cite this version:**

Arnaud Stanczak. La méthode de la "classe puzzle" est-elle efficace pour améliorer l'apprentissage ?. Psychologie. Université Clermont Auvergne [2017-2020], 2020. Français. NNT : 2020CLFAL013 . tel-03170791

HAL Id: tel-03170791

<https://theses.hal.science/tel-03170791>

Submitted on 16 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Clermont Auvergne

U. F. R. Psychologie, Sciences Sociales et Sciences de l'Éducation

Laboratoire de Psychologie Sociale et Cognitive (LAPSCO) - CNRS, UMR 6024

École Doctorale Lettres, Sciences Humaines et Sociales



La méthode de la « classe puzzle » est-elle efficace pour améliorer l'apprentissage ?

Thèse de Doctorat présentée par Arnaud Stanczak

En vue de l'obtention du titre de Docteur en Psychologie spécialisé en Psychologie Sociale

Sous la direction de Céline Darnon et la co-direction de Pascal Huguet

18 Décembre 2020

Membres du jury

Céline BUCHS	Maître d'enseignement et de recherche, Université de Genève	Rapporteure
Fabrizio BUTERA	Professeur, Université de Lausanne	Rapporteur
Benoît GALAND	Professeur, Université Catholique de Louvain	Examinateur
Isabelle RÉGNER	Professeure, Université Aix-Marseille	Examinatrice

Avant-propos

Cette thèse a été financée dans le cadre du volet ProFan « Répondre au besoin de nouvelles compétences pour l'emploi dans la société numérique » de l'action Innovation Numérique pour l'Excellence Éducative du Programme d'Investissements d'Avenir. Adossée à la mission interministérielle Monteil sur le numérique pour l'éducation, l'expérimentation ProFan a pour objectif de promouvoir et de qualifier de nouveaux contextes d'apprentissage et d'enseignement auprès d'élèves de lycées professionnels. ProFan regroupe autour d'un protocole expérimental commun 109 établissements (et des entreprises partenaires), plusieurs milliers d'élèves sur trois années scolaires, des enseignant-es, inspecteur-rices, chef-fes d'établissement, et sept équipes de recherche en France et à l'étranger (LAPSCO UMR 6024 tête de réseau). Ce consortium a généré un grand nombre de données (plusieurs millions) dont le recueil a pris fin au terme de l'année scolaire 2019-2020. Dans cette thèse, nous ne présentons pas de résultats directement issus de ProFan, mais de plusieurs études expérimentales conduites en parallèle auprès de collégiens sur le mécanisme d' « interdépendance positive » testé à beaucoup plus grande échelle au sein de ProFan avec des lycéens. Nos travaux complètent donc ceux conduits au titre de ProFan

Remerciements

Merci. A toutes et tous ceux qui ont rendu ce travail possible. Cette aventure de trois ans n'aurait pas pu voir le jour sans vous, alors encore une fois, merci. Retranscrire mes pensées dans cette première partie n'est pas chose aisée : comment signifier ma gratitude envers toutes celles et ceux qui m'ont aidé, rassuré, guidé et surtout supporté (!) pendant tout ce temps ?! Obligeamment, mes premières pensées vont à Céline, Pascal, Anaïs et Marie, la « #TeamClermont »¹ de l'équipe ProFan, ainsi que tou-te-s les collègues, camarades acteurs et actrices qui ont rendu ce projet possible. Non content de faire partie d'une équipe solidaire, d'avoir rencontré des personnes formidables, aussi professionnellement qu'humainement, j'aimerais également remercier les collègues du LAPSCO, doctorant-es, post-doctorant-es, enseignant-es et tou-te-s les autres pour leur accueil et leur soutien (mention particulière à Eva, Nonore et Camille pour leur aide précieuse sur la rédaction de ce manuscrit !).

Dédicace spéciale à l'équipe 2 (« *the best* »), Jean-Claude et Alice pour les discussions passionnantes et l'initiation à un certain jeu de revalorisation de notre chère langue française. Ensuite, mes pensées vont à ma famille et mes ami-es, pour leur support infailible, l'amour que je leur porte, et tous ces moments qui ont embelli cette histoire. Merci beaucoup, et aussi désolé de vous avoir continuellement parlé du boulot et de vous avoir paru acablé à certains moments (*oups*) ! Il est évident que ces quelques lignes ne suffiront pas à exprimer toute la gratitude que j'ai pour vous et à quel point je suis fier d'avoir vécu cette expérience à vos côtés. Sans surprise, je vous dédie aussi une part statistiquement significative de ce travail de thèse qui, pour moi, ne se résume pas seulement à ce qui transparait dans ce manuscrit, mais également à tous les changements vers lesquels vous m'avez guidé. Sur une petite dernière note musicale, j'aimerais encore vous remercier :

*“ It's no place for the faint-hearted, but my heart is strong,
because now I know where I belong,
it's you and I against the world ”*²

Officialisons donc la fin du début de ce manuscrit qui, je l'espère, vous apportera un certain plaisir intellectuel. Nonobstant ces tergiversations, je vous souhaite une très bonne lecture.

¹« *Guillemeté* » dans le texte, se référer à M. CROIZET J-C ou Me. KHAMZINA K. pour les règles du jeu.

² Muse - Aftermath de l'album Drones sorti en 2015, également « *guillemeté* » dans le texte. Le « *live* » était franc « *cool* ».

Résumé

Ce travail de recherche a pour objectif de tester les effets de la classe puzzle, ou « *Jigsaw classroom* », sur les apprentissages. La méthode Jigsaw est une pédagogie coopérative créée par Aronson et collaborateurs dans les années 1970, afin de favoriser l'inclusion des minorités ethniques (e.g, Mexicain·es et Afro-Américain·es) dans les écoles nouvellement déségrégues. Selon la théorie de l'interdépendance sociale les effets positifs de l'apprentissage coopératif dépendent de la structuration des interactions entre individus (Deutsch, 1949 ; Johnson & Johnson, 1989). Dans Jigsaw, la structuration de cette interdépendance provient essentiellement de la distribution de ressources complémentaires : chaque individu dispose d'une « pièce du puzzle » à reconstituer à l'aide des autres membres du groupe. La coordination des efforts entre membres devrait amener ces dernier·es à mettre en place des interactions facilitatrices (e.g., comportements d'entraide, explications et questionnements) et aboutir à un meilleur apprentissage. Toutefois, bien que cette méthode soit présentée par ses concepteur·rices comme un outil efficace pour améliorer l'apprentissage des élèves, les preuves empiriques tendent à manquer. Dans cette thèse, l'efficacité de Jigsaw sera questionnée à travers une analyse de la littérature scientifique, ainsi qu'une méta-analyse sur les travaux récents et un ensemble d'études expérimentales menées auprès d'élèves de sixième. À notre connaissance, bien que certaines recherches testant les effets de Jigsaw soient compilées dans des méta-analyses (Kyndt et al., 2013), il n'existe pas à ce jour de méta-analyses testant spécifiquement les effets de Jigsaw sur les apprentissages.

À travers six chapitres, nous tenterons d'apporter des éléments permettant d'évaluer l'efficacité de la méthode Jigsaw sur les apprentissages. Dans le chapitre 1, nous présentons la théorie de l'interdépendance sociale, plusieurs définitions et manières de structurer de la coopération entre élèves ainsi qu'une revue des leurs effets sur les apprentissages. Nous développons l'idée qu'il existe des différences d'efficacité entre les pédagogies coopératives créées entre les années 1960 et 2000 (Newmann & Thompson, 1987 ; Johnson et al., 2000), et que certaines d'entre elles n'ont pas encore fait l'objet d'une validation empirique solide. Le chapitre 2 examine l'une d'elle en détail : *Jigsaw* (Aronson et al., 1978 ; Aronson & Patnoe, 2011). Nous y décrivons l'évolution des études empiriques menées depuis sa création jusqu'à ce jour. Le chapitre 3 pointe certaines limites de cette littérature, notamment par rapport à la puissance statistique et les procédures méthodologiques, ainsi que les impacts qu'elles peuvent avoir sur l'estimation de l'efficacité de Jigsaw sur les apprentissages. Nous y développons aussi notre hypothèse de recherche, son opérationnalisation ainsi que les outils et procédures

statistiques que nous utilisons dans les chapitres empiriques : tests d'équivalence (Lakens, 2017), plus petite taille d'effet d'intérêt (Hattie, 2009) et méta-analyses (Borenstein et al., 2010 ; Goh et al., 2016). Le chapitre 4 présente les résultats d'une méta-analyse des effets de Jigsaw sur les apprentissages, à travers des articles empiriques publiés entre les années 2000 et 2020. Nous testons plusieurs modérateurs (e.g., niveau scolaire, discipline étudiée, type de Jigsaw, localisation des recherches) afin de quantifier la dispersion des effets de Jigsaw et de mieux comprendre l'hétérogénéité entre les études. Le chapitre 5 synthétise cinq études menées auprès de populations de collégien·nes français·es dans lesquelles nous testons l'efficacité de Jigsaw sur les apprentissages comparativement à des conditions de travail individuelles (études 1 et 2), ou d'enseignement habituel avec des enseignant·es volontaires (études 3A, 3B et 3C). Les résultats de ce chapitre sont interprétés à la lumière de la méta-analyse et des controverses liées à la structure de Jigsaw. Dans le sixième et dernier chapitre de ce manuscrit, nous synthétisons les résultats développés dans les chapitres théoriques et empiriques. Les contributions et limites de ce travail de recherche y sont développées, ainsi que des perspectives théoriques et pratiques pour les dépasser en vue de futures recherches sur cette thématique.

Mots-clés	Jigsaw classroom ; apprentissage coopératif ; interdépendance positive ; méta-analyse.
------------------	--

Abstract

The objective of this thesis is to test the effect of the Jigsaw classroom on learning. The Jigsaw classroom is a cooperative technique created by Aronson and his colleagues in the 1970s to promote the inclusion of ethnic minorities (e.g., Mexican and African-American) in desegregated schools. Although this method is presented by its developers as an effective tool for improving student learning, empirical evidence is lacking. According to the social interdependence theory, the structure of interactions between individuals determine the effects of cooperative learning (Deutsch, 1949; Johnson & Johnson, 1989). In Jigsaw, this structure comes from the distribution of complementary resources: each individual owns a “*jigsaw piece*”, namely a piece of information which requires the coordination of efforts among members to answer a problematic. With the help of other group members, promotive interactions (e.g., helping behaviors, explanations and questioning) should emerge which results in a better learning for the members.

In this thesis, Jigsaw's effectiveness will be evaluated through a review of the scientific literature, as well as a meta-analysis of recent research and a set of experimental studies conducted among french sixth graders. To our knowledge, the experimental study of Jigsaw's effects on learning in student populations is almost non-existent in the scientific literature and even though some research testing these effects is compiled in meta-analyses (Kyndt et al., 2013), there are no meta-analyses to date that specifically adress the question of Jigsaw's effects on learning. Hence, the research presented in this manuscript will attempt to evaluate the effectiveness of the Jigsaw method on learning. In Chapter 1, we present “social interdependence theory” (Johnson & Johnson, 1989, 2002, 2005), several definitions and ways of structuring cooperation between students, as well as a review of their effects on learning. Chapter 2 examines one of these cooperative technique in detail: *Jigsaw* (Aronson et al., 1978; Aronson & Patnoe, 2011). We describe the evolution of empirical studies conducted from its conception to the present day. Chapter 3 points out some of the limitations of this literature, particularly in terms of statistical power, and the impacts it may have on the estimation of Jigsaw's effectiveness on learning. We also develop our main hypothesis, its operationalization and the statistical tools and procedures we use in the empirical chapters: equivalence tests (Lakens, 2017), smallest effect size of interest (Hattie, 2009) and meta-analyses (Borenstein et al., 2010; Goh et al., 2016). Chapter 4 presents the results of a meta-analysis of Jigsaw's effects on learning, which synthesized empirical articles published between 2000 and 2020. We test several moderators (e.g., grade level, discipline, type of Jigsaw, location of research) in order

to quantify the dispersion of Jigsaw effects and to assess heterogeneity between studies. Chapter 5 compiles five studies conducted among french sixth graders in which we test the effectiveness of Jigsaw on learning, compared to an “individual” (studies 1 and 2) or a “teaching as usual” condition (studies 3A, 3B and 3C). The results of this chapter are interpreted with regard to the meta-analysis and the debates related to the structure of Jigsaw. In the last chapter of this manuscript, we summarize the main results developed trough the theoretical and empirical chapters. The contributions and limitations of our research are developed, as well as theoretical and practical perspectives to overcome them in view of future research.

Keywords Jigsaw classroom; cooperative learning; positive interdependence; meta-analysis.

Table des matières

Table des matières	8
Table des tableaux	13
Table des figures	14
Introduction générale	15
Chapitre 1. La théorie de l'interdépendance positive et l'apprentissage coopératif : vue d'ensemble	17
1. Objectif du chapitre	17
2. Historique et enjeux actuels	17
3. La théorie de l'interdépendance positive	20
3.1. L'interdépendance sociale (Deutsch, 1949)	20
3.2. La théorie de l'interdépendance positive (Johnson & Johnson, 1989 ; 2002)	21
3.3. Les éléments essentiels à la coopération	24
3.3.1. L'interdépendance positive	24
3.3.2. Responsabilité individuelle et contribution identifiable	26
3.3.3. L'utilisation de compétences sociales	27
3.3.4. Réflexion au niveau du groupe	28
3.3.5. Interactions facilitatrices	28
4. Un aperçu des principales pédagogies coopératives	29
5. Les effets de l'apprentissage coopératif sur les performances scolaires	32
6. De la théorie à la pratique	36
Chapitre 2. La classe puzzle ou « Jigsaw » : Approche historique, théorique et empirique	38
1. Objectif du chapitre	38
2. Contexte historique	38
3. Déroulement de Jigsaw	40
4. Les études menées sur Jigsaw, par période	43
4.1. Revues systématiques de la littérature	43
4.2. La genèse : 1970-1980	45
4.3. La deuxième version de Jigsaw : 1980-1990	47
4.4. Une validation empirique de la méthode ? 1990-2000	48
4.5. L'exportation de la pédagogie : 2000-2010	51
4.6. La littérature récente : 2010-2020	55
5. Une littérature inconsistante	60
Chapitre 3. Les limites de la littérature Jigsaw : Considérations méthodologiques et opérationnalisation de l'hypothèse de recherche	62
1. Objectif du chapitre	62
2. Une puissance statistique souvent faible	62

2.1.	La question de la puissance statistique -----	62
2.2.	La taille d'effet probable de Jigsaw -----	64
2.3.	Un probable biais de publication en faveur des résultats positifs -----	67
3.	Des analyses statistiques inadéquates pour évaluer l'efficacité de Jigsaw -----	69
3.1.	Procédures statistiques et design expérimental/indépendance des données -----	69
3.2.	L'indépendance des données -----	70
3.3.	Un manque de transparence : le cas de l'université d'Attaturk -----	72
4.	Tenir compte des limites méthodologiques -----	73
4.1.	Définition des problématiques et de l'hypothèse de recherche -----	74
4.2.	Spécification de la « plus petite taille d'effet d'intérêt » (SESOI) -----	74
4.3.	Présentation de la procédure de méta-analyse interne -----	75
5.	Synthèse et transition -----	76
Chapitre 4. L'utilisation de Jigsaw en classe permet-elle de favoriser les apprentissages ? Une méta-analyse des travaux récents -----		78
1.	Objectif du chapitre -----	78
2.	Méthode -----	78
2.1.	Critères d'inclusion et d'exclusion -----	78
2.2.	Mesure de l'apprentissage -----	80
2.3.	Groupe contrôle -----	81
2.4.	Effet global et dispersion -----	84
2.4.1.	Calcul de l'effet moyen et spécification du modèle -----	84
2.4.2.	Ajustements -----	85
2.4.3.	Pondération -----	85
2.5.	Hétérogénéité et tests de modération -----	86
2.5.1.	Type de Jigsaw -----	86
2.5.2.	Localisation de l'article -----	87
2.5.3.	Durée de l'intervention -----	87
2.5.4.	Discipline -----	87
2.5.5.	Niveau scolaire -----	88
2.5.6.	Biais de publication -----	88
3.	Résultats -----	89
3.1.	Effet global de Jigsaw -----	89
3.2.	Dispersion de l'effet et tests de modérateurs -----	91
3.2.1.	Type de Jigsaw -----	92
3.2.2.	Localisation de l'article -----	92
3.2.3.	Durée de l'intervention -----	92
3.2.4.	Discipline -----	93
3.2.5.	Niveau scolaire -----	93

3.3.	Corrélations entre modérateurs -----	93
3.4.	Tests de puissance statistique appliqués à la méta-analyse -----	95
4.	Discussion -----	97
4.1.	Synthèse des principaux résultats -----	97
4.2.	Un modérateur culturel ? -----	99
4.3.	Conclusions intermédiaires-----	102
Chapitre 5. Jigsaw améliore-t-il les apprentissages en science d'élèves de 6 ^e ? Cinq expériences et une méta-analyse interne -----		104
1.	Objectif du chapitre -----	104
2.	Vue d'ensemble-----	104
2.1.	Choix de la population -----	104
2.2.	Calcul de puissance a priori -----	106
3.	Expérience 1 -----	108
3.1.	Participant·es -----	108
3.2.	Procédure -----	108
3.3.	Matériel -----	109
3.3.1.	Contenu de la leçon -----	109
3.3.2.	Évaluation de connaissance-----	109
3.3.3.	Niveau scolaire initial -----	110
3.4.	Résultats-----	110
3.4.1.	Analyses préliminaires -----	110
3.4.2.	Effet de Jigsaw sur la performance académique (E1) -----	110
3.5.	Discussion intermédiaire (E1) -----	111
4.	Expérience 2 -----	112
4.1.	Participant·es -----	112
4.2.	Procédure -----	112
4.3.	Matériel -----	112
4.3.1.	Contenu de la leçon -----	112
4.3.2.	Évaluation de connaissances -----	113
4.3.3.	Niveau scolaire initial -----	113
4.4.	Résultats-----	113
4.4.1.	Analyses préliminaires -----	113
4.4.2.	Effet de Jigsaw sur la performance académique (E2) -----	113
4.5.	Discussion intermédiaire (E2) -----	114
5.	Expérience 3A -----	115
5.1.	Participant·es -----	115
5.2.	Procédure -----	115
5.3.	Matériel -----	116

5.3.1.	Contenu des leçons-----	116
5.3.2.	Évaluations de connaissances -----	117
5.3.3.	Niveau scolaire initial -----	118
5.4.	Résultats-----	118
5.4.1.	Analyses préliminaires -----	118
5.4.2.	Effet de Jigsaw sur les performances académiques (E3A)-----	118
6.	Expérience 3B -----	119
6.1.	Participant·es -----	119
6.2.	Procédure -----	119
6.3.	Matériel -----	119
6.3.1.	Contenu des leçons-----	119
6.3.2.	Évaluation de connaissances -----	120
6.3.3.	Niveau scolaire initial -----	120
6.4.	Résultats-----	120
6.4.1.	Analyses préliminaires -----	120
6.4.2.	Effet de Jigsaw sur les performances académiques (E3B)-----	120
7.	Expérience 3C -----	121
7.1.	Participants -----	121
7.2.	Procédure -----	121
7.3.	Matériel -----	121
7.3.1.	Contenu des leçons-----	121
7.3.2.	Évaluation de connaissance-----	122
7.4.	Résultats-----	122
7.4.1.	Analyses préliminaires.-----	122
7.4.2.	Effet de Jigsaw sur les performances académiques (E3C)-----	122
7.5.	Discussion intermédiaire (E3A, E3B et E3C) -----	123
8.	Méta-analyse interne-----	123
9.	Discussion générale-----	125
9.1.	Synthèse des principaux résultats -----	125
9.2.	Contributions et limites -----	125
9.3.	Viser le « <i>gold standard</i> » des expérimentations randomisées à grande échelle	130
Chapitre 6. Repenser le puzzle : Synthèse, limites et perspectives-----		134
1.	Objectif du chapitre -----	134
2.	Résumé des chapitres précédents -----	134
3.	Quantifier et qualifier les effets de Jigsaw sur les apprentissages-----	135
3.1.	La nécessité de tester empiriquement une pratique pédagogique-----	135
3.2.	Des modérateurs de l'efficacité de Jigsaw ? -----	137
3.3.	La structure de Jigsaw : frein ou levier aux interactions facilitatrices ? -----	141

4. Perspectives et ouvertures-----	143
4.1. Jigsaw et l'élaboration cognitive -----	143
4.2. Retourner aux fondamentaux? -----	146
4.3. Mener des études à grande échelle pour détecter les effets de Jigsaw -----	149
5. Améliorer les apprentissages avec la coopération ?-----	153
5.1. L'obsession des performances académiques-----	153
5.2. Implémenter efficacement un changement-----	155
6. Conclusion -----	156
Bibliographie -----	158
Table des annexes -----	176
Annexes -----	i

Table des tableaux

Tableau 1: Méta-analyses et effets de l'apprentissage coopératif sur la réussite académique (Hattie, 2009)	32
Tableau 2: Compilation de méta-analyses portant sur les effets de l'apprentissage coopératif en comparaison à un groupe contrôle sur les apprentissages	34
Tableau 3: Structuration de l'interdépendance positive dans Jigsaw	40
Tableau 4: Les différentes versions de Jigsaw (Robert et al. en préparation)	52
Tableau 5: Caractéristiques et résultats principaux des expérimentations menées sur la période 2010-2020 (sources : Robert et al., en préparation ; Vives et al., en préparation)	58
Tableau 6: Scenarii possibles d'un test d'hypothèse nulle (Lakens: Improving your Statistical Questions; coursera.org).....	63
Tableau 7: Simulations des tailles d'effet possibles pour un t-test apparié (Sahin, 2010)	71
Tableau 8: Synthèse des perspectives engagées pour répondre aux limites de la littérature ..	77
Tableau 9: Liste des études incluses dans la méta-analyse	83
Tableau 10: Formules de calcul des estimations de taille d'effet moyenne (ES)	84
Tableau 11: Résultats des analyses de modération	91
Tableau 12: Corrélations et intervalles de confiance des modérateurs de la méta-analyse	94
Tableau 13: Caractéristiques des études présentées dans le chapitre 6.....	105
Tableau 14: Distribution des classes selon le design expérimental (E3A)	115
Tableau 15: Synthèse des résultats (expériences 3A, 3B et 3C).....	123
Tableau 16: Estimation du nombre total de participant-es en fonction de la puissance statistique du test et de la taille d'effet à détecter (G*Power, Faul et al., 2007)	149

Table des figures

Figure 1: Schéma de la théorie de l'interdépendance sociale (Johnson & Johnson, 2005)	22
Figure 2: Tailles d'effets (d de Cohen) de l'interdépendance positive sur plusieurs variables (Johnson & Johnson, 1989).....	23
Figure 3: Tailles d'effet et nombre d'études associées aux pédagogies coopératives comparativement à une approche individuelle (Johnson, Johnson & Stanne, 2000).....	31
Figure 4: Étape 1 de Jigsaw (travail individuel)	41
Figure 5: Étape 2 de Jigsaw (groupe d'expert-es)	42
Figure 6: Étape 3 de Jigsaw (phase puzzle)	43
Figure 7: Évolution du nombre total d'articles (k) et des échantillons (N) dans les recherches Jigsaw (données provenant de Robert et al., en préparation).....	44
Figure 8: Taille d'échantillon et nombre d'articles empiriques publiés après 2000 (Robert al., en préparation).....	65
Figure 9: Courbe de puissance de test simulée pour une taille d'effet de $\delta = 0.40$, au seuil $\alpha = .05$ (JAMOVI 1.1.8.0)	66
Figure 10: Analyse de sensibilité (puissance de test a posteriori) sur la base du nombre moyen de participants par condition dans la littérature Jigsaw (JAMOVI 1.1.8.0).....	67
Figure 11: Tri des articles en fonction des critères d'éligibilité à la méta-analyse.....	79
Figure 12: Forest-plot des ES observées	89
Figure 13: Funnel-plot des ES observées.....	90
Figure 14: Analyse de puissance a posteriori pour méta-analyse (Borenstein et al., 2010)....	96
Figure 15: Forest-plot de tailles d'effet observées dans les expériences E1 à E3C	124

Introduction générale

« Outre la transmission des connaissances, la Nation fixe comme mission première à l'école de faire partager aux élèves les valeurs de la République. Le service public de l'éducation fait acquérir à tous les élèves le respect de l'égalité des êtres humains, de la liberté de conscience et de la laïcité. Par son organisation et ses méthodes, comme par la formation des maîtres qui y enseignent, il favorise la coopération entre les élèves. ». Article L111-1 du Code de l'Éducation.

La coopération, où l'apprentissage coopératif, est un monument de la psychologie sociale et de l'éducation. Johnson et Johnson (2009) recensent plus d'un millier d'articles empiriques rédigés depuis les années 1960, compilés dans plusieurs méta (Johnson et al., 2000 ; Kyndt et al., 2013) et méga-analyses (Hattie, 2009) qui montrent que l'apprentissage coopératif permettrait d'améliorer l'apprentissage des élèves ainsi que leur bien-être à l'école, leur motivation et d'autres variables socioaffectives. On peut définir l'apprentissage coopératif comme un ensemble de situations dans lesquelles des individus interagissent et apprennent ensemble : selon plusieurs recherches en didactique, sociologie ainsi qu'en psychologie, cette pratique est supposément ancrée chez les enseignant·es qui encouragent leurs élèves à s'entraider et à échanger les uns avec les autres (Slavin, 2011).

Pourtant, la mise en place de l'apprentissage coopératif au sein des classes rencontre un certain nombre de barrières. Malgré un consensus établi sur l'efficacité de l'apprentissage coopératif sur la base de plusieurs décennies de recherches empiriques et théoriques (Hattie, 2009, 2015 ; Johnson & Johnson, 2009), Pianta et collaborateurs (2007) observent que moins de 10% du temps d'enseignement dans les classes américaines est consacré au travail de groupe. Plus récemment et en France, le rapport EPODE (Enquête PériODique sur l'Enseignement, DEPP, 2018) s'est intéressé aux pratiques enseignantes de plusieurs pays européens, dont la France. Ce rapport indique que, malgré l'utilité perçue et des attitudes positives envers l'apprentissage coopératif, moins d'un·e enseignant·e sur deux mettait en place une telle pratique dans sa classe. Pour Baloché et Brody (2017), une des limites que rencontrent les enseignant·es souhaitant mettre en place l'apprentissage coopératif dans leurs classes réside dans leur manque de formation : en effet, la littérature scientifique portant sur les effets bénéfiques de l'apprentissage coopératif suggère que le simple fait de faire coopérer des individus, ou des les assigner à un travail de groupe, ne suffit pas pour faire émerger un apprentissage de meilleure qualité (Cohen, 1994 ; Gillies, 2003 ; Johnson & Johnson, 1989 ; 2005). Plusieurs chercheur·es développent l'idée selon laquelle la coopération peut être

structurée de plusieurs manières, en mettant en place différentes techniques pédagogiques dans les classes (Sharan, 1980 ; Slavin, 1980 ; Topping et al., 2017). Pour Johnson et Johnson (2000), l'évaluation de l'efficacité de ces techniques est primordiale : seules les interventions ayant fait l'objet d'une validation empirique rigoureuse devraient être déployées dans les établissements afin d'en faire bénéficier les élèves. Or, parmi les différentes techniques pédagogiques développées jusqu'à ce jour, la classe puzzle, ou Jigsaw (Aronson et al., 1978) semble déroger à cette règle : la popularité de la méthode ne correspond pas à la quantité de preuves empiriques validant son efficacité. Au contraire, il semble y avoir une disparité entre son efficacité, telle que rapportée par ses concepteurs, et les recherches qui tendent à rapporter des effets beaucoup plus mitigés (Moskowitz et al., 1983 ; 1985 ; Roseth et al., 2018). La littérature scientifique portant sur la méthode Jigsaw est aussi marquée par d'autres inconsistances : les premières synthèses de travaux menés dans les années 1980 à 2000 tendent à évaluer son efficacité comme étant relativement faible, comparativement à d'autres pédagogies coopératives (Newmann & Thompson, 1987 ; Johnson et al., 2000). Pourtant, depuis les années 2000, une recrudescence d'articles rapportant des effets positifs de la méthode ont permis à Jigsaw à se hisser dans le haut du classement de John Hattie (2017), qui recueille l'efficacité de différentes interventions académiques sur les apprentissages.

Plusieurs zones d'ombre semblent exister dans cette littérature : l'objectif de cette thèse est d'apporter des éléments empiriques et théoriques pour tenter d'évaluer l'efficacité de la méthode Jigsaw sur les apprentissages. Dans le cadre de cette thèse, nous présenterons une revue critique de la littérature, une méta-analyse des travaux récents et cinq expérimentations menées auprès de collégien·nes français·es afin de contribuer à une réévaluation de l'efficacité de Jigsaw.

Chapitre 1. La théorie de l'interdépendance positive et l'apprentissage coopératif : vue d'ensemble

1. Objectif du chapitre

L'objectif de ce chapitre sera de présenter les principaux éléments théoriques, historiques et de définitions de l'apprentissage coopératif. Nous présenterons un aperçu historique que cette recherche puis nous nous intéresserons particulièrement à la théorie de l'interdépendance positive telle que développée par Deutsch (1949) puis par Johnson et Johnson (1989 ; 2005), ainsi que sa relation avec l'apprentissage coopératif. Pour finir, nous illustrerons, à l'aide de méta-analyses, les effets que peut avoir la coopération sur les apprentissages.

2. Historique et enjeux actuels

Dans leur ouvrage « *Educational Psychology, A Century of Contributions* » Zimmerman et Schunk (2002) situent l'apparition des premiers travaux en psychologie de l'éducation dans les années 1890. D'après les auteurs, ceux-ci seraient influencés par la philosophie humaniste et pragmatique de John Dewey (1916) et des premières recherches psychologiques menées par William James (1890). Dewey fut l'un des premiers à développer l'argument que la coopération, c'est-à-dire des interactions positives et facilitatrices entre individus, pouvait permettre d'améliorer leur développement et leur éducation. Pour Lev Vygotsky (1978), le développement et l'apprentissage d'un individu dépendent de l'interaction entre son environnement social et ses pairs. Que ce soit par expérience (i.e., théorie des stades développementaux, Piaget, 1926), par observation et imitation (i.e., théorie sociocognitive, Bandura, 1986) ou par instruction par un pair plus compétent (Vygotsky, 1978), le rôle des pairs est au coeur de sa théorie socioculturelle des apprentissages. Dans cette théorie, Vygotsky (1978) développe l'idée selon laquelle les interactions entre étudiant·es et leurs professeur·es, ainsi que les interactions avec les étudiant·es entre eux, permet l'émergence de l'apprentissage, c'est à dire de la construction de connaissances et d'idées. Plus particulièrement, Vygotsky conceptualise la « zone proximale de développement », qui définit la différence entre l'état des apprentissages d'un individu, et les acquisitions possibles à accomplir avec l'aide d'une personne plus compétente que l'individu. Pour Vygotsky, la relation entre étudiant·e et professeur·e est primordiale puisqu'elle permet d'atteindre un équilibre (i.e., « *scaffolding* ») qui consiste pour l'étudiant·e à progressivement atteindre le résultat de l'apprentissage souhaité par le professeur·e. Contrairement à Piaget (1926) qui affirmait que l'apprentissage des individus, notamment des enfants lors de leur développement,

s'effectuait de manière individuelle par confrontation entre des expériences de pensées et le « réel », Vygostky (1978) fait l'emphase sur la nature sociale des apprentissages : ce sont les interactions entre pairs permettent l'émergence de nouvelles idées et perspectives, et donc, l'apprentissage. La perspective socioconstructiviste constitue la base théorique de l'apprentissage coopératif : les interactions entre enfants favorisent leur développement social et cognitif, puisque les enfants d'âges similaires sont susceptibles d'interagir dans des zones proximales de développement relativement proches les unes des autres.

Il faudra attendre les années 1940 et les travaux de Morton Deutsch pour voir émerger une des théories les plus influentes en psychologie de l'éducation, la théorie de l'interdépendance sociale. Inspiré par la psychologie gestaltiste et par les travaux de Lewin (1936 ; 1942) et de Kofka (1935), Deutsch (1949) formalise sa théorie sur la continuité des travaux de Lewin et des dynamiques de groupe. Pour Lewin (1936) les relations entre les membres déterminent « l'essence » du groupe, de telle sorte à ce qu'un changement dans l'état d'un des membres va impacter les interactions avec les autres. Ces derniers sont placés dans un « état de tension » et c'est la présence de buts ou d'objectifs communs qui détermine la façon dont les individus interagissent entre eux. La théorie de l'interdépendance sociale de Deutsch (1949) reprend ces éléments et décrit les phénomènes, activités et mécanismes qui prennent place lorsque des individus interagissent entre eux au sein de groupes (voir Figure 1). Plus tard, ce sont les frères David et Roger Johnson qui élargissent la théorie de l'interdépendance sociale de Deutsch en la nommant « théorie de l'interdépendance positive » (Johnson & Johnson, 1989). Cette théorie figure aujourd'hui parmi les trois principales approches théoriques de l'apprentissage coopératif en psychologie, telles que reprises par Slavin (2016) :

- L'approche de la *cohésion sociale*, c'est-à-dire celle de la théorie de l'interdépendance sociale, propose que les individus bénéficient de l'apprentissage coopératif parce qu'ils partagent un but commun et s'entraident pour accomplir ce dernier. Johnson et Johnson (1989 ; 1995 ; 2005) en sont les principaux défenseurs.
- L'approche *motivationnelle* suppose que les effets de l'apprentissage coopératif sont déterminés par les intérêts des individus au sein de leur groupe. Les tenants de cette approche sont davantage intéressés par les manières de structurer l'interdépendance positive de récompenses.
- L'approche *cognitive* se focalise sur les interactions entre individus, avec l'idée que la nature des interactions qui déterminent les effets de l'apprentissage coopératif. Cette approche est donc plutôt centrée sur les *processus*.

Selon la perspective de cohésion sociale, les interactions entre les membres du groupe sont déterminées par la cohésion du groupe. En d'autres termes, lors d'un apprentissage coopératif, les individus s'engagent dans la tâche et s'entraident parce qu'ils s'identifient au groupe et souhaitent que les autres réussissent. Selon cette perspective, c'est principalement le but commun, c'est-à-dire la structure des objectifs de la coopération qui fait émerger les interactions dites « facilitatrices ». C'est quand les membres du groupe se sentent responsables vis-à-vis de leurs apprentissages et de ceux d'autrui que les effets positifs de la coopération apparaissent. La perspective motivationnelle suppose que ce ne sont pas les objectifs ou les buts, mais plutôt les récompenses et les intérêts qu'ont les individus à coopérer qui expliquent les bénéfices de l'apprentissage coopératif. Slavin (1980 ; 1983) affirme que la structure des récompenses individuelles et de groupe crée une situation dans laquelle les individus ont intérêt à coopérer : les interactions facilitatrices émergent alors pour accomplir la tâche et être récompensés en fonction de sa contribution. La principale différence entre ces deux approches se situe dans la raison qui justifie les interactions entre individus : l'approche motivationnelle soutient que les individus coopèrent parce que c'est dans leur propre intérêt, tandis que l'approche de la cohésion sociale, ou de l'interdépendance positive, soutient que c'est parce que les individus se soucient de leur groupe et de la réussite des autres. L'approche cognitive, quant à elle, s'intéresse non pas aux raisons (i.e., intérêt ou responsabilité) ou aux conditions (i.e, structuration des buts et/ou récompenses), mais aux interactions telles qu'elles émergent pendant l'apprentissage coopératif. Pour les tenants de cette approche, c'est la qualité des interactions entre individus qui explique les effets positifs ou négatifs de l'apprentissage coopératif. C'est l'élaboration cognitive, c'est-à-dire la restructuration des connaissances et l'engagement des processus cognitifs profonds, qui prédit un apprentissage de meilleure qualité. D'après cette approche, l'un des moyens les plus efficaces pour qu'il y ait une élaboration cognitive est d'expliquer le matériel à quelqu'un d'autre, c'est-à-dire de mettre en place des relations de tutorat par les pairs (Wittrock, 1992).

Entre les années 1970 et 2000, les chercheurs issus de ces différents courants sont parvenus au consensus selon lequel la coopération avait des effets positifs sur de nombreuses variables, notamment l'apprentissage des individus (Johnson & Johnson, 1989 ; Johnson et al., 2000 ; Newmann & Thompson, 1987 ; Springer et al., 1999). Durant cette période, plusieurs pédagogies coopératives ont été développées, se basant sur ces différents courants de pensée (voir Sharan, 1980 et Slavin, 1980 pour des revues détaillées). L'objectif de la plupart des études était alors de comparer l'efficacité des différentes pédagogies, c'est-à-dire de mettre en

évidence les conditions de structuration de l'interdépendance positive qui entraînent les gains les plus importants. Plus récemment, Johnson et Johnson (2009) s'accordent à dire que l'apprentissage coopératif dans son ensemble est une véritable « *success story* » : avec plus de 1000 études menées depuis les années 1970, combinées dans plusieurs méta et « méga » analyses (Lou et al., 1996 ; Springer et al., 1999 ; Hattie, 2009), et des effets positifs sur la réussite académique, l'estime de soi et la motivation, le consensus semble être établi.

Dans les prochains paragraphes, nous présenterons la théorie de l'interdépendance positive (i.e., l'approche par cohésion sociale), une des théories les plus influentes dans la psychologie de l'éducation qui a permis de décrire les différentes conditions nécessaires pour que l'apprentissage coopératif ait des effets bénéfiques sur les performances scolaires et d'autres variables socioaffectives.

3. La théorie de l'interdépendance positive

3.1. L'interdépendance sociale (Deutsch, 1949)

D'après Deutsch (1949), une interdépendance sociale existe quand les actions et les conséquences d'un individu sont affectées par celles d'autrui. Cette interdépendance peut être de deux natures différentes : positive, dans le cas où les interactions entre individus sont orientées vers un but commun (e.g., un attaquant de football qui aide son équipe à remporter le match) ou négative, lorsque la réussite d'un individu diminue les chances de réussite des autres ou suppose l'échec des autres (e.g., concours d'entrée dans une école). Autrement dit, la coopération résulte d'une relation d'interdépendance positive entre individus, tandis que la compétition résulte d'une relation d'interdépendance négative. Cette théorie a largement été inspirée des mouvances gestaltistes qui présentent les groupes comme des entités à part entière, dont les dynamiques entre les membres peuvent varier en fonction de leur structuration (Kofka, 1935 dans Johnson & Johnson, 2005). Ainsi, d'après la conceptualisation de cette théorie (voir Figure 1), les actions des individus à l'intérieur de groupes sont déterminées par leur type « d'interdépendance » : si les individus disposent d'un but commun, c'est-à-dire de raisons valables pour coopérer (e.g., si la réussite des uns est reliée positivement à la réussite des autres) alors leur interdépendance est positive. À l'inverse, dans le cas d'une corrélation négative entre les buts des individus, l'interdépendance est négative. D'après cette théorie, le type d'interdépendance va influencer la nature des interactions entre individus, à travers trois processus psychologiques : la « cathexis » ou « l'investissement d'énergie dans un objet autre que soi » (Deutsch, 1949 dans Johnson & Johnson, 2005), c'est-à-dire un concept proche de

l'engagement, la substituabilité ou la mesure dans laquelle les actions d'un individu peuvent se substituer à celles d'autrui (e.g., une activité terminée par un membre du groupe n'a pas besoin d'être répétée par un autre individu) et l'inductibilité, c'est-à-dire le degré d'influence exercée et reçue par les individus.

Pour résumer le modèle de Deutsch (1949), des individus reliés par un but positif, c'est-à-dire l'expectation de la réalisation d'un objectif commun, devraient mettre en place des interactions facilitatrices (e.g., entraide, réflexion, explication, comportements de coopération) afin d'augmenter la probabilité de réalisation de ce but. À l'inverse, des buts négatifs, ou une interdépendance négative, aura pour conséquence de générer des interactions dites « oppositionnelles », c'est-à-dire compétitives, où les individus tenteront de réduire les probabilités de réussite d'autrui. Pour résumer, cette théorie suppose que la présence d'intérêts communs, structurés par les buts et la tâche, permet l'émergence d'interactions facilitatrices entre les membres. Appliquées à l'éducation, ces interactions facilitatrices vont entraîner un meilleur apprentissage.

3.2. La théorie de l'interdépendance positive (Johnson & Johnson, 1989 ; 2002)

La théorie de l'interdépendance positive étend les premières formulations de Morton Deutsch (1949) tout en précisant les conditions nécessaires à la structuration de l'interdépendance positive entre les individus au sein d'un groupe. Elle repose notamment sur une compilation d'études mettant en évidence la supériorité de l'interdépendance positive, comparativement à l'interdépendance négative ou bien l'indépendance (i.e., une structure individualiste), sur une gamme de variables comprenant la réussite scolaire, le bien-être psychologique ou encore l'estime de soi (Johnson & Johnson, 1989). Johnson et Johnson développent l'idée selon laquelle la coopération, qui renvoie à une interdépendance positive entre individus, doit être structurée de manière à remplir différentes conditions afin d'être réellement bénéfique pour les individus. Ils soutiennent que la qualité des interactions des individus au sein de leur groupe est déterminée par des conditions en qui permettent l'émergence, notamment la cohésion du groupe et un sentiment de responsabilité vis-à-vis des autres membres et de la tâche à accomplir. Dit autrement, si le cadre est favorable à des interactions positives et facilitatrices, les individus devraient développer une cohésion de groupe importante, ce qui devrait faciliter la réussite de la tâche assignée au groupe (voir Figure 1 ci-dessous).

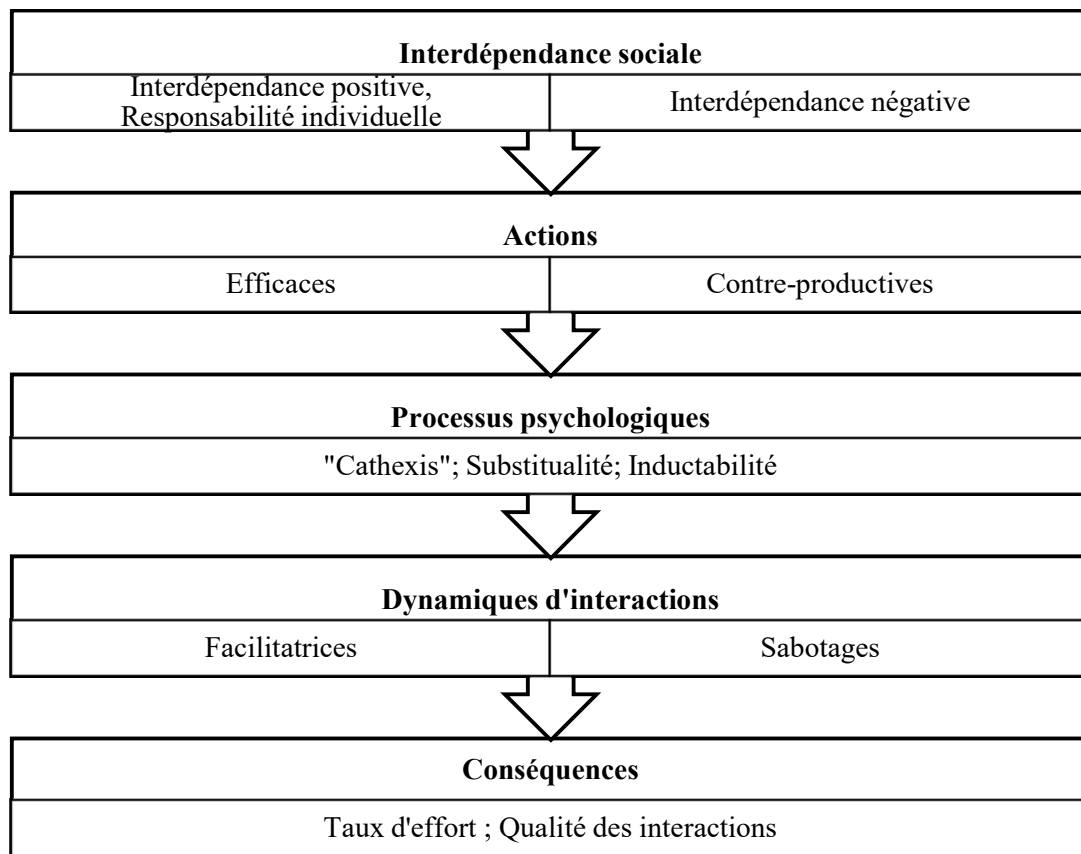


Figure 1: Schéma de la théorie de l'interdépendance sociale (Johnson & Johnson, 2005)

La théorie de l'interdépendance sociale met l'accent sur des explications en termes d'identification au groupe et de motivation à la réussite commune : les membres du groupe s'entraident et s'engagent dans la tâche commune parce qu'ils s'y identifient et veulent que les autres membres réussissent. Selon cette théorie, c'est donc l'articulation conjointe des efforts, sous la volonté d'accomplir un but commun, qui explique les bénéfices de la coopération sur de multiples variables (voir Figure 2).

L'apprentissage coopératif selon Johnson et Johnson repose principalement sur l'idée que la présence d'un but commun et d'un sentiment de responsabilité sont les principales conditions à remplir pour faire émerger les interactions facilitatrices. Dans la perspective motivationnelle, Slavin (1983) argumente que la présence de récompense de groupe est au moins aussi importante que le but commun, puisque les individus sont principalement guidés par leurs propres intérêts à coopérer. Ces différentes perspectives ont amené plusieurs équipes de chercheur·es à concevoir différentes pédagogies coopératives, c'est-à-dire des manières de structurer ces relations d'interdépendance positive, qui reposent globalement sur une « base commune ». Cette base commune de la coopération se retrouve notamment dans les définitions que l'on peut trouver, dans la littérature scientifique, sur le concept même d'apprentissage

coopératif. Johnson et collaborateurs (2000) décrivent par exemple l'apprentissage coopératif comme « [...] un terme générique qui renvoie à différentes manières d'organiser et d'enseigner une classe de manière coopérative » (p.3). Pour Newmann et Thompson (1987), l'apprentissage coopératif renvoie à « des approches pédagogiques qui varient, mais qui ont toutes en commun de proposer une alternative à « l'enseignement frontal », où l'enseignant-e fait cours à toute la classe en même temps. À la place, les approches coopératives mettent les étudiants en petits groupes, sur la supposition qu'ils vont s'entraider et se motiver, résultant en des gains sur les apprentissages » (p.5). Buchs et Butera (2015) définissent l'apprentissage coopératif comme des « situations dans lesquelles un-e enseignant-e structure un travail en groupe avec l'objectif de maximiser les gains sociaux et cognitifs » (p.1). Enfin, Slavin (1983) offre une définition relativement large qui permet de regrouper différentes pratiques sous la même dénomination de l'apprentissage coopératif : « Ce qui caractérise ces méthodes est que les étudiants passent du temps à travailler en petits groupes - hétérogènes - dans lesquels ils doivent interagir et aider autrui. Il existe néanmoins une multitude de façon de structurer ces interactions, d'où un nombre important de « pédagogies coopératives » (p.431). Quoiqu'il en soit, l'intégralité des chercheur-es s'intéressant à l'apprentissage coopératif s'accordent pour dire qu'un certain nombre d'éléments sont essentiels à structurer pour s'assurer de son efficacité (Cohen, 1994 ; Gillies & Ashman, 2003 ; Johnson & Johnson, 2009 ; Slavin, 2016).

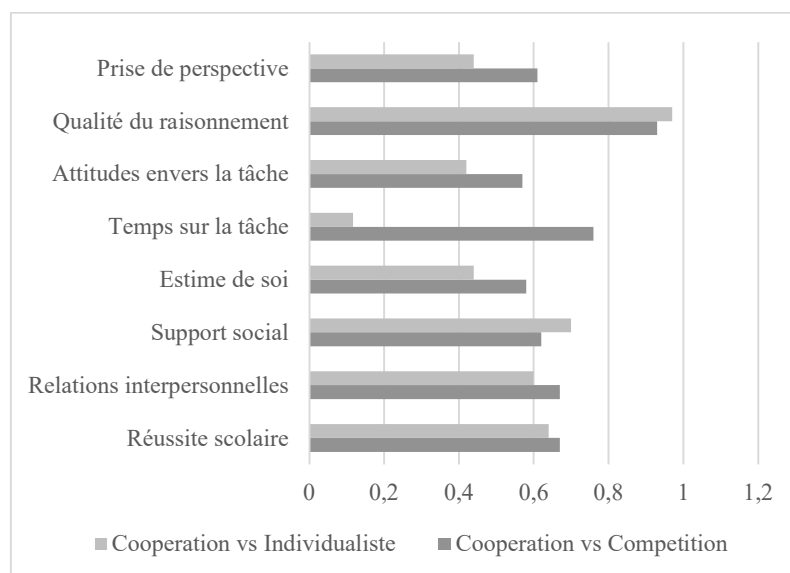


Figure 2: Tailles d'effets (*d* de Cohen) de l'interdépendance positive sur plusieurs variables (Johnson & Johnson, 1989)

3.3. Les éléments essentiels à la coopération

D'après Johnson et Johnson (2005 ; 2009), il existe au moins cinq conditions à remplir afin que l'apprentissage coopératif ait des effets positifs sur les apprentissages. Ces conditions font référence à la nature de l'interdépendance positive entre individus, à la responsabilité individuelle et la visibilité de la contribution de chacun·e, aux compétences sociales mises en jeu lors des interactions, à la réflexion au niveau du groupe et à la nature des interactions facilitatrices entre les membres.

3.3.1. L'interdépendance positive

Les relations d'interdépendance positive entre individus, au sein d'un groupe, peuvent être mises en place de plusieurs manières différentes. D'après Johnson et Johnson (1995 : 2002 ; 2005), le concept d'interdépendance positive regroupe différents éléments qui caractérisent à la fois les objectifs du groupe et la façon dont ces derniers sont évalués (i.e., « *goal* » et « *reward interdependence* »), la manière dont les ressources et les rôles sont répartis (i.e., « *mean* » et « *resource interdependence* ») entre les individus ainsi que l'environnement physique et social dans lequel ont lieu les interactions (i.e., « *boundary interdependence* »).

Les interdépendances de ressources et de rôles renvoient à la façon dont est structurée la tâche à effectuer par les individus dans leur groupe, c'est-à-dire les « moyens » dont ils disposent afin de la résoudre. Par exemple, les individus ne disposent pas toujours des mêmes ressources, comme c'est le cas dans le cadre de *Jigsaw* (Aronson et al., 1978) où les membres du groupe doivent mettre en commun différents éléments d'information pour répondre à une problématique générale (voir chapitre 2). Dans cette approche, chaque membre du groupe dispose d'un élément différent, et la coordination des efforts et des connaissances de chacun·e est nécessaire pour arriver à la résolution de la tâche, souvent une reconstitution d'une leçon scindée en sous-parties (Aronson & Patnoe, 2011). D'autres approches ne reposent pas sur une telle complémentarité de ressources, mais distribuent différents « rôles » au sein du groupe : une personne pourrait être responsable de la transcription du rapport collectif, une autre de la manipulation expérimentale, et une dernière de la coordination entre les membres (Sharan, 1980 ; Slavin, 1983 ; Topping et al., 2017). Il convient de noter qu'il est tout à fait possible de cumuler à la fois une interdépendance de ressources et de rôles. Pour reprendre le cas de *Jigsaw*, les individus endossent successivement le rôle « d'expert·es », qui maîtrisent la sous-partie qu'ils ont étudiée, et le rôle de « novice », à qui on explique et présente des éléments de connaissance dont iel ne dispose pas.

Un autre facteur susceptible de générer de l'interdépendance positive se situe dans les buts et les récompenses à répartir suite à la réalisation de la tâche. Johnson et Johnson (2005) définissent le but des individus en situation d'interdépendance comme « *a desired state of affair* », qu'on pourrait traduire par « un état attendu de la complétion d'une tâche ». Dans les contextes scolaires, buts et récompenses sont intrinsèquement reliés : le but principalement rencontré par les étudiant·es consiste en l'apprentissage d'une ou plusieurs leçons et leur « récompense » correspond à la nature de l'évaluation (e.g., individuelle, collective ou hybride) qui mesure cet apprentissage. Slavin (1983 ; 2011) affirme que les bénéfices associés à l'apprentissage coopératif dépendent fortement de la façon dont sont récompensés les individus qui ont contribué à la réussite du groupe. Pour ce dernier, il est important que les individus soient évalués en fonction de leur performance individuelle, tout en prenant également en compte celle du groupe. Pour ce faire, il utilise une méthode qui consiste à assigner au groupe une note moyenne qui correspond à la pondération des scores individuels des membres le constituant. Une autre manière de récompenser le groupe est d'évaluer un produit de groupe, comme un exposé oral ou un rapport collectif. De cette manière, tous les individus disposent de la même évaluation. D'après Johnson et Johnson (1990 ; 2005), la combinaison d'une interdépendance de moyens et de buts est additive : les individus seront d'autant plus efficaces s'ils disposent de moyens différents, sont récompensés en fonction de leurs performances individuelles et évalués en tant que groupe. Ces auteurs suggèrent également que, sans la présence de buts communs ou récompenses de groupe, la simple distribution de ressources complémentaires peut avoir des effets délétères sur les apprentissages. Ainsi, l'apprentissage coopératif améliore les performances académiques quand les élèves qui travaillent ensemble sont récompensé·es en tant que groupe, tandis que la contribution de chacun·e est visible. Slavin (1983) affirme également que la tâche doit être structurée de manière à ce qu'elle ne soit réalisable que par l'interaction entre les différents individus qui composent le groupe. Pour lui, même s'il existe un but ou une récompense commune à toutes et tous, la tâche perd de son intérêt si elle peut être réalisée intégralement par une seule personne.

Johnson et Johnson (2005) décrivent un dernier facteur à prendre en considération lors de la structuration de l'apprentissage coopératif : les « frontières ». L'interdépendance de frontière détermine qui interagit avec qui, et dans quelles circonstances (Kofka, 1935 cité dans Johnson & Johnson, 2009). Les caractéristiques de ces frontières peuvent dépendre des facteurs environnements (e.g., classe entière, salles différentes, îlots de tables), physiques (e.g., proximité immédiate avec d'autres membres) et sociaux (e.g., compétition ou indépendance

entre les groupes). Pour résumer, l'interdépendance de frontière détermine qui est un membre du groupe et qui ne l'est pas, ainsi que les mesures dans lesquelles les membres interagissent entre eux. Pour reprendre l'exemple de Jigsaw, les élèves coopèrent au sein de groupes de 3 à 5, leurs « groupes puzzle », mais également avec d'autres élèves dans une phase dite « d'expert-es », où se regroupent les individus qui possèdent la même pièce d'information. Chaque individu interagit au sein de son groupe « puzzle », mais également de manière plus temporaire avec son groupe « d'expert-es ». Dans Jigsaw, les groupes « puzzle » et « d'expert-es » sont indépendants les uns des autres : l'interdépendance de frontière est donc principalement axée sur l'individu en relation avec son groupe, et non pas sur les groupes entre eux. Il est tout à fait possible de structurer l'interdépendance de frontières en mettant, par exemple, des groupes entiers en compétition, comme c'est le cas pour la technique du « Team Game Tournament » (deVries & Slavin, 1978). L'environnement dans lequel évoluent les individus peut servir de marqueurs à ces relations d'interdépendance de frontières, en rendant saillantes les différences entre les individus ou les groupes, et en modifiant l'espace physique dans lequel les interactions prennent place.

3.3.2. Responsabilité individuelle et contribution identifiable

Outre la structure de la tâche, des moyens alloués et des récompenses associées à sa réalisation, il existe d'autres conditions pour que l'apprentissage coopératif soit efficace. L'interdépendance positive, telle que structurée par les moyens, buts ou frontières, est censée garantir l'émergence de comportements responsables et altruistes entre les membres. Si la réussite des uns dépend de la réussite de l'autre, chaque individu est en situation de « *sink or swim* » (Johnson & Johnson, 2005), c'est-à-dire responsable à la fois de sa propre réussite, mais aussi de celle des autres : soit tous les membres œuvrent pour la réussite du groupe, et « flottent » à la surface de la tâche, soit ils ne parviennent pas à coordonner leurs efforts, et « coulent ». En d'autres termes, l'interdépendance positive vise à créer et renforcer le sentiment de responsabilité des membres du groupe pour accomplir leur part du travail et faciliter celui des autres membres du groupe. Pour rendre les interactions entre membres plus efficaces, Johnson et Johnson (1989 ; 2002 ; 2005) estiment que les individus doivent disposer d'un sentiment de responsabilité envers leur groupe, en partie stimulé par la visibilité de la production de chacun·e. Pour ces auteurs, cette visibilité peut provenir à la fois de la manière dont sont évalués et récompensés les individus, mais également de leurs sentiments et de leur responsabilité vis-à-vis d'autrui. Dans le cas contraire, un sentiment de responsabilité trop faible ou une situation où les contributions individuelles ne sont plus visibles peuvent entraîner des

comportements contre-productifs pour le groupe. Karau & Williams (1993) parlent de « paresse sociale » pour désigner un comportement de laisser-faire, c'est-à-dire un désengagement dans l'activité du groupe. Dans l'apprentissage coopératif, la paresse sociale peut apparaître lorsque la taille du groupe augmente, puisque la contribution individuelle devient moins visible (Huguet et al., 1999). La taille des groupes est un critère important lors de la mise en place d'une pédagogie coopérative, puisque la coordination entre un nombre important d'individus peut représenter un coût cognitif trop important pour les individus et ainsi impacter négativement la réussite du groupe (voir Nokes-Malach et al., 2015 pour une synthèse). Si la plupart des chercheur·es s'intéressant à l'apprentissage coopératif situent la taille optimale des groupes entre 3 et 5 élèves, c'est parce qu'au-delà d'un certain nombre, le risque de paresse sociale augmente mécaniquement. Outre le contrôle de la composition des groupes, Slavin (1983) préconise l'utilisation d'évaluations collectives et individuelles afin de réduire le risque d'émergence de ce phénomène. Dans la théorie de l'interdépendance positive, de tels comportements ne sont pas censés émerger si les individus partagent le même but commun et se sentent responsables les uns des autres. Pour Johnson et Johnson (2005), c'est le sentiment d'unité et de cohésion sociale dans le groupe qui fait émerger un sentiment d'entraide, et donc protège contre la paresse sociale. De manière générale, le sentiment de responsabilité dépend d'un certain nombre de facteurs, dont l'un des plus critiques renvoie à la taille même du groupe ainsi qu'à l'utilisation convenable de compétences sociales.

3.3.3. L'utilisation de compétences sociales

Dans le cas d'une relation d'interdépendance positive entre individus, chacun est supposé s'entraider et orienter ses efforts vers l'accomplissement du but commun. Ces comportements sont appelés « facilitateurs » puisqu'ils permettent de favoriser la réussite de tout le groupe, à travers des actions altruistes envers les membres qui en ont besoin. De tels comportements génèrent en retour de la confiance et de la motivation qui influencent les relations entre les membres. Pour que ces comportements facilitateurs émergent du travail coopératif, selon ces auteur·es, les individus qui interagissent entre eux doivent maîtriser un certain nombre de compétences sociales (i.e., communication, empathie, écoute à tour de rôle). En effet, Johnson et Johnson (2002 ; 2005 ; 2009) rapportent que l'interdépendance positive est plus forte quand les individus parviennent à établir des relations de confiance et à résoudre des conflits de manière constructive. Ils développent l'idée selon laquelle les individus doivent avoir été formés au préalable à l'utilisation de ces compétences. Buchs et collaborateurs (2016) développent l'idée selon laquelle il est nécessaire de former les étudiant·es avant de les mettre

dans des situations d'apprentissage coopératif : dans leur étude, iels mettent en évidence qu'une simple explication de *pourquoi* et *comment* coopérer permet aux étudiant-es d'interagir de manière plus efficace. Johnson et Johnson (2005) soulignent également l'importance de préparer aussi bien les élèves que leurs enseignant-es à la coopération, à l'aide d'exercices de « *team-building* » permettant de favoriser la cohésion au sein des groupes.

3.3.4. Réflexion au niveau du groupe

Enfin, la réflexion au niveau du groupe, selon Johnson et Johnson (2005 ; 2009) doit remplir plusieurs objectifs. Les individus doivent établir s'ils sont parvenus à remplir les objectifs fixés (i.e., résolution de la tâche, production de groupe) et identifier les contributions respectives de chacun-e. Lorsqu'elle est appliquée pendant le fonctionnement du groupe, une telle réflexion (i.e., « *feedback* ») permet d'orienter les décisions et les stratégies à appliquer pour persister dans la tâche. C'est pendant ce processus de réflexion que les contributions respectives des membres sont mises en évidence. Dans leur étude, Johnson et collaborateurs (1990) observent que les individus ont davantage progressé lorsque les enseignant-es leur communiquaient un feedback au niveau du groupe et de manière individuelle : ce feedback consistait à expliquer les compétences sociales à utiliser (e.g., écoute attentive, explication, correction), puis à observer les individus pour ensuite évaluer leur utilisation de ces compétences. Comparativement à un groupe coopératif où le feedback était communiqué de manière globale, ou bien à une condition de travail individuelle sans feedback, la réussite des individus dans la condition de feedback individuel et de groupe était supérieure. Pour Archer-Kath (1994), un tel feedback permet d'améliorer la motivation et de réduire la paresse sociale, puisque les contributions individuelles sont rendues visibles au moment de l'évaluation de l'enseignant-e. Idéalement, cette réflexion permet de quantifier si les interactions facilitatrices entre individus ont bien eu lieu, dans le cadre de leur travail en groupe.

3.3.5. Interactions facilitatrices

Selon Johnson et Johnson (1989 ; 2002), les conditions de structuration de l'interdépendance positive telles que décrites précédemment permettent de faire émerger des interactions facilitatrices entre les individus, ce qui explique les gains sur les apprentissages. Les auteurs définissent les interactions facilitatrices comme : « *des interactions à travers lesquelles les individus encouragent et facilitent les efforts de chacun-es pour accomplir des tâches et atteindre les objectifs du groupe.* » (2002). Pour ces auteurs, ces interactions peuvent prendre la forme de comportements d'assistance, d'aide, et d'échange de ressources et d'informations. Dans la théorie de l'interdépendance positive, les interactions facilitatrices sont

la conséquence logique d'une structuration de l'apprentissage coopératif efficace : c'est parce que le cadre est propice à la coopération (e.g., « *sink or swim* ») que les membres du groupe vont s'engager dans les compétences interpersonnelles nécessaires à un travail d'équipe efficace.

4. Un aperçu des principales pédagogies coopératives

Comme développé précédemment, il existe de multiples pédagogies coopératives dont l'objectif commun est de proposer une manière de structurer l'interdépendance positive entre les membres du groupe. Dans le cadre de la théorie de l'interdépendance sociale positive, le but commun des membres a souvent une portée pédagogique (i.e., comprendre une leçon, rendre un rapport écrit, faire un exposé sur une thématique). Selon cette théorie, c'est parce que les individus interagissent entre eux de manière positive - donc facilitatrice - que des bénéfices sur les apprentissages peuvent être observés (Johnson & Johnson, 1989 ; 2005 ; 2009). Or, il existe plusieurs manières de cadrer, ou structurer, ces interactions.

Sharan (1980) offre une des premières revues des différentes pédagogies coopératives mises en place par les chercheur.es en psychologie sociale et de l'éducation. Il distingue les techniques STAD (*Student Team Achievement Division*, Slavin, 1978) et TGT (*Team Game Tournament*, DeVries & Slavin, 1978 cité par Sharan, 1980), qui structurent à la fois une interdépendance positive pour les membres d'un groupe, mais également de la compétition entre les groupes, aux autres approches qui n'utilisent pas de compétition. Dans l'approche STAD, les élèves travaillent en groupe de 4 à 5 individus sur une même leçon et doivent s'assurer que tout le monde parvient à un niveau de compréhension satisfaisant. L'évaluation se fait de manière individuelle, à l'aide de quiz portant sur la leçon, mais les scores prennent en compte la progression relative de chacun des membres par rapport à leur niveau de connaissance initial. Les groupes sont également évalués en agrégeant la somme des scores individuels, et les récompenses peuvent leur être attribuées sous forme de certificats qui attestent de leur réussite (Slavin, 1994 dans Slavin, 2011). D'après le concepteur de la méthode, STAD est particulièrement adapté aux contenus scientifiques, tels que les mathématiques et les contenus scientifiques, mais peut être utilisé dans une grande variété de disciplines, notamment les arts du langage et les études sociales. Plusieurs études sur le STAD ont rapporté des effets positifs sur les apprentissages en mathématiques, en langues, en sciences et dans d'autres matières, comparativement à des groupes sans coopération (Mevarech, 1985 ; Slavin, 1995 ; Slavin & Karweit, 1984).

Le TGT reprend la même structure que la STAD, en ajoutant un élément de compétition entre les différents groupes. En effet, plutôt que d'effectuer des quiz de manière individuelle, les groupes doivent « s'affronter » et gagner des points qui seront assignés à tous les autres membres, de manière équivoque. D'autres techniques, comme Jigsaw (Aronson et al., 1978) n'utilisent ni compétition, ni récompense de groupe. Dans cette approche, les élèves disposent de ressources complémentaires et suivent une procédure en plusieurs étapes dans lesquelles ils assurent successivement le rôle « d'expert », sur la ressource qu'ils possèdent, et de « novice », qui reçoit une information de la part de ses camarades. Cette pédagogie est particulièrement populaire, notamment grâce aux prises de positions politiques de ses créateurs (e.g., sur la déségrégation des écoles américaines, Aronson & Bridgeman 1979; et sur la tuerie de Columbine, voir Gilbert, 2001 pour l'interview de Elliot Aronson) et à la visibilité du site internet jigsaw.org (i.e., plus de 5 millions de visites), qui contribuent à son exportation (voir chapitre 2). L'approche de Sharan et Sharan (1976), dénommée « Group Investigation » reprend quelques éléments de Jigsaw dont la répartition des ressources. Dans ces groupes, les étudiant.es doivent étudier un sujet divisé en plusieurs parties. Chaque étudiant.e est libre de choisir l'information qu'il souhaite étudier, avec pour objectif final de présenter un produit de groupe (par exemple, un exposé) à toute la classe. Enfin, Johnson & Johnson (1975, dans Newmann & Thompson, 1987) utilisent « Learning Together » pour faire coopérer des étudiant.es sur un sujet commun, avec les mêmes libertés que dans le « Group Investigation ».

Slavin (1983) regroupe ces différentes pédagogies selon deux principaux critères : la structure de la tâche, avec ou sans spécialisation (i.e., de rôles ou de structure) et le type de récompense attribuée. Pour lui, une pédagogie coopérative doit être articulée autour d'une tâche et d'une structure qui encouragent la coopération entre membres. Il suggère notamment que la coopération ne peut être efficace que si la tâche requiert une coordination des efforts, c'est-à-dire a minima une interdépendance des moyens. Slavin (1983) inclut dans sa revue des expérimentations qui testent l'efficacité des approches coopératives en classe réelle, comparativement à un groupe contrôle sans interdépendance positive. Les résultats de cette méta-analyse indiquent que 63% des études (29 sur 46) rapportent un effet positif de l'apprentissage coopératif, tandis que 33% (15/46) ne rapportent pas d'effet et 4% (2/46) un effet négatif. Il met en contraste ces effets globalement positifs au faible taux de réussite pour Jigsaw (i.e., 67% d'effets nuls, 16% d'effets positifs et 16% d'effets négatifs) qu'il attribue à l'absence de récompense de groupe. Bien qu'il rapporte un effet positif de l'approche Jigsaw II - une version qu'il développe en 1980 et qui intègre une évaluation de groupe similaire à

l'approche STAD - cette observation ne porte que sur une étude (Ziegler, 1981) et il est donc difficile de conclure quant à son efficacité (voir chapitres 2 et 4). Dans une autre méta-analyse s'intéressant aux différences d'efficacité entre ces approches, Johnson et collaborateurs (2000) apportent des éléments proches de ceux de Slavin (1980 ; 1983) : les techniques qui cumulent la présence d'interdépendance positive de buts et de récompenses de groupe sont celles qui sont reliées à des gains plus importants sur les apprentissages (voir Figure 3). Bien que ces éléments soient purement descriptifs, Johnson et collaborateurs (2000) invitent les chercheurs à multiplier les recherches expérimentales sur les approches qui n'ont pas encore fait l'objet d'une validation empirique jugée suffisante, notamment Jigsaw, Group Investigation (GI) et TGT.

Il est important de noter que, à notre connaissance, la méta-analyse de Johnson et al. (2000) est la seule à avoir comparé l'effet de différentes méthodes sur les apprentissages. En effet, la plupart des méta-analyses récentes rapportent des effets généraux de l'apprentissage coopératif, quelle que soit la méthode utilisée. Nous en présenterons les principaux résultats dans le paragraphe suivant.

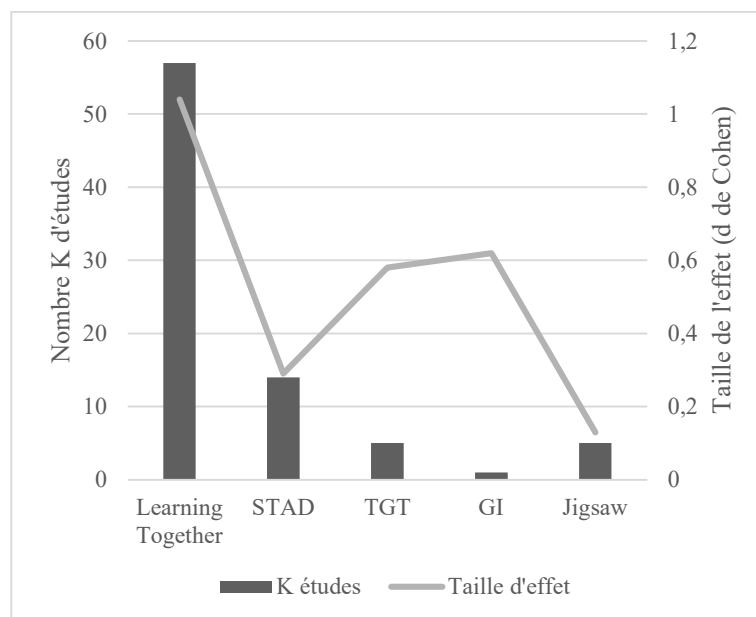


Figure 3: Tailles d'effet et nombre d'études associées aux pédagogies coopératives comparativement à une approche individuelle (Johnson, Johnson & Stanne, 2000)

5. Les effets de l'apprentissage coopératif sur les performances scolaires

Depuis sa théorisation et son opérationnalisation dans la recherche expérimentale, plusieurs méta-analyses regroupant les études sur l'apprentissage coopératif ont été publiées dans des revues à comité de lecture. Le Tableau 1 ci-dessous reprend les principaux résultats de ces méta-analyses. Pris dans leur ensemble, ces méta-analyses apportent un soutien empirique robuste à la théorie de l'interdépendance sociale positive : malgré des dénominations différentes (e.g., apprentissage coopératif, travail en petit groupe, apprentissage guidé par les pairs), les effets rapportés dans ces six méta-analyses sont globalement positifs et d'amplitude moyenne ($min = 0.17$; $max = 1.19$; $médiane = 0.33$). Les travaux de John Hattie (2009), qui synthétisent plusieurs méta-analyses sur différents types d'intervention visant à améliorer la réussite scolaire, rapportent des effets de l'apprentissage coopératif d'une amplitude proche (voir Tableau 1). C'est sur la base d'un nombre important d'études que Hattie (2015) classe l'apprentissage coopératif parmi les interventions les plus prometteuses pour améliorer la réussite des étudiant·es.

Intervention	<i>N</i> méta-analyses	<i>k</i> études	<i>ES</i> (<i>se</i>)
Apprentissage coopératif	10	306	0.41 (.06)
Apprentissage coopératif vs compétitif	7	1024	0.54 (.11)
Apprentissage coopératif vs individualiste	4	774	0.59 (.09)

Note : *ES* = estimation de taille d'effet ; *se* = erreur standard (précision) de l'estimation.

Tableau 1: Méta-analyses et effets de l'apprentissage coopératif sur la réussite académique (Hattie, 2009)

Toutes ces méta-analyses s'intéressent aux effets de l'apprentissage coopératif 1) en milieu de classe naturelle, 2) sur des mesures d'apprentissages via des indicateurs de performances scolaires (e.g., tests de connaissances, évaluations sommatives), 3) en comparaison à un groupe contrôle d'instruction « frontale », sans interdépendance positive et 4) sur des designs expérimentaux ou quasi expérimentaux. Il existe toutefois de nombreuses différences, notamment dans les comparaisons et les modérateurs pris en considération dans les analyses et les descriptions, c'est pourquoi nous détaillerons les principaux résultats de ces méta-analyses dans les prochains paragraphes (pour une synthèse, voir Tableau 2).

La méta-analyse de Newmann et Thompson (1987) cherche principalement à tester les différences d'efficacité entre les principales approches coopératives présentées précédemment. En comparant le nombre d'effets positifs rapportés dans la littérature, les chercheurs observent

que les pédagogies STAD (89%), Learning Together (73%) et Group Investigation (67%) ont des taux généralement plus élevés de bénéfiques sur les apprentissages, comparativement à Jigsaw (17%) qu'ils présentent comme « la moins efficace des pédagogies coopératives » (Newmann & Thompson, 1987). Néanmoins, ces comparaisons sont effectuées de manière strictement descriptives (i.e., comparaison de pourcentage, en absolu), et reposent sur une méthode de « *vote counting* » qui ne permet pas de tenir compte statistiquement des différences entre et à l'intérieur même des expérimentations (voir Borenstein et al., 2010 pour la critique d'une telle approche). En effet, la méthode du « *vote counting* » considère qu'un effet n'existe pas quand le résultat d'une étude est « non significatif » et ne permet pas de quantifier la dispersion des effets (voir chapitre 4 et 5), ce qui la rend insuffisante selon ces auteur·es. La méta-analyse de Lou et collaborateurs (1996) offre une vision un peu plus détaillée de ces différences, bien que l'objectif des auteur·es n'est pas de comparer différentes pédagogies entre elles, mais plutôt de rapporter un effet global de l'apprentissage en groupe sur les performances académiques. À l'aide d'analyses d'hétérogénéité, plusieurs modérateurs de l'efficacité du travail en groupe sont mis en évidence : iels observent notamment que l'apprentissage coopératif, caractérisé par la présence d'interdépendance positive entre individus, a des effets plus élevés sur la performance ($\delta = 0.28$, 95% CI [0.21 ; 0.34]) qu'un travail de groupe sans interdépendance positive ($\delta = 0.15$, 95% CI [0.11 ; 0.19]), ce qui supporte empiriquement la théorie de l'interdépendance positive. Les auteur·es observent également des différences en fonction du niveau de scolarisation, avec des effets plus élevés chez les enfants de CM1 à 6^e ($\delta = .29$, 95% CI [0.24 ; 0.34]) que ceux de l'enseignement secondaire ($\delta = 0.17$, 95% CI [0.12 ; 0.23]) ou du supérieur ($\delta = 0.19$, 95% CI [-0.03 ; 0.42]).

Springer et collaborateurs (1999) apportent des éléments méta-analytiques propres aux étudiant·s nord-américain·es dans l'enseignement supérieur. L'effet global observé sur la base de 37 études est positif et d'amplitude « moyenne » ($\delta = 0.51$). Des différences sont observées en fonction de la composition des groupes : les effets observés sur la réussite académique sont plus élevés dans les groupes composés majoritairement d'étudiant·es d'origine minoritaire (i.e., Mexicains et Afro-Américains ; $\delta = 0.76$) que dans les groupes composés majoritairement d'étudiant·es d'origine ethnique majoritaire (i.e., blanche et caucasienne ; $\delta = 0.46$).

Auteur.es	<i>k</i> études	Période	Niveau scolaire	Disciplines	Dénomination	Durée	Résultats principaux	Taille d'effet
Newmann & Thompson (1987)	27	1972-1986	Secondaire	Toutes disciplines confondues	« Apprentissage coopératif »	> 2 semaines	<i>STAD</i> est la structure la plus efficace (89% de résultats positifs) et <i>Jigsaw</i> la moins efficace (17%)	1.19
Lou et al. (1996)	51	1942-1996	Élémentaire, Secondaire et Supérieur	Mathématique et scientifique vs littéraire et sociale	« Regroupement intraclasse »	> 1 jour	L'apprentissage coopératif a en moyenne des effets positifs sur les apprentissages	0.28
Springer et al. (1999)	37	1980-1996	Supérieur	Science, Technologie, Ingénierie, Mathématiques	« Apprentissage en petit groupe »	Non précisé	Effets positifs et moyens : des différences en fonction du type de mesure	0.51
Rohrbeck et al. (2003)	81	1972-2000	Élémentaire	Mathématiques vs « autres »	« Apprentissage assisté par les pairs »	> 1 semaine	Effets positifs et moyens : bénéficie davantage aux élèves défavorisés et aux minorités ethniques.	0.33
Puzio & Colby (2013)	18	1980-2009	Élémentaire et Secondaire	Littéraire	« Apprentissage coopératif »	Non précisé	Effets positifs, mais faibles sur les mesures de lecture, compréhension et vocabulaire.	0.19
Kyndt et al. (2013)	65	1995-2010	Élémentaire, Secondaire et Supérieur	Mathématique, scientifique vs littéraire et sociale	« Apprentissage coopératif en face à face »	> 1 jour	Effets moyens sur l'apprentissage, de multiples modérations : discipline, âge et "culture"	0.54

Tableau 2: Compilation de méta-analyses portant sur les effets de l'apprentissage coopératif en comparaison à un groupe contrôle sur les apprentissages

La méta-analyse de Rohrbeck et collaborateurs (2003) porte sur une version spécifique de l'apprentissage coopératif : le tutorat par les pairs. Bien que les objectifs d'une telle intervention soient souvent comparables à ceux qui sous-tendent la mise en place d'une pédagogie coopérative, la particularité du tutorat réside dans le fait que les individus collaborent en dyades, c'est-à-dire en groupes de deux. Les résultats de la méta-analyse de Rohrbeck et collaborateurs (2003), qui se focalisent sur les populations d'élèves d'école primaire élémentaire, montrent le tutorat par les pairs est efficace : la taille d'effet rapportée est de $\delta = 0.33$, 95% CI [0.29 ; 0.37], avec des effets positifs sur toutes les disciplines étudiées (e.g., mathématiques, $\delta = 0.22$; études sociales, $\delta = 0.49$; lecture, $\delta = 0.27$). Avec des analyses d'hétérogénéité, les auteur·es mettent également en évidence des différences en fonction de la composition des groupes. Comme dans les travaux de Springer et collaborateurs (1999), les effets du tutorat semblent plus efficaces pour les élèves issues de minorités ethniques ($\delta = .51$, 95% CI [.36 ; .60]) que pour les élèves de majorité ethnique ($\delta = 0.23$, 95% CI [0.16 ; 0.29]). Le statut socio-économique a également un effet modérateur sur l'efficacité d'une telle intervention, puisque les élèves d'origine sociale défavorisée bénéficient davantage de l'apprentissage par les pairs ($\delta = 0.45$, 95% CI [0.32 ; 0.57]) que les élèves d'origine sociale favorisée ($\delta = 0.32$, 95% CI [0.24 ; 0.39]). Pour ces auteur.es, de tels résultats sont prometteurs pour la prise en compte de la vulnérabilité de certaines populations, notamment les minorités ethniques de niveau socio-économique défavorisé.

Kyndt et collaborateurs (2013), quant à elleux, ne testent pas de modération par la composition du groupe, sur la base des caractéristiques démographiques *per se*, mais utilisent un critère « culturel » pour regrouper les études effectuées dans les sociétés « occidentales » (i.e., États-Unis et Europe) ou « orientales » (i.e., Moyen-Orient et Asie). Leurs résultats indiquent que l'apprentissage coopératif en « face à face » a des effets plus marqués au sein des sociétés « orientales » qui sont plutôt collectivistes, que les sociétés « occidentales » qui sont plutôt individualistes (Kitayama et al., 1997 ; Oyserman et al., 2002), avec une différence statistiquement significative de l'ordre de $\delta = .38$, 95% CI [.25 ; .53]. De manière plus globale, les auteur.es rapportent que l'apprentissage a des effets positifs sur la réussite académique, $\delta = 0.54$, 95% CI [0.47 ; 0.60]. Pour finir, la méta-analyse de Puzio et Colby (2013) s'intéresse aux effets de l'apprentissage coopératif sur la réussite académique dans les matières littéraires, chez les élèves scolarisés à l'école primaire. Les auteur·es rapportent des effets positifs, mais d'amplitude relativement faibles en lecture, $\delta = 0.16$, 95% CI [0.07 ; 0.25], en compréhension, $\delta = 0.20$, 95% CI [0.13 ; 0.28], ainsi qu'en vocabulaire, $\delta = 0.22$, 95% CI [0.13 ; 0.32].

Cependant, aucun facteur modérateur (e.g., facteurs démographiques) n'est testé dans cette méta-analyse.

Pris dans leur ensemble, ces méta-analyses dressent un tableau plutôt positif des effets de l'apprentissage coopératif sur la réussite académique. Les effets observés sont systématiquement positifs et leur amplitude varie entre des amplitudes « faibles » (i.e., $\delta = 0.16$ dans Puzio & Colby, 2013) jusqu'à « moyennes » (i.e., $\delta = 0.54$ dans Kyndt et al., 2013). Néanmoins plusieurs questions subsistent, notamment celle de l'efficacité relative de chaque pédagogie coopérative. En effet, aucune comparaison d'efficacité n'a été faite entre les différentes manières de structurer l'interdépendance positive depuis la méta-analyse de Johnson et collaborateurs (2000). En conséquence, de nombreux points d'ombres subsistent sur les relations entre les conditions de mise en place de ces pédagogies et leurs effets sur les apprentissages, tels que mesurés par des indices de performances académiques.

6. De la théorie à la pratique

Comme développé dans l'introduction de ce manuscrit, il semble qu'il existe aujourd'hui une disparité entre la popularité et la robustesse de la théorie de l'interdépendance positive telle que décrite par Johnson et Johnson (2009) et sa mise en place au sein des classes par les enseignant·es (Enquête PériODique sur l'Enseignement, 2018 ; Pianta et al., 2017). Ce problème n'est pas spécifique à la discipline et peut traduire une difficulté d'importer des concepts de recherche à un milieu professionnel (voir Burkhardt, 2003 pour une discussion relative à la psychologie de l'éducation).

Il existe également une autre rupture qui concerne particulièrement une méthode coopérative : *Jigsaw*. En effet, d'après ses concepteur·rices, cette pédagogie permet de réduire les conflits ethniques parmi les élèves, d'améliorer la motivation ainsi que les attitudes envers l'apprentissage et de promouvoir un meilleur apprentissage (Aronson & Patnoe, 2011). Or, de telles affirmations, relayées de manière importante par les chercheur·es (e.g., Hattie, 2017) et par les médias (e.g., Arenas & Meirieu, 2020 ; Le Monde, 2018) ne correspondent pas à la réalité empirique caractérisée par des résultats beaucoup plus mitigés (Newmann & Thompson, 1987 ; Johnson et al., 2000). Bien que ces auteur·es attestent l'existence d'un consensus des effets positifs de l'apprentissage coopératif en général, des écarts importants d'efficacité subsistent entre les différentes pédagogies coopératives (Johnson & Johnson, 2005 ; Slavin, 2016). Il est nécessaire d'évaluer l'efficacité d'une intervention et d'examiner les conditions permettant une bonne implémentation avant de la mettre en place sur le terrain. Comme le

conseillent Johnson et collaborateurs (2000) dans leurs travaux : « [...] *Si tout enseignant-e peut développer sa version de l'apprentissage coopératif « efficace », sans études ni recherche, on ne peut pas savoir si les autres enseignant-es peuvent s'attendre à des résultats fiables lorsque cette méthode est utilisée. Ainsi, les méthodes non évaluées doivent être utilisées avec prudence. Il apparaît nécessaire de disposer d'une nouvelle génération de chercheur-es qui formuleront une nouvelle opérationnalisation de la coopération dans la classe et l'école, et qui soumettent leurs formulations à une évaluation empirique rigoureuse* » (p.15). Dans le chapitre suivant, nous présenterons une revue de littérature des travaux menés sur Jigsaw, depuis sa création jusqu'à aujourd'hui.

Résumé du chapitre 1

- La théorie de l'interdépendance sociale positive suggère qu'il est possible d'améliorer l'apprentissage des individus en les faisant coopérer sous certaines conditions, notamment la présence d'un but commun et la responsabilité individuelle.

- Les nombreux travaux et méta-analyses mettent en évidence des effets positifs de l'apprentissage coopératif - comparativement à un groupe contrôle individualiste ou compétitif - dont l'amplitude moyenne est proche de $\delta = 0.40$.

- Certaines pédagogies, c'est-à-dire certaines manières de structurer l'interdépendance n'ont pas reçu une validation empirique importante : c'est notamment le cas de la méthode *Jigsaw*, marquée par une popularité importante, mais un faible nombre d'études testant son efficacité dans des protocoles randomisés et puissants.

Chapitre 2. La classe puzzle ou « Jigsaw » : Approche historique, théorique et empirique

1. Objectif du chapitre

Dans ce chapitre, nous ferons un état de l'art de la littérature scientifique s'intéressant à la méthode Jigsaw. Comme vu dans le chapitre précédent, Jigsaw est une pédagogie coopérative relativement populaire (Aronson & Patnoe, 2011), malgré ses effets assez mitigés sur les apprentissages (Newmann & Thompson, 1987 ; Johnson et al., 2000). Nous commencerons par présenter le contexte historique dans lequel cette méthode a émergé, ainsi que les principales étapes de son fonctionnement. Ensuite, nous présenterons plusieurs études empiriques - en commençant par les études *princeps* jusqu'aux travaux plus récents - et conclurons sur les enjeux actuels de la recherche sur Jigsaw.

2. Contexte historique

L'origine de *Jigsaw* - ou « classe puzzle » - remonte aux travaux d'Elliot Aronson et de ses collègues dans les années 70 aux États-Unis. Dans cette période, la déségrégation des écoles, c'est-à-dire l'inclusion des minorités ethniques, venait d'être mise en place, et cette situation faisait émerger des conflits entre élèves. Inspirés par la théorie du contact d'Allport (1954) et par la dynamique des groupes de Sherif et collaborateurs (1961), Elliott Aronson et ses collègues ont décidé de mettre en place une pédagogie coopérative pour réduire ces conflits interethniques : *Jigsaw* (Aronson et al., 1978 ; Aronson & Bridgeman, 1979 ; Lucker et al., 1976). L'hypothèse du contact d'Allport (1954) établit que des relations intergroupes, c'est-à-dire des interactions entre individus issus de groupes sociaux différents, peuvent permettre de réduire les préjugés lorsque certaines conditions sont remplies : parmi elles, la présence d'un but commun qui nécessite une coordination des efforts, des statuts égaux entre les individus au sein du groupe qui suggère une minimisation des différences d'origine sociale et de compétences, ainsi que la coopération et le support institutionnel. Pour Aronson et ses collaborateurs (1978), Jigsaw devait permettre de réduire les préjugés ethniques entre élèves en les faisant coopérer sur une tâche commune où chacun·e pouvait apporter une contribution unique, et donc être sur un pied d'égalité (i.e., un statut égal). Dans le même ordre d'idées, les travaux sur la théorie des conflits intergroupes développent l'argument selon lequel la coopération ainsi que la présence de buts communs (i.e., buts d'intérêts supérieurs ou « superordonnés ») permettent de réduire les conflits entre groupes. Les expérimentations de Sherif prenaient place dans un camp de vacances où des groupes d'enfants étaient créés puis

mis en compétition à travers une série de jeux. La mise en concurrence des groupes, avec l'attribution d'une récompense unique pour le groupe vainqueur de la compétition, a rapidement dégénéré et des comportements hostiles (e.g., stéréotypes, raids dans les camps adverses) sont apparus. Afin de réduire cette hostilité, Sherif et ses collègues ont mis en place une série de buts communs qui nécessitaient la coordination d'individus issus des deux camps : cette coopération a permis de réduire l'hostilité et les conflits intergroupes initialement causés par la compétition.

La proposition d'Elliott Aronson et de ses collègues s'intègre dans la continuité de ces travaux sur la coopération. Pour ces auteurs, la prédominance de la compétition au sein des classes peut amener à des émotions négatives (e.g., jalousie, suspicion, sabotage) et en conséquence, des opportunités d'apprentissage moindres. Pour ces auteur-es, bien que la coopération entraîne de meilleurs apprentissages que la compétition (Qin et al., 1995 ; Stanne et al., 1999), la structure des classes dans les pays occidentaux favorise les contextes individualistes et compétitifs ce qui peut avoir pour effet de mettre en difficulté les élèves (Butera et al., 2006 ; Di Stasio, 2016). Selon elleux, un contexte compétitif n'encourage ni le soutien, ni l'entraide ou le développement d'un climat de classe favorable, mais entraîne plutôt un phénomène pervers qui tire les élèves vers le bas. Pour Aronson et ses collaborateurs (2011), la mise en place d'un contexte d'interdépendance positive entre les élèves à travers Jigsaw doit s'articuler selon les préceptes suivants :

1. L'apprentissage doit se faire de manière coopérative : la coordination des efforts est essentielle pour que tous les élèves réussissent.
2. La structure de Jigsaw doit permettre une telle articulation des efforts : la réussite du groupe doit passer par la réussite individuelle de chacun des membres.
3. Chaque élève - quel que soit son statut dans la classe - doit être en position d'apporter à ses camarades une partie de l'information nécessaire, une « expertise », au groupe.

C'est un changement de paradigme important au sein de la classe : plutôt que d'avoir un système favorisant la compétition et le mérite « individuel », c'est-à-dire la capacité à fournir des réponses correctes et attendues par l'enseignant e, la méthode Jigsaw récompense les comportements d'aide, l'attention portée au pair, l'inclusion et le partage (Aronson & Bridgeman, 1979 ; Aronson & Patnoe, 2011 ; Robert et al., *en préparation*). Puisqu'il est attendu de cette méthode qu'elle améliore les relations interpersonnelles, en modifiant le cadre scolaire dans lequel ont lieu les interactions sociales, ainsi que le sentiment d'appartenance et l'estime de soi, il est avancé que cette méthode devrait également avoir des effets positifs sur la réussite académique des élèves (Lucker et al., 1976 ; Aronson et al., 1978). Selon la théorie

de l'interdépendance positive, c'est l'interdépendance de ressources (i.e., complémentarité des contenus) et de buts (i.e., accéder à la représentation complète de la leçon) qui expliqueraient, dans un tel contexte, les effets positifs sur les performances scolaires (voir Tableau 3). Dans la partie suivante, nous développerons le fonctionnement de Jigsaw, tel que pensé par ses concepteur·es (Aronson et al., 1978 ; Aronson & Patnoe, 2011).

Piliers de l'interdépendance positive (Johnson & Johnson, 1989)	Application à Jigsaw (Aronson et al., 1978)
Relations positives entre membres	Interdépendance des ressources
But commun	Compréhension de la leçon
Compétences sociales	Entraides dans les phases « puzzle » et « expert »
Responsabilité individuelle	Complémentarité des ressources (« <i>sink or swim</i> »)
Réflexion au niveau du groupe	Non spécifié dans les premières versions de Jigsaw

Tableau 3: Structuration de l'interdépendance positive dans Jigsaw

3. Déroulement de Jigsaw

De manière générale, la méthode Jigsaw comprend trois étapes principales qui structurent la coopération entre les élèves : une phase « individuelle », une phase de « groupe d'expert » et une phase de « groupe puzzle ». Le site internet jigsaw.org propose une version détaillée de la procédure en 10 étapes, bien que la plupart des recherches effectuées sur Jigsaw ne fassent mention que des trois étapes principales de la pédagogie (voir Robert et al., *en préparation*, pour une revue).

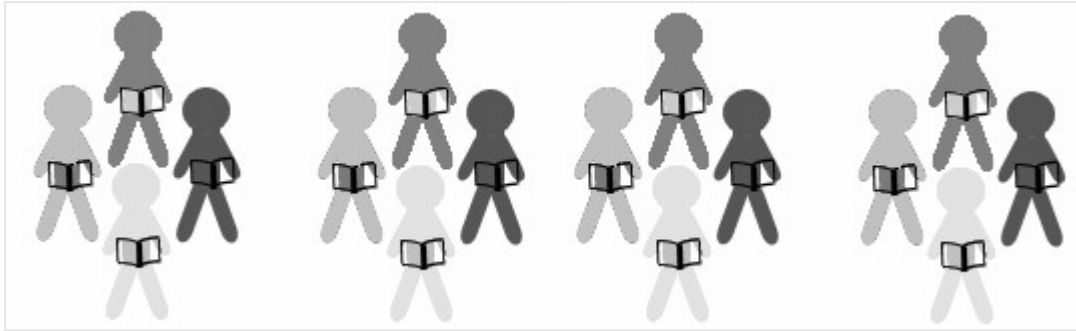


Figure 4: Étape 1 de Jigsaw (travail individuel)

La première étape d'une classe puzzle consiste en la création de groupes d'élèves : les groupes « puzzle » (voir Figure 4 ci-dessus). Cette étape implique que l'enseignant·e en charge du cours ait préparé, au préalable, un contenu pédagogique divisé en plusieurs sous-parties ; en effet, chaque membre du groupe puzzle devra recevoir une sous-partie du contenu pédagogique traité pendant le cours. Prenons l'exemple d'un contenu de Sciences de la Vie et de la Terre (SVT) de niveau 6e dont la thématique repose sur les liens de parenté entre espèces. L'enseignant·e choisit de faire travailler ses élèves sur quatre activités différentes : le remplissage de tableau sur la base de caractéristiques possédées par l'animal (activité A), la création de classification par groupes emboîtés (activité B), le classement des espèces en fonction de leurs attributs (activité C) et la description d'un milieu de vie (activité D). Ainsi, les groupes puzzle sont constitués de quatre élèves, disposant chacun d'une sous-partie de la thématique générale (i.e., la parenté entre espèces) à traiter. Selon Aronson et Patnoe (2011), la constitution des groupes doit être relativement hétérogène, de telle manière à ce que des élèves de différents niveaux et de différents statuts travaillent ensemble³.

En garantissant une interdépendance positive entre les élèves, les plus faibles peuvent bénéficier de l'aide des plus forts, qui eux-mêmes progressent en expliquant les contenus pédagogiques à leurs pairs. Les objectifs de cette première étape sont multiples. Tout d'abord, chez les jeunes enfants, une telle répartition des tâches permet d'identifier les membres de son groupe, c'est-à-dire les camarades avec qui auront lieu les interactions. Cette première étape sert également aux élèves à prendre connaissance de leur sous-partie. De manière individuelle, ils doivent se familiariser avec le contenu. À des niveaux d'enseignement secondaire ou primaire, l'usage de guides, sous forme de questions et de points importants, a été rapporté par certains auteurs afin de faciliter la compréhension des élèves (Souvignier & Kronenberger,

³ Voir Hooper et Hannafin (1988) ou Baer et collaborateurs (2010) pour des arguments contraires.

2007). Une fois que cette première étape a été effectuée, les élèves sortent temporairement de leur groupe puzzle pour former des groupes « d'expert·es » (voir Figure 5).

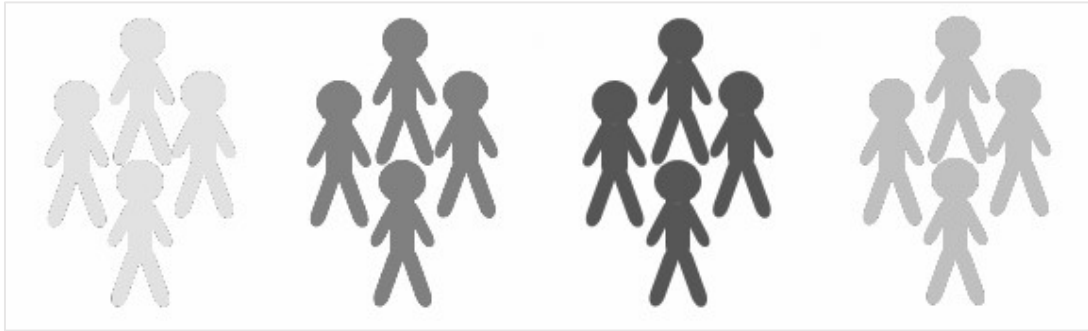


Figure 5: *Étape 2 de Jigsaw (groupe d'expert·es)*

Dans cette seconde étape, l'enseignant·e forme des groupes « experts » constitués d'élèves qui ont travaillé sur la même sous-partie. Idéalement, la taille des groupes est similaire à celle des groupes puzzle (e.g., quatre ou cinq élèves) puisque les risques de comportements contre-productifs peuvent mécaniquement émerger avec l'augmentation de la taille du groupe (voir « paresse sociale », Karau & Williams, 1993 dans le chapitre 1). Dans Jigsaw cette étape est particulièrement importante, car elle fournit aux élèves les plus en difficulté l'occasion d'être aidé·es par leurs camarades plus expérimenté·es. L'objectif des groupes « expert » est de s'assurer que tout le monde maîtrise et comprend sa sous-partie. Pour ce faire, les élèves doivent s'entraider, synthétiser le contenu sur lequel ils ont travaillé, donner des conseils et des feedbacks aux autres expert·es et s'autoréguler. Les membres de ce groupe temporaire doivent également se préparer à « enseigner » leur sous-partie à leur camarade des groupes puzzle. Une fois que les expert·es se sont mis d'accord sur la meilleure façon de transmettre les informations, chaque élève retourne dans son groupe puzzle initial et présente à tour de rôle les informations spécifiques à sa sous-partie. C'est la préparation à cette restitution de connaissances qui constitue l'enjeu de la phase « expert ». Dans notre exemple, cette partie consisterait à comparer les réponses, notamment les classements des espèces dans les différents tableaux, et la réalisation de classification par groupes.

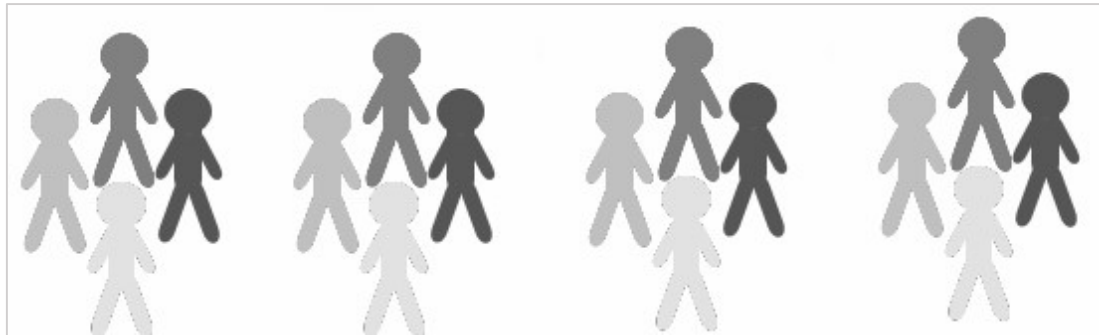


Figure 6: Étape 3 de Jigsaw (phase puzzle)

Enfin, les élèves retournent dans leurs groupes d'origine et présentent le matériel sur lequel ils sont devenus « experts » à leurs camarades (voir Figure 6). À tour de rôle, les élèves sont donc « enseignant·es » (ou « expert·es ») et « enseigné·es » (ou « novices »). En plus d'une interdépendance de ressources, mise en place par la distribution de contenus différents entre les élèves, il existe une interdépendance de rôles dans cette étape. Il est donc attendu que les élèves présentent un certain nombre de compétences sociales : l'écoute attentive, l'assiduité, la prise de parole en tour par tour sont des comportements qui favorisent l'émergence d'interactions facilitatrices, et assurent la réussite du groupe. Les présentations sont effectuées à tour de rôle et supervisées par l'enseignante·e, mais Aronson et ses collègues (1978) conseillent de prévoir un temps collectif pour revenir sur les différentes parties, en classe entière. Ainsi, chaque groupe accède à la totalité du contenu de cours en mettant en commun les connaissances des différents membres, à la manière d'un puzzle. Dans notre exemple, une fois les activités expertes réalisées, les élèves doivent remplir un tableau de classification et réaliser un arbre de parenté d'espèces qui n'ont pas été traitées dans les activités « d'expert ». C'est donc la mise en commun de connaissances et de compétences qui permet de remplir cet objectif.

4. Les études menées sur Jigsaw, par période

4.1. Revues systématiques de la littérature

Dans cette partie, nous présenterons les différents articles empiriques testant l'efficacité de Jigsaw sur différentes variables (e.g., réussite académique, réduction des préjugés, attitudes envers l'école et les pairs) et par période de publication. En effet, la littérature Jigsaw est marquée par au moins deux périodes distinctes (voir Figure 7). On peut y constater, dans un premier temps, un faible nombre d'articles empiriques rédigés entre les années 1970 et 2000, puis une augmentation rapide des publications, au détriment de la taille moyenne des échantillons (Robert et al., *en préparation*). Dès les années 2000, plusieurs versions « révisées »

de Jigsaw ont vu le jour (voir Tableau 4). Par exemple, Jigsaw II a été créée par Slavin (1980 ; 1983) afin d'ajouter une évaluation de groupe à Jigsaw, et ainsi accroître le sentiment d'interdépendance positive des individus. Jigsaw III rajoute une phase d'entraînement collective, sous forme de quiz distribué à tous les élèves, afin d'ajouter un feedback sur la performance individuelle de ces dernières avant leur évaluation individuelle (Stahl, 1994). Dans Jigsaw IV, chaque changement d'étape est marqué par la présence de quiz individuels, permettant de disposer des informations à communiquer aux autres membres du groupe, et donc de vérifier la responsabilité individuelle de chacun-e (Holliday, 2002). Enfin, Subject Jigsaw est développé spécialement pour l'enseignement de la physique et de la chimie, et comprend une phase de préparation du matériel en dehors du cours, à la manière d'une classe inversée (Doymus, 2007). Il faut cependant noter que seules les versions Jigsaw II et Subject Jigsaw ont fait l'objet d'une certaine validation empirique : les versions III et IV n'ont été développées qu'à travers des essais théoriques ou des articles présentant leur implémentation, mais ne rapportant pas d'analyses statistiques permettant d'évaluer leur efficacité.

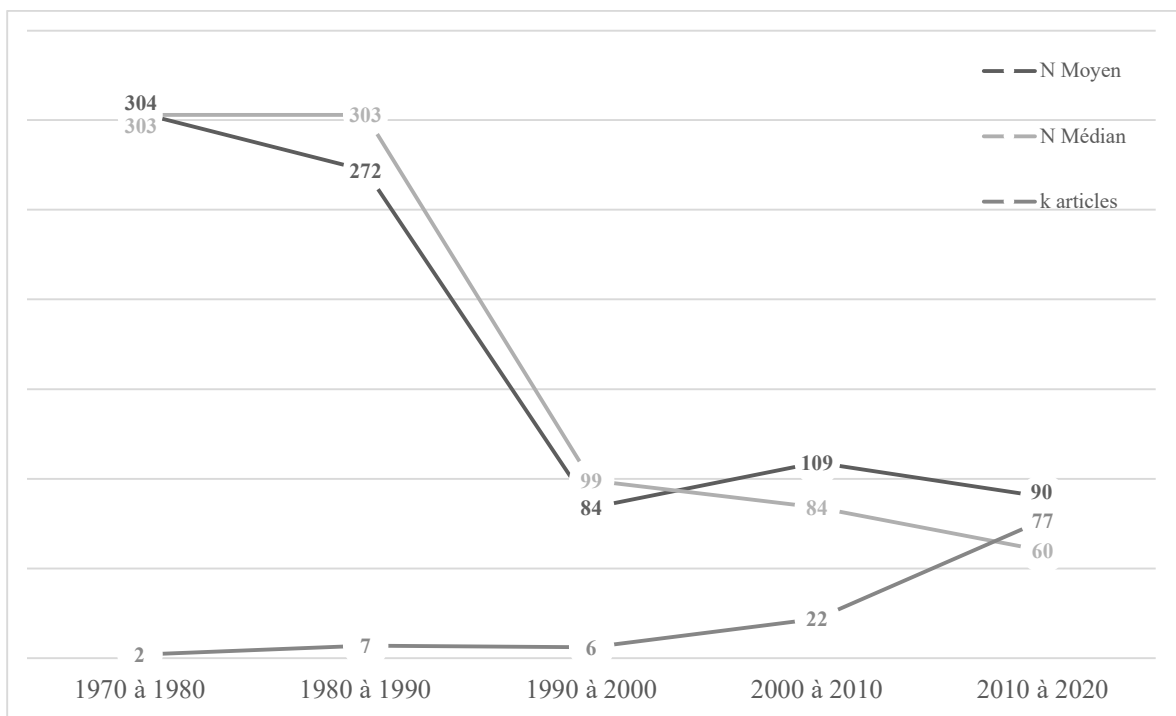


Figure 7: Évolution du nombre total d'articles (k) et des échantillons (N) dans les recherches Jigsaw (données provenant de Robert et al., en préparation)

Dans ce manuscrit, nous nous focaliserons particulièrement sur les articles empiriques, avec présence de groupe contrôle, publiés dans des revues à comité de lecture référencées dans des bases de données d'articles scientifiques (e.g., Scopus, Web of Science). Les articles ne rapportant pas d'analyses statistiques inférentielles ne seront pas abordés, au même titre que les articles publiés dans des revues non référencées ou ceux décrivant l'implémentation de Jigsaw. L'application de ces filtres consécutifs a pour conséquence d'exclure de notre analyse approximativement une centaine d'articles, soit 66% des articles rédigés jusqu'à ce jour : ces articles constituent des recherches non expérimentales ($k = 37$ articles d'implémentations et essais théoriques) ou bien ne sont pas publiés dans des revues à comité de lecture ($k = 65$). Ces derniers sont traités de manière détaillée dans des récentes revues systématiques de littérature (voir Vives et al., *en préparation* ainsi que Robert et al., *en préparation*). Il existe également une part non négligeable d'articles publiés avant et pendant les années 1970, qui ne figurent sur aucune base de données et ne seront par conséquent pas présentés dans ce manuscrit. De tels articles sont cependant mentionnés dans des revues et méta-analyses comme celles de Sharan (1980), Newmann et Thompson (1987) ou bien encore Slavin (1983).

4.2. La genèse : 1970-1980

Comme évoqué précédemment, les études *princeps* de Jigsaw avaient pour principal objectif de tester les effets de l'utilisation de Jigsaw sur la réduction des conflits interethniques. Toutefois, rapidement, des gains sur l'apprentissage ont été observés : les auteur·es ont alors suggéré l'existence de relations positives entre les variables socioaffectives (e.g., attitudes envers l'école et les pairs, estime de soi générale) et l'engagement scolaire, les gains sur les premières pourraient donc être associés à des gains en matière d'apprentissages (Aronson et al., 1978 ; Aronson & Bridgeman, 1979).

L'une des premières études menées avec la méthode Jigsaw et s'intéressant à ces gains provient de Lucker et collaborateurs en 1976. Elle consiste en une étude de terrain, chez des élèves de CM2 et 6^e, qui s'intéresse aux performances académiques des minorités ethniques (i.e., Mexicain·es et Afro Américain·es), comparativement à celles des majorités (i.e., « Anglo » soit anglophones, ou blanc·hes caucasien·nes). Sur la base de la théorie de l'interdépendance positive et suite à une étude pilote menée sur une centaine de participants, les auteurs ont fait l'hypothèse que les étudiant·es issu·es de minorités ethniques pourraient bénéficier de la coopération selon Jigsaw, sans détriment pour les majorités ethniques. Pour la tester, ils ont mis en place un protocole expérimental dans lequel des classes d'élèves étaient réparties aléatoirement dans une condition Jigsaw ou une condition d'enseignement habituel,

pour une durée de deux semaines. Dans la condition Jigsaw, les enseignant·es appliquaient la procédure telle que décrite par Aronson et collaborateurs (1978 ; 1997), alors que la condition d'enseignement habituel était davantage centrée sur la transmission de connaissances « frontale » de l'enseignant·e vers ses élèves, sans qu'il y ait de coopération entre eux. Leur apprentissage était mesuré à l'aide d'une évaluation de connaissances portant sur la leçon vue en cours, composée de 37 items de questions à choix multiples. Les résultats cette étude mettent en évidence un effet d'interaction entre l'origine ethnique et la condition expérimentale : les élèves en situation de minorité ethnique ont obtenu de meilleurs résultats lorsqu'ils ont étudié avec la méthode Jigsaw, comparativement à ceux qui ont suivi le cours de manière habituelle (Lucker et al., 1976). Ils n'observent pas de différences significatives entre les conditions Jigsaw et de classe habituelle chez les élèves en majorité ethnique, c'est donc principalement aux minorités qu'a bénéficié la méthode. D'après les auteurs, de tels gains reposent en partie sur le fait que les élèves minoritaires ont été davantage inclus et proches des élèves majoritaires : iels auraient alors bénéficié du support et de l'aide apportée par les autres élèves. Pour Lucker et collaborateurs (1976), le sentiment d'auto-efficacité et une motivation accrue seraient également des causes possibles expliquant les gains de performance observés chez cette population. D'autres explications sont également évoquées, telles que l'augmentation du taux de participation et d'engagement, c'est-à-dire l'assiduité dans la scolarité (Aronson et al., 1978).

Des résultats similaires sont rapportés par Blaney et collaborateurs (1977) qui utilisent une procédure similaire (i.e., design expérimental Jigsaw vs Enseignement habituel) sur des élèves de CM2 dans un cours de sciences sociales. D'après elleux, la coopération est reliée à des meilleures relations interpersonnelles, un climat d'apprentissage plus favorable et de la cohésion sociale. Pour ces auteur·es, l'utilisation de Jigsaw traduit une volonté d'instaurer des relations d'interdépendance positive via système de « tutorat par les pairs », responsables à la fois de leur propre apprentissage et de celui de leurs camarades. À travers leur étude, Blaney et collaborateurs (1977) s'intéressent à l'évaluation qu'ont les élèves de leurs pairs et du fonctionnement de leur classe. Leurs variables dépendantes sont donc centrées sur les aspects sociaux et affectifs, plutôt que sur les performances académiques *per se*. Iels rapportent que les élèves minoritaires ayant étudié avec Jigsaw appréciaient davantage leurs pairs et avaient une meilleure estime de soi générale (e.g., « *À quel point apprécies-tu être toi-même ?* ») et de meilleures attitudes envers l'école (e.g., « *À quel point es-tu heureux·se d'être en classe ?* ») que les élèves minoritaires de la condition contrôle. Iels formulent toutefois une limite observée chez les élèves d'origine Mexicaine-Américaine d'une des classes de leur étude, qui rapportent

une estime de soi moins élevée que les autres élèves. Selon les auteur·es, devoir expliquer son contenu aux autres peut exacerber la frontière langagière et menacer l'estime de soi de ces élèves (Blaney et al., 1977), mais cela n'explique pas pourquoi les effets sont positifs dans les 9 autres classes de leur étude.

4.3. La deuxième version de Jigsaw : 1980-1990

À la suite des premiers travaux menés par Aronson et collaborateurs (pour une synthèse, voir, Aronson & Bridgeman., 1979 ou Sharan, 1980), d'autres chercheur·es sont intéressé·es à Jigsaw comme manière de structurer l'interdépendance positive en classe. Dans la continuité de ces recherches, Diane Bridgeman (1981) s'intéresse aux effets de Jigsaw sur les compétences sociales et l'empathie. Pour mesurer le niveau d'empathie, elle utilise plusieurs dessins animés dans lesquels des personnages vivent des péripéties ; la tâche des participant·es consiste à inférer l'état mental d'un personnage prenant connaissance de différentes situations, c'est-à-dire de se mettre à la place de ces personnages. Les résultats de cette recherche indiquent que les élèves ayant profité de l'enseignement Jigsaw ont des scores d'empathie plus élevés que les élèves des classes contrôles, sans coopération. Néanmoins, de tels effets ne sont pas observés sur une autre mesure d'empathie, à savoir la qualité du jugement moral, mesuré à l'aide de situations de dilemmes moraux. Sur cette étude, les effets de Jigsaw sont donc mitigés.

Un des apports les plus importants de cette période réside dans les travaux de Moskowitz et collaborateurs (1983 ; 1985), qui se caractérisent par une tentative de réplique « mécanique » des études *princeps*. À travers deux études menées durant une année scolaire auprès d'une large cohorte d'élèves de 6^e et de CM2 (étude 1 : $N = 261$, $k_{\text{classes}} = 13$; étude 2 : $N = 384$, $k_{\text{classes}} = 24$) les chercheur·es ont tenté de répliquer les effets positifs de Jigsaw, tels que présentés par Aronson et ses collègues, sur les relations interpersonnelles et la réussite académique. Ces travaux sont parmi les premiers à remettre en question l'efficacité de Jigsaw. En effet, dans leurs tentatives de réplique, Moskowitz et collaborateurs (1983 ; 1985) n'observent pas les gains attendus, que ce soit sur les variables socioaffectives (i.e., estime de soi scolaire et sociale, attitudes envers l'école) ou sur les performances académiques. Même dans les classes ayant implémenté Jigsaw de manière fidèle (i.e., 3 classes sur 7 pour l'expérience de 1983 et 5 classes sur 11 pour l'expérience de 1985), les bénéfices de la méthode n'apparaissent pas. Pour ces auteur·es, l'absence d'effets positifs, même dans les groupes expérimentaux fidèles à la méthode, suggère que Jigsaw n'est peut-être pas si efficace que rapporté par ses concepteur·rices. D'après eux, ceci pourrait être dû à un problème dans la structure de Jigsaw : l'absence de production ou de récompense commune aurait pour effet de

réduire drastiquement l'interdépendance positive entre les élèves. Sans but commun, le système de récompense est alors individualiste (e.g., « chacun pour soi ») voire compétitif (e.g., « mieux réussir que les autres »). Outre le fait que cette recherche soit l'une des premières tentatives de réplification des effets de Jigsaw, Moskowitz et collaborateurs (1983 ; 1985) ont également apporté d'autres contributions, notamment au regard du critère d'implémentation de la méthode. En effet, après avoir observé la manière dont Jigsaw avait été mise en place dans les différentes classes - alors que chaque enseignant-e avait suivi la même formation au préalable - les chercheur-es ont reporté de larges différences dans leur pratique : « *Dans la présente étude, les enseignants ont rapporté avoir maîtrisé la mise en place de Jigsaw et trouvé la méthode utile. [...]. Néanmoins, beaucoup d'enseignants ont substantiellement modifié Jigsaw lors de sa mise en place dans la classe.* » (Moskowitz et al., 1983, p. 694, traduit en Français). Pour eux, la structure de Jigsaw peut s'avérer trop difficile et contraignante pour être mise en place tel quel dans les classes. Or, si les modifications apportées à la méthode par les enseignant-es sont trop conséquentes, il peut être difficile d'établir quelle part de Jigsaw ou des modifications apportées ont des effets sur les variables dépendantes (Robert et al., *en préparation*).

Cette limite de la méthode Jigsaw a également été constatée par Slavin (1978 ; 1980) qui propose d'ajouter une récompense de groupe afin d'accroître l'interdépendance positive de récompense. Dans cette nouvelle version, appelée « Jigsaw II », le groupe puzzle est évalué collectivement, sur la base des scores individuels des membres : les individus ne sont plus seulement en charge de leur propre réussite (i.e., but individuel), mais doivent s'assurer que les autres réussissent également afin de maximiser la note du groupe. L'étude expérimentale de Susanne Ziegler (1981) est une des premières à s'intéresser aux effets de Jigsaw II. Mise en place dans six classes canadiennes d'élèves de CM2 et de 6^e ($N = 146$) pendant 8 semaines, cette étude s'intéressait à la fois à des éléments socioaffectifs (e.g., attitudes envers les pairs, nombre d'amis) et de performance académique via des évaluations de connaissances. Les résultats de cette expérimentation mettent en évidence des effets positifs sur toutes les variables : les élèves dans la condition expérimentale obtiennent de meilleurs scores aux évaluations de connaissance, rapportent un nombre d'amis plus élevé et des attitudes envers leurs pairs plus positives que les élèves de la condition contrôle.

4.4. Une validation empirique de la méthode ? 1990-2000

La période des années 1990 à 2000 est marquée par des recherches menées à la fois sur la première version de Jigsaw et sur Jigsaw II (Mattingly & VanSickle, 1991 ; Dori et al., 1995 ; Lazarowitz et al., 1994 ; Desforges et al., 1991 ; Walker & Crogan, 1998). Par exemple, dans

leur étude menée chez des étudiants de 1^{ère} et de terminale, Lazarowitz et collaborateurs (1994) rapportent des effets positifs de Jigsaw sur les performances académiques et des variables socioaffectives telles que la motivation, l'estime de soi, et le nombre d'amis comparativement à un groupe contrôle d'enseignement individualisé. Bien que les auteur·es rapportent des effets positifs sur ces variables, d'autres restaient in affectées par l'expérimentation : par exemple, les étudiant·es qui étaient dans la condition Jigsaw n'ont pas perçu que leur environnement de travail était plus coopératif et moins compétitif, contrairement aux attentes des chercheurs et aux effets rapportés par Aronson et collaborateurs (1978). Dans une autre étude, Dori et ses collègues (1995) trouvent des effets similaires sur les apprentissages, en suivant une procédure quasi identique à celle de Lazarowitz et collaborateurs (1994). L'étude de Desforges et collaborateurs (1991) apporte également des éléments empiriques en faveur de l'hypothèse de la réduction des préjugés, basée sur l'hypothèse du contact d'Allport (1954). Dans leur paradigme, ces auteur·es font coopérer des participant·es avec des compères se faisant passer pour des patients atteints de trouble mental, soit à l'aide de Jigsaw, d'un script coopératif structuré ou sans intervention. Suite à leur coopération sur une tâche, les participant·es étaient interrogé·es sur leurs attitudes envers les personnes atteintes de trouble mental, puis certain·es étaient recontacté·es un mois plus tard pour des mesures du même type. Les résultats de cette étude sont les suivants : les participant·es qui rapportaient initialement avoir des attitudes plutôt négatives envers les patient·es atteint·es de trouble mental rapportent des attitudes (via des mesures autorapportées) plus positives après avoir coopéré selon Jigsaw, en comparaison avec une absence d'intervention. Pour finir, Walker et Crogan (1998) s'intéressent aux relations interethniques entre les élèves d'origines ethniques différentes (i.e., Asiatiques, Aborigènes, Australiens, Européens) de CE2 à 6^e. Menée au sein de deux écoles ($N_{\text{école A}}=60$ et $N_{\text{école B}}=43$) en Australie, l'étude a duré respectivement quatre et trois semaines, en raison d'implémentation différente selon les établissements⁴. Pendant cette période, les élèves de la condition expérimentale étudiaient en suivant la procédure Jigsaw, dans des groupes d'environ 4 élèves. Ces dernier·es devaient donc travailler individuellement, puis en phase « expert » et présenter leurs documents en phase « puzzle » afin de reconstituer la leçon. Les auteur·es rapportent des effets positifs de Jigsaw sur l'évaluation des pairs ainsi que sur la réduction de préjugés. Cependant, ces effets ne sont rapportés qu'au travers d'interactions multiples, parfois entre les établissements et les conditions expérimentales et d'autres fois avec l'intégralité des

⁴ L'étude portait également sur les performances académiques, mais aucun·e des enseignantes n'a finalement communiqué les données aux chercheur·es responsables de l'étude.

variables expérimentales (e.g., ANOVA à quatre modalités : *condition*école*ethnie*genre*). Les résultats d'analyses effectuées avec une telle flexibilité dans les tests statistiques et sur un échantillon assez faible (i.e., inférieur à 100 participant·es) doivent donc être examinés avec prudence (voir Bratt, 2008 pour une critique de l'étude de Walker & Crogan, 1998).

Une seule étude publiée sur cette période s'intéresse à Jigsaw II, celle de Mattingly et VanSickle (1991). Dans cette étude, les auteurs argumentent que Jigsaw II peut permettre de dépasser les limites de la première version, considérée comme moins efficace que d'autres approches coopératives (voir chapitre 1). Selon ces auteurs, l'ajout d'une récompense de groupe basée sur la somme des contributions individuelles est censé améliorer l'interdépendance de récompenses, et donc de renforcer le but commun. D'après eux, « *cette structuration remplit la condition de responsabilité individuelle et de récompense de groupe, telle que formulée par Slavin* » (Mattingly & VanSickle, 1991, p. 6). Dans leur étude, 45 élèves de 3^e participent à un enseignement d'histoire-géographie pendant 9 semaines, soit de manière habituelle (i.e., enseignement en classe entière, devoirs et activités individuelles), soit avec la méthode Jigsaw II. Les résultats des évaluations de connaissance sur plus de 100 questions à choix multiples, obtenus au bout des 9 semaines d'expérimentation montrent que les élèves ayant étudié en suivant la méthode Jigsaw II obtiennent de meilleurs scores que les élèves en situation d'enseignement habituel (i.e., $d = .81$, soit un effet considéré comme « large » selon les classifications de Cohen, 1962 ; 1992). D'après eux, les résultats ne s'arrêtent pas seulement aux simples notes obtenues aux tests ; la qualité et la quantité des interactions entre les élèves de la classe coopérative ont été une autre conséquence bénéfique de Jigsaw II (Mattingly & VanSickle, 1991 dans Stahl & VanSickle, 1992). Selon eux, la note commune a stimulé les élèves dans leur participation, en rendant encore plus visible la contribution de chaque membre.

Cependant, ni ces recherches ni les méta-analyses et revues publiées jusqu'à cette période ne cherchent à comparer l'efficacité de la première comparativement à la deuxième version de Jigsaw, telle que proposée par Slavin (1980 ; 1985). Les quelques revues ayant été publiées sur cette période ne font pas cette distinction et regroupent aussi bien dans l'appellation « Jigsaw » des études menées selon la méthode d'Aronson et collaborateurs (1978) que celle de Slavin (voir Newmann & Thompson, 1987 ou Johnson et al., 2000). C'est également dans cette période qu'une troisième version de Jigsaw est conceptualisée par Stahl (1994), qui propose d'ajouter une phase de correction en classe entière, à l'aide de quiz adressés aux différents groupes puzzles. Cette procédure ne fera cependant pas l'objet de recherche

empirique sur cette période. La supériorité de Jigsaw II par rapport à Jigsaw n'est donc pas clairement établie (voir chapitre 4 pour un argument similaire).

4.5. L'exportation de la pédagogie : 2000-2010

La période post-2000 est marquée par un tournant dans la littérature Jigsaw : alors que les études provenant des États-Unis, berceau de la pédagogie, dominent les publications scientifiques jusqu'en 2000 avec plus de 70% (soit 11 études sur 15) des articles provenant des USA, ce pourcentage tombe à moins de 20% (soit 5 études sur 22) sur la période 2000 à 2010, puis à moins de 10% (soit 5 études sur 77) sur la période de 2010 à 2020 (Robert et al., *en préparation*). Cette évolution peut s'interpréter de différentes manières : on peut y voir l'exportation de la méthode Jigsaw dans d'autres pays, notamment ceux du Moyen-Orient qui ont multiplié les recherches durant ces dix dernières années, ainsi qu'une perte d'intérêt dans ce domaine de recherche aux États-Unis. On peut également constater que sur cette période, la plupart des études utilisent la première procédure de Jigsaw d'Aronson et ses collègues (1978 ; 1979) et beaucoup moins la procédure Jigsaw II telle que conceptualisée par Slavin (1980). La version IV de Jigsaw voit également le jour grâce à l'ajout de plusieurs quiz de connaissances entre les étapes principales de Jigsaw, mais seulement à travers un essai d'implémentation et non pas une recherche expérimentale (Holliday, 2002). Au Moyen-Orient, particulièrement en Turquie, les auteurs utilisent également une version spécifique de Jigsaw, appliquée exclusivement aux disciplines de la physique et de la chimie (i.e., Subject Jigsaw, voir Doymus, 2007 ; 2008 ; Karacop & Doymus, 2013). Ainsi, même si d'autres propositions d'améliorations de la pédagogie sont proposées, les versions I et II sont les plus couramment utilisées dans la littérature scientifique puisqu'elles recouvrent environ 70% des articles publiés (voir Robert et al., *en préparation* pour une revue de la littérature). Les principales révisions de la méthode sont synthétisées dans le Tableau 4 ci-dessous.

Version	Conception	Caractéristiques principales
Jigsaw I	Aronson et al. (1978)	Complémentarité des ressources et de rôles. Structure en trois étapes majeures : individuelle, expert, puzzle
Jigsaw II	Slavin (1980)	Ajout d'une récompense de groupe basée la somme des performances individuelles dans le groupe
Jigsaw III	Stahl (1994)	Correction globale (i.e., quiz de groupe) avant la phase puzzle
Jigsaw IV	Holliday (2002)	Quiz et tests de connaissances entre chaque étape
Subject Jigsaw	Doymus (2007 ² ; 2008a ^{1,3} ; 2008 ¹ ; 2010 ^{1,3})	Spécifique à la physique-chimie Préparation du matériel en dehors de la classe ¹ Chaque individu a l'occasion de voir tous les contenus dans une phase pré-expert ² Discussion en classe entière ³

Tableau 4: Les différentes versions de Jigsaw (Robert et al. en préparation)

Durant cette période, les recherches portent sur une grande variété de populations et de disciplines. Par exemple, Ghaith et Abd El-Malak (2004) testent l'efficacité de Jigsaw II chez des élèves Libanais de 4^e dans le cadre d'un cours d'anglais langue étrangère. Ils rapportent des effets positifs sur les apprentissages mesurés par des évaluations de connaissances sur les leçons traitées. Shaaban (2006) ainsi que Gömleksiz (2007) observent des résultats analogues avec la même méthode et sur des disciplines littéraires, auprès d'élèves de CM2 et d'Université respectivement. Cependant, les résultats observés sur ces études sont obtenus sur des échantillons généralement assez faibles ($N_{\text{Ghaith}} = 48$; $N_{\text{Shaaban}} = 44$; $N_{\text{Gömleksiz}} = 66$). Dans une autre étude, Artut et Karim (2007) utilisent la version II de Jigsaw dans le cadre de la formation de professeurs de mathématiques. Au bout d'une formation de 9 semaines effectuée soit en Jigsaw II, soit en condition d'enseignement en classe entière, ces professeurs étaient évalués sur leurs connaissances en mathématiques ; les résultats indiquent des gains importants en faveur du groupe expérimental, par rapport au groupe contrôle⁵ (i.e., effet « très large » $d = 1.06$).

⁵ L'estimation de taille d'effet étant potentiellement biaisée (voir Borenstein et al., 2011 et Funder & Ozer., 2019 pour une discussion), un tel indicateur doit être interprété avec précaution. Le chapitre 3 développera cette limite de façon plus détaillée.

Des effets plus mitigés sont également rapportés : Souvignier et Kronenberg (2007) n'observent aucun gain, voire des performances moins bonnes en Jigsaw comparativement à un enseignement habituel, sur quatre séquences de contenus scientifiques (i.e., astronomie et mathématiques). Même dans leur troisième condition expérimentale, une version « améliorée » de Jigsaw avec des questions pour aider les élèves lors des phases « expert » et « puzzle » (i.e., c'est-à-dire proche d'un Jigsaw IV, Holliday, 2002), les performances étaient moins bonnes au sein du groupe expérimental qu'au sein du groupe contrôle. Bratt (2008), à travers deux études menées chez des élèves norvégiens de 8 à 15 ans, n'observe pas d'effets bénéfiques sur les relations interpersonnelles (i.e., attitudes positives et expression de préjugés). Pour lui, l'absence de résultats malgré une implémentation de la technique fidèle aux considérations d'Aronson (1978 ; 1997) à travers deux études ($N_{\text{étude 1}} = 68$; $N_{\text{étude 2}} = 129$) remet en question l'efficacité de Jigsaw dans la réduction des préjugés interethniques.

D'autres auteurs, comme Martin Hänze et Roland Berger soulèvent également des doutes sur l'efficacité globale de la méthode dans l'amélioration des apprentissages. Leurs études menées auprès d'élèves de Terminale (Hänze et Berger, 2007 ; Berger & Hänze, 2009) dans des cours de physique-chimie mettent en évidence que les effets de Jigsaw ne s'observent que dans certaines conditions, à travers des médiations par le sentiment d'auto-efficacité. Selon ces auteurs, la méthode Jigsaw devrait améliorer les trois « besoins fondamentaux » que sont le sentiment de proximité sociale, d'autonomie et de compétence, basés sur la théorie de l'autodétermination (Deci & Ryan, 1985 cités par Hänze et Berger, 2007). D'après ces auteurs, Jigsaw pourrait répondre au besoin de compétence et d'autonomie grâce à l'interdépendance des ressources (i.e., compétence dans la phase « expert » et autonomie dans la phase individuelle), et satisfaire le sentiment de proximité sociale grâce à la coopération avec autrui. Dans leurs expérimentations, les étudiant·es dans la condition de Jigsaw se sentaient plus compétents, autonomes et socialement proches de leurs camarades : les gains rapportés sur les apprentissages n'étaient observés que pour les étudiant·es qui rapportaient des améliorations dans ces besoins fondamentaux. Pour les autres, les effets de Jigsaw étaient nuls, voire négatifs. Enfin, l'étude de Moreno (2009) n'apporte pas de soutien empirique à l'hypothèse selon laquelle Jigsaw améliore les apprentissages des étudiant·es en études supérieures. Au contraire, les participant·es de cette étude (étudiant·es dans un cours de botanique) exposé·es à la méthode Jigsaw rapportent une charge cognitive plus élevée et une performance moindre que dans la condition contrôle. Cependant, l'auteure note une limite importante de sa recherche, puisque Jigsaw n'était utilisé que sur une seule séance : « les performances moins élevées pour le groupe

Jigsaw pourraient être interprétées comme les meilleures tentatives des étudiant·es d'interagir avec une pédagogie non familière » (Moreno, 2009, p. 441).

En Turquie, Doymus (2007 ; 2008a ; 2008b) utilise des versions modifiées de Jigsaw (i.e., « Subject Jigsaw », voir Robert et al., en préparation) spécialement dédiées à l'enseignement de la physique-chimie, sur des séquences couvrant plusieurs semaines de cours. Ces versions autorisent des modifications importantes par rapport aux premières expérimentations d'Aronson et collaborateurs (1978) : bien que la structure en différentes étapes soit conservée, en suivant le principe d'interdépendance de ressources et de rôles les évaluations sont toujours effectuées de manière individuelle sur des tests de connaissances relatives aux contenus pédagogiques vus en cours (i.e., questions choix multiples, résolution de problème). Les modifications concernent principalement 1) la substitution de la phase « individuelle » en classe à une phase de préparation du matériel pédagogique chez soi, en amont de la classe et 2) la possibilité d'une restitution et de discussion en classe entière, après la phase « puzzle ». Par exemple, dans son étude menée auprès d'étudiant·es sur un cours de chimie sur les états de la matière, Doymus (2007) n'utilise pas l'étape 1 de Jigsaw qui consiste à créer des groupes « puzzle » où chacun dispose d'une pièce d'information différente. À la place, les étudiant·es sont tout de suite répartis dans trois groupes qui correspondent normalement à la phase « expert » (voir Doymus, 2008a ; 2008b ; 2010, pour une procédure similaire). Ces trois groupes renvoient à un état différent de la matière (i.e., groupe S : solide ; groupe L ; liquide ; groupe G : gazeux). La phase suivante, ou phase « puzzle », est également modifiée puisque quatre groupes intermédiaires sont formés, et correspondent à des regroupements par paires entre étudiant·es : SL pour la paire « solide-liquide », LG pour « liquide gazeux », SG pour « solide gazeux » et TP pour « triple point », constitué par les étudiant·es restant·es. De plus, dans cette version, la restitution de connaissances ne s'effectue pas au niveau du groupe, mais devant la classe entière, et les groupes interagissent entre eux après avoir présenté leurs parties. Doymus (2007) observe que, suite à cette nouvelle procédure, les évaluations individuelles sur deux tests de chimie sont meilleures pour les étudiant·es en Jigsaw comparativement à une approche habituelle, frontale et en classe entière. Dans d'autres recherches, Doymus (2008a ; 2008b) utilise une autre procédure légèrement différente : cette fois-ci les étudiant·es doivent effectuer le travail « individuel » (i.e., l'étape 1 selon Aronson et al., 1978) chez eux, à la maison, en amont de la classe. Dans cette version, les étudiant·es commencent le cours directement en phase « expert », que Doymus (2008a ; 2008b) nomme « home group ». Ensuite, les groupes « puzzle » sont constitués en regroupant les étudiant·es ayant travaillé sur une sous-partie

spécifique, comme dans la version originelle de Jigsaw. Dans les expérimentations de Doymus (2008a ; 2008b) et aussi de Doymus et collaborateurs (2010), la phase de reconstitution, qui correspond à la phase « puzzle » est différente, puisqu'elle s'effectue en classe entière. Les variations subtiles de procédure dans les études de ces auteurs ajoutent une hétérogénéité non négligeable au terme « Subject Jigsaw ». Bien que la quasi-intégralité de leurs recherches rapporte des effets positifs et larges (i.e., $d > .80$, Cohen, 1962, 1994), on ne peut pas raisonnablement prédire quelle part des modifications apportées à la procédure Jigsaw permet d'expliquer ces gains. On peut également noter que ces modifications ne concernent pas l'attribution de récompenses de groupe, qui constituent le « problème » de Jigsaw selon Slavin (1980) ou Moskowitz et collaborateurs (1983 ; 1985). Dans Subject Jigsaw, les modifications majeures proviennent de la préparation du travail chez soi et de la présentation du contenu à la classe entière. On peut supposer que cette dernière modification puisse rajouter une condition supplémentaire à la visibilité de la contribution de chacun·e (i.e., responsabilité individuelle), mais elle ne constitue pas un changement dans l'attribution de récompense ou de but commun entre les étudiant·es. Pour conclure à la supériorité de Subject Jigsaw sur les versions antérieures de la méthode, il faudrait tester empiriquement, sur un contenu pédagogique et une évaluation identique, l'effet respectif de chacune de ces méthodes.

4.6. La littérature récente : 2010-2020

Sur la période la plus récente, l'augmentation du nombre d'articles sur Jigsaw s'accompagne d'une réduction de la taille des échantillons des études (voir Figure 7). Pourtant, l'écrasante majorité des articles réalisés sur la période 2010-2020 rapporte des effets positifs de la méthode sur les performances académiques, que ce soit les versions Jigsaw I, II ou même Subject Jigsaw (voir Tableau 5 pour une synthèse) ce qui contraste fortement avec le faible taux de réussite rapporté dans les revues publiées de 1980 à 2000 (e.g., Sharan, 1980 ; Newmann & Thompson, 1987 ; Slavin, 1983 ; Johnson et al., 2000). Dans leur revue critique sur la littérature sur Jigsaw, Vives et collaborateurs (*en préparation*) montrent que les articles récemment publiés sur Jigsaw tendent à rapporter une très large quantité d'effets positifs, parfois larges, voire très larges (i.e., $d > 1.00$, soit plus d'un écart-type standardisé sur la note, Cohen, 1962), alors que les premiers articles rapportaient des effets beaucoup plus faibles (i.e., d [.13 ; .29], Johnson et al., 2000). Robert et collaborateurs (*en préparation*) observent des corrélations entre l'année de publication des recherches concernant la méthode Jigsaw et la taille d'échantillon (r

= -.31), la qualité d'implémentation de la méthode⁶ ($r = .29$) et l'estimation des tailles d'effets observées dans les expérimentations ($r = .30$). Pour ces auteur·es, la quantité des recherches effectuées ces dernières années sur la méthode Jigsaw semble s'être faite au détriment de leur qualité méthodologique. Plus particulièrement, iels notent que ces articles proviennent majoritairement du Moyen-Orient, avec une recrudescence d'articles réalisés en Turquie sur la période des années 2000 à 2020 (Doymus et al., 2010 ; Koc et al., 2010 ; Sahin, 2011 ; Tarhan & Sesen, 2012 ; Tarhan et al., 2013 ; Yoruk, 2016 ; Basyah et al., 2018).

Par exemple, Koc et collaborateurs (2010) comparent les effets de Subject Jigsaw avec ceux de la méthode Group Investigation (GI) et d'un groupe contrôle sans intervention. Suivant une procédure proche de celle de Doymus et collaborateurs (2010), les auteurs rapportent des effets positifs et d'amplitude similaire pour Subject Jigsaw et GI, comparativement au groupe contrôle. D'autres chercheur·es utilisent la procédure de Jigsaw et observent des effets positifs sur les apprentissages tels que mesurés par des évaluations de connaissances, en chimie (Tarhan & Sesen, 2012 ; Yoruk, 2016), en physique (Tarhan et al., 2013) et en économie (Basyah et al., 2018). Au niveau secondaire, Sahin (2011) utilise la procédure Jigsaw III et observe également des effets positifs chez des élèves de sixième dans l'apprentissage de langue turque. Dans une autre étude, Law (2011) s'intéresse aux effets de Jigsaw sur la réussite et la motivation d'élèves Hong-Kongais dans un cours de lecture. Il met en place trois groupes expérimentaux : un groupe Jigsaw, un groupe d'apprentissage par le théâtre et un groupe contrôle, en classe entière sans intervention. À la suite des cinq sessions, il observe que les élèves tendent à être plus performants sur une évaluation de connaissance, et davantage centrés sur la « maîtrise » de leur contenu, après avoir étudié avec Jigsaw. En d'autres termes, ces derniers obtiennent de meilleurs résultats que les autres groupes, et rapportent plus d'adhésion aux « buts de maîtrise » (e.g., « *Je résous les exercices en lien avec la leçon parce que ça m'intéresse* ») que les élèves issus du groupe contrôle. Dans la littérature scientifique, les buts de maîtrise sont le plus souvent associés à un plus grand désir d'apprendre et de développer des nouvelles compétences, ainsi qu'un à une coopération plus importante (Darnon et al., 2006 ; Poortvliet & Darnon, 2014). Law (2011) rapporte toutefois que les corrélations entre les buts de maîtrise et la réussite à l'évaluation de connaissances sont assez faibles ($r = .13$).

⁶ Score continu élaboré sur la base de la nature du groupe contrôle et du design expérimental, la qualité de l'implémentation et des mesures, la durée de l'expérimentation et la taille de l'échantillon (Robert et al., *en préparation*). Les auteur·es rapportent que la qualité de l'implémentation de Jigsaw est « forte ou excellente » dans 21% des cas (11 articles sur 53), « intermédiaire » dans 55% des cas (29 sur 53) et « médiocre à faible » dans 25% de cas (13 articles sur 53).

D'autres chercheur·es se sont intéressées aux effets de Jigsaw sur d'autres variables que celles des performances académiques. Darnon et collaborateurs (2012) ont étudié l'effet de Jigsaw sur le sentiment d'auto-efficacité en mathématiques et en français des élèves de lycée professionnel. Les auteur·es rapportent qu'en seulement quatre sessions de 2 heures, le sentiment d'auto-efficacité des élèves ayant travaillé en coopération avec Jigsaw s'était amélioré, par rapport à un groupe contrôle de travail individuel. Crone et Portillo (2013) rapportent des résultats similaires, sur une population d'étudiants en psychologie à l'université : les auteur·es mesuraient la confiance des étudiant·es dans leur capacité à communiquer oralement des explications (e.g., « *Je me sens capable d'expliquer un cours de psychologie que j'ai suivi* »), ainsi que leur sentiment d'auto-efficacité académique (e.g., « *J'ai confiance en mes capacités d'étudiant·e* ») avant et après avoir suivi un cours avec Jigsaw ou sans intervention. Les résultats de cette étude indiquent des bénéfices de Jigsaw sur la confiance et le sentiment d'auto-efficacité des étudiant·es, mais pas de bénéfices sur l'examen terminal du cours.

Pour revenir aux études portant sur des contenus scolaires, Roseth et collaborateurs (2018), offre une réinterprétation particulièrement intéressante de Jigsaw au regard de théories sociocognitives des apprentissages (i.e., perspective cognitive, chapitre 1). Plutôt que de considérer la méthode Jigsaw comme étant purement coopérative, les auteur·es argumentent que Jigsaw génère différentes dynamiques en fonction des étapes de la pédagogie :

- Même si la phase individuelle a pour vocation de créer le groupe « puzzle », la nature individuelle de la tâche (i.e., prendre connaissance de sa sous-partie) peut être insuffisante pour créer un sentiment d'unité du groupe (Johnson & Johnson, 2005).
- La phase « expert » serait reliée à des dynamiques individualistes puisque les membres sont indépendants d'un point de vue des ressources (Slavin, 2016).
- Pour finir, la phase « puzzle » impliquerait non seulement des relations d'interdépendance positive - grâce à la complémentarité des ressources - mais également des relations compétitives à cause de l'indépendance des récompenses. Sans récompense de groupe, il est tout à fait possible que les individus récupèrent les informations dont disposent les autres membres, sans partager les leurs (Buchs et al., 2004 ; Buchs & Butera, 2019).

Auteur·es	Année	Pays	Type de Jigsaw	Population	N _{total}	Discipline	Durée	Effets positifs		
								Groupe contrôle	Performances académiques	Socio-affectives
Doymus et al.	2010	Turquie	Subject Jigsaw	Enseignement supérieur	122	Chimie	5 semaines	Assisté par ordinateur ; Travail individuel	+	/
Koc et al.	2010	Turquie	Subject Jigsaw	Enseignement supérieur	106	Chimie	4 semaines (4h/semaine)	Enquête de groupe ; Travail individuel	+	/
Sahin	2011	Turquie	Jigsaw III	Secondaire (6 ^e)	71	Expression écrite	6 semaines (5h/semaine)	Travail individuel	+	Attitudes envers Jigsaw
Law	2011	Chine	Jigsaw	Primaire (CM2)	279	Lecture	5 x 1 h	Groupe "théâtre" ; Travail individuel	+	Buts d'accomplissement
Tarhan & Sesen	2012	Turquie	Jigsaw	Enseignement supérieur	38	Chimie	1 Session (1h15)	Travail individuel	+	Attitudes envers Jigsaw
Darnon et al.	2012	France	Jigsaw	Lycéen.n.es professionnels	33	Mathématiques et Français (Expression orale)	4 sessions (2h/ session)	Travail individuel	Non mesuré	Sentiment d'auto-efficacité (mathématiques et français)
Tarhan, et al.	2013	Turquie	Jigsaw	Secondaire (6 ^e)	61	Physique	2 semaines (4h/semaine)	Travail individuel	+	Attitudes envers Jigsaw
Crone & Portillo	2013	USA	Jigsaw	Enseignement supérieur	70	Psychologie cognitive	1h10	Jigsaw "réduit" ; Travail individuel	+(tendanciel)	Sentiment d'auto-efficacité académique
Yoruk	2016	Turquie	Jigsaw	Enseignement supérieur	63	Chimie	6 semaines (1x/semaine)	Travail individuel	+	/
Theobald et al.	2017	USA	Jigsaw	Enseignement supérieur	770	Biologie	4 semaines (3h/ semaine)	Coopération non structurée	+	± Perceptions du travail de groupe
Basyah, et al.	2018	Indonésie	Jigsaw	Tertiaire (Premières)	60	Économie	14 semaines	Travail individuel	+	/
Roseth et al.	2018	USA	Jigsaw (modifié)	Enseignement supérieur	258	Anatomie	1 Semestre (5h/ semaine)	Travail individuel	+	± Régulation conflit socio-cognitif, motivation

Tableau 5: Caractéristiques et résultats principaux des expérimentations menées sur la période 2010-2020 (sources : Robert et al., en préparation ; Vives et al., en préparation).

Leur étude porte sur 258 étudiants sur un cours universitaire d'anatomie humaine d'une durée d'un semestre (i.e., 75 minutes quatre fois par semaine). Bien qu'il semble y avoir de faibles effets de Jigsaw sur les performances académiques (voir chapitre 4), les auteurs montrent surtout la dynamique des perceptions des étudiant·es au sein de Jigsaw : iels percevaient davantage de compétition et moins de coopération entre eux au début de l'expérimentation - c'est-à-dire au début du semestre - par rapport à un groupe d'enseignement habituel, puis cette tendance s'inversait à la fin de l'expérimentation. De manière générale, Roseth et collaborateurs (2018) amènent à repenser Jigsaw comme une structure qui génère divers types de relations entre les individus. Cette reconsidération permettrait de mieux comprendre les effets mitigés observés dans les premières tentatives de réplification des effets de Jigsaw. En effet, Moskowitz et ses collègues (1983 ;1985) soulignaient déjà le problème lié à l'absence de récompense de groupe dans Jigsaw, problème supposément réglé par la création de Jigsaw II (Slavin, 1980). Dans la recherche de Roseth et collaborateurs (2018), les auteurs utilisent une version de Jigsaw plus proche de celle de Slavin (1980), puisqu'iels ajoutent une récompense de groupe basée sur la performance individuelle de chaque membre.

Dans une autre étude, Voyles et ses collègues (2015) apportent d'autres modifications à Jigsaw afin d'accroître la responsabilité individuelle, et donc réduire la paresse sociale, des individus. Pour ce faire, les chercheur·es ont demandé aux individus de prendre des notes dans les phases « expert », pour les partager aux autres membres du groupe et de les transmettre à l'enseignant·e pour les évaluer. Ensuite, chacun devait présenter un produit commun à l'intégralité de classe, afin de mettre en place un but commun plus visible que celui proposé dans Jigsaw II. Pour les auteur·es, ces modifications permettent d'accroître le sentiment de responsabilité individuelle et faciliter l'évaluation et le feedback au sein du groupe : puisque chaque individu doit soumettre un produit (i.e., une prise de note personnelle) à l'enseignant·e, la contribution individuelle est rendue explicite. L'enseignant·e peut aussi évaluer cette prise de note, et ainsi rendre un feedback aux individus dans leurs groupes.

Bien que de tels ajouts semblent être tout à fait en phase avec l'amélioration de la structure de Jigsaw, les propositions de Voyles et collaborateurs (2015) n'ont pas l'objet d'une validation empirique par l'expérimentation. Leur recherche consiste davantage en une description de l'implémentation de la version modifiée de Jigsaw, mais cette dernière n'est pas évaluée empiriquement. D'autres méthodes comme l'évaluation par les pairs (Kao, 2012) ou la distribution de rôles (Topping et al., 2017) pendant l'apprentissage coopératif ont également été proposées pour augmenter l'interdépendance positive entre les individus. Dans cette même

période, nous pouvons aussi noter l'apparition de plusieurs recherches combinant Jigsaw et les technologies de communication comme l'usage de blog (Huang et al., 2011), de tablettes (Huang et al., 2014) ou d'applications web (Parsazadeh et al., 2018) qui contribuent à enrichir la diversité des procédures. Nous ne traiterons cependant pas de manière détaillée ces études, puisqu'elles dévient assez fortement du cadre scolaire « typique » des études présentées jusqu'à présent. De plus, ces études s'inscrivent dans des perspectives centrées sur l'utilisation et l'acceptabilité des outils numériques, plutôt que dans l'évaluation de l'efficacité de Jigsaw *per se*. De la même manière, nous mentionnerons pour son originalité, mais ne développerons pas l'étude de Nebel et collaborateurs (2017), qui utilise le jeu vidéo *Minecraft* et met en évidence des meilleurs comportements de coopération (i.e., répartition des tâches, coordination des efforts) suite à l'entraînement des participants selon la méthode Jigsaw.

Pour conclure, Vives et collaborateurs (*en préparation*) constatent que de ces multiples modifications et ajustements de la version de base parsèment la littérature scientifique, rendant parfois les comparaisons entre études difficiles compte tenu de l'hétérogénéité des mises en pratique. Il semble difficile d'établir l'efficacité de Jigsaw de manière générale si des modifications sont apportées d'une expérimentation à une autre, d'autant plus que certains articles ne clarifient pas spécifiquement les procédures suivies lors de leurs recherches (Robert et al., *en préparation* ; Vives et al., *en préparation*).

5. Une littérature inconsistante

Pour résumer, il subsiste encore aujourd'hui des enjeux importants pour la recherche concernant la méthode Jigsaw. La littérature récente, marquée par de forts effets principaux sur les apprentissages observés sur toutes les versions de Jigsaw semble contraster avec les travaux pionniers, et surtout avec les conclusions de Newmann et Thompson (1987) ou Moskowitz et collaborateurs (1983 ; 1985) qui affirmaient que Jigsaw n'avait pas d'effet notable sur l'apprentissage des élèves. Un des objectifs de ce travail de thèse sera d'apporter des éléments de réponses pour mieux considérer cette disparité. L'analyse de la littérature a révélé des tendances différentes entre deux périodes de recherche, avant et après les années 2000. La recrudescence d'articles récents publiés sur des échantillons faibles, mais rapportant des effets positifs et larges doit être examinée avec prudence, compte tenu des limites soulevées par Slavin (1980) ou Johnson et collaborateurs (2000) sur l'efficacité de Jigsaw. Dans le prochain chapitre, nous développerons l'argument que les travaux les plus récents semblent souffrir d'un certain nombre de limites méthodologiques, particulièrement liées à la puissance et aux procédures statistiques déployées pour évaluer Jigsaw.

Résumé du chapitre 2

- Les effets positifs de Jigsaw sont marqués par une forte hétérogénéité (i.e., parfois négatifs, nuls, ou positifs de faible ou de très large amplitude). L'efficacité de la méthode est loin de faire consensus : le nombre d'études rigoureuses semble relativement bas et les preuves de l'efficacité de Jigsaw manquent.
-
- La littérature Jigsaw est marquée par une rupture entre deux périodes de publication : alors que la période allant de 1970 à 2000 est plutôt marquée par un faible nombre d'études menées auprès d'échantillons assez importants, le nombre d'études a explosé depuis 2000, au détriment des tailles d'échantillon.
-
- Plusieurs versions de la méthode Jigsaw ont été développées, reposant toutes sur le principe d'interdépendance de ressources. Cependant, la plupart ne sont pas validées empiriquement : l'absence de comparaison entre les différentes méthodes Jigsaw ne permet pas de conclure à la supériorité d'une version sur une autre.

Chapitre 3. Les limites de la littérature Jigsaw : Considérations méthodologiques et opérationnalisation de l'hypothèse de recherche

1. Objectif du chapitre

Dans ce chapitre, nous nous intéresserons aux limites méthodologiques observées dans la littérature traitant de la méthode Jigsaw. Nous développerons l'argument que cette littérature souffre d'un certain nombre de défauts : sous-puissance de test, manque de transparence dans les procédures expérimentales et tests statistiques inadaptés. Comme dans le chapitre précédent, nous ne traiterons que des articles expérimentaux (i.e., comparaison entre une version de Jigsaw et un groupe contrôle) publiés dans des revues à comité de lecture. Nous présenterons également l'hypothèse de recherche principale du manuscrit de thèse ainsi que les outils déployés pour la tester, notamment la spécification d'une plus petite taille d'effet d'intérêt et des tests d'équivalence (Lakens, 2017 ; Lakens et al., 2018) et la procédure de méta-analyse interne (Goh et al., 2016).

2. Une puissance statistique souvent faible

2.1. La question de la puissance statistique

L'une des limites importantes identifiées dans la littérature présentée dans le chapitre 2 concerne la taille des échantillons des études sur Jigsaw. Dans le chapitre 2, nous avons observé que les tailles d'échantillon moyennes étaient relativement importantes (i.e., entre 200 et 300 participant·es par condition, voir Figure 7, chapitre 2) sur les périodes de 1970 à 1990, puis enregistraient une baisse importante jusqu'à aujourd'hui. Rappelons que lors d'une expérimentation cherchant à tester une hypothèse H_1 comme par exemple, « *l'apprentissage coopératif aura un effet positif sur les performances académiques* » contre une hypothèse nulle H_0 : « *l'apprentissage coopératif n'a pas d'effet positif sur les performances académiques* », quatre scénarii sont possibles (voir Tableau 6). Dans une large majorité des cas, les scientifiques tentent de détecter de *vrais* effets, en rejetant l'hypothèse nulle H_0 au profit de l'hypothèse alternative H_1 , c'est-à-dire à se placer dans le cas du scénario du « vrai positif ». Dans ce cas de figure, les chercheur·es doivent principalement s'intéresser au facteur $(1 - \beta)$ qui renvoie à la probabilité de détecter un effet, lorsque ce dernier existe. Ce facteur est nommé « puissance de test » ou « puissance statistique » (Cohen, 1962 ; 1992) et dépend principalement de deux éléments : la taille de l'effet à détecter ainsi que la taille de l'échantillon sur lequel l'expérimentation a lieu.

Des effets de petite taille (i.e., proches ou inférieur à $\delta = 0.20$) nécessiteront des échantillons importants pour disposer d'une puissance de test suffisante (Cohen, 1992). Par convention, certains auteur·es évaluent une puissance de test comme étant satisfaisante à au moins 80% de chance de détecter un effet existant, mais en pratique, la puissance statistique en psychologie est estimée comme étant plus proche des 50% (Fanelli, 2010 ; Funder & Ozer, 2019 ; Tressoldi, 2012). On peut aussi noter que dans la littérature récente sur Jigsaw, aucun article publié à ce jour n'a effectué d'analyse de puissance de test pour déterminer la taille de l'échantillon. Faire une prédiction sur une taille d'effet probable permet de mettre en place une expérimentation qui aura plus de chances d'être « informative », sur la base d'un calcul de puissance de test *a priori* (Lakens & Evers, 2014). Dit autrement, spécifier l'amplitude d'un effet probable, c'est-à-dire un effet sur lequel on dispose à la fois de raisons théoriques et méthodologiques pour mener une expérimentation, permet d'avoir une idée sur le nombre de participant·es requis pour observer des différences entre groupes expérimentaux dans une étude, et donc de réduire le risque d'erreur de Type II.

	H ₀ vraie	H ₁ vraie
Résultat significatif	Faux positif (α)	Vrai positif (1 - β)
Résultat non significatif	Vrai négatif (1 - α)	Faux négatif (β)

Tableau 6: Scénarii possibles d'un test d'hypothèse nulle (Lakens: *Improving your Statistical Questions*; [coursera.org](https://www.coursera.org))

Revenons au Tableau 6 et évoquons rapidement les autres scénarii : le « faux négatif » dépend du facteur β et renvoie à la situation dans laquelle l'effet existe, mais nous ne disposons pas de suffisamment de puissance de test. Il est directement lié à la décision prise par les chercheur·es par rapport à la puissance de test choisie : si ces dernier·es décident que 80% de chance de détecter l'effet (1 - β) est une puissance satisfaisante, alors le risque de « faux négatif » est de 20%. On appelle ce risque « erreur de Type II ». Ce risque peut être réduit à 10% si les chercheur·es décident d'utiliser une puissance de 90%, à 5% si la puissance est de 95%, etc. Les autres scénarii concernent un autre type d'erreur, « l'erreur de Type I », situation dans laquelle un test statistique (e.g., une analyse de variance, un *t*-test indépendant ou une régression linéaire) rapporte une *p*-value inférieur à ce seuil, alors que l'effet « n'existe pas ». Il s'agit tout simplement d'une fluctuation due au hasard, d'un bruit statistique que le test aurait détecté (Goodman, 2008 ; Halsey et al., 2015). L'utilisation de la simple *p*-value comme critère

de rejet de l'hypothèse nulle H_0 fait l'objet de critiques depuis sa création parce qu'elle ne permet pas de qualifier la fluctuation des valeurs en fonction de la puissance statistique (Pek & Flora, 2017). Nous reviendrons sur ce point en chapitre dans le chapitre 4 qui présentera la procédure de méta-analyse.

2.2. La taille d'effet probable de Jigsaw

Comme évoqué précédemment, nous argumentons que la littérature Jigsaw présente probablement un risque élevé d'erreur de Type II, c'est-à-dire de sous-puissance statistique, notamment sur les études publiées après 2000. En effet, une grande majorité des études récentes ne dispose pas d'une puissance de test satisfaisante pour détecter des effets « modérés » et encore moins des effets de « petite taille », c'est-à-dire des effets proches de $\delta = 0.40$ ou $\delta = 0.20$ (Robert et al., *en préparation*). De manière générale, les méta-analyses publiées jusqu'à ce jour nous informent de la taille d'effet probable de Jigsaw et de l'apprentissage coopératif de manière générale. Comme discuté dans le chapitre 1, ces données méta-analytiques situent l'effet moyen de Jigsaw autour de $\delta \pm 0.20$ (Johnson et al., 2000) et l'effet de l'apprentissage coopératif sur la performance académique, en comparaison à un apprentissage individuel, aux alentours de $\delta = 0.40$ (Hattie, 2009). Dans le domaine des interventions académiques visant à améliorer l'apprentissage des étudiant·es et élèves, Hattie (2009) préconise l'utilisation d'un seuil de « signification pratique ». Concrètement, il s'agit de fixer un seuil à partir duquel une intervention sera jugée comme étant utile ou « d'intérêt pratique » pour les enseignant·es et leurs étudiant·es. Pour Hattie (2009), ce seuil doit être fixé à $\delta = 0.40$, soit une progression de presque la moitié d'un écart-type sur une mesure de performance scolaire puisqu'il renvoie à la taille d'effet médiane des interventions académiques sur la réussite académique rapportée dans son ouvrage. D'autres chercheur·es situent ce « seuil de signification pratique » vers une taille d'effet de $\delta = 0.33$ (e.g., Gall et al., 1996 cité dans Kyndt et al., 2013), ou dans un intervalle de prédiction compris entre $\delta = 0.25$ et $\delta = 0.50$ (Springer et al., 1999). Il convient de noter que jusqu'à présent, dans la littérature sur la méthode Jigsaw, de tels éléments ne sont pas pris en compte : certain·es auteur·es justifient la taille de leur échantillon sur des heuristiques (i.e., « *au moins 30 participant·es par condition* », voir Tarhan et collaborateurs, 2013), mais dans la l'écrasante majorité des cas, aucune justification n'est apportée. Un des arguments que nous souhaitons développer dans ce chapitre est que la puissance de test disponible pour avoir une chance raisonnable de détecter un effet de Jigsaw situerait situer aux alentours de $\delta \pm 0.20$ pour des raisons « historiques » (Johnson et al., 2000) ou de $\delta \pm 0.40$ pour des raisons « pratiques » (Hattie, 2009) est insuffisante. De manière plus précise, nous argumentons que cette sous-

puissance de test concerne particulièrement les articles scientifiques récents et réalisés au Moyen-Orient, notamment en Turquie.

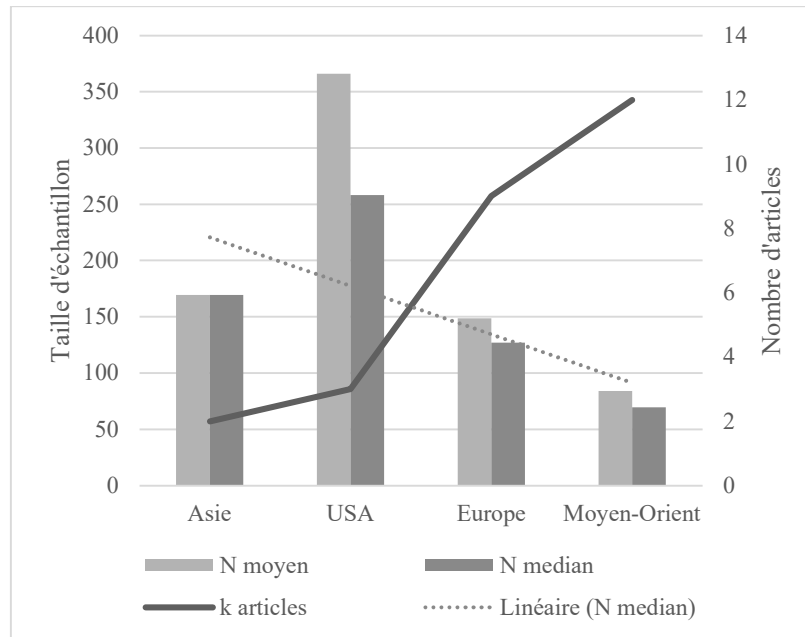


Figure 8: Taille d'échantillon et nombre d'articles empiriques publiés après 2000 (Robert al., en préparation)

La Figure 8 décrit une tendance forte dans la littérature Jigsaw, à savoir une relation négative entre le nombre d'études réalisées et la taille moyenne des échantillons de ces études. Même s'il ne s'agit que d'éléments descriptifs, il semble que les articles provenant du Moyen-Orient (i.e., Turquie et Liban) soient particulièrement concernés par la sous-puissance statistique. Dans ces articles (Doymus, 2007 ; 2008a ; 2008b ; Sahin, 2010 ; Yoruk 2016 ; Artut & Tarim, 2007 ; Gömleksi'z, 2007 ; Karacop et al., 2010 ; Tarhan et al., 2013, Koc et al., 2010 ; Tarhan & Sesen, 2012), la taille moyenne de l'échantillon total est de 84 participant-es par condition ($N_{\text{médian}} = 70$) soit quasiment deux fois moins que la moyenne globale de 143 participant-es par condition ($N_{\text{médian}} = 81$) sur cette même période. Or, sur un tel échantillon, la puissance statistique pour détecter un effet de faible ou de moyenne amplitude n'est pas suffisante. Pour illustrer ce manque de puissance, nous avons effectué plusieurs simulations, à l'aide du logiciel JAMOVI (voir Figure 9 ci-dessous).

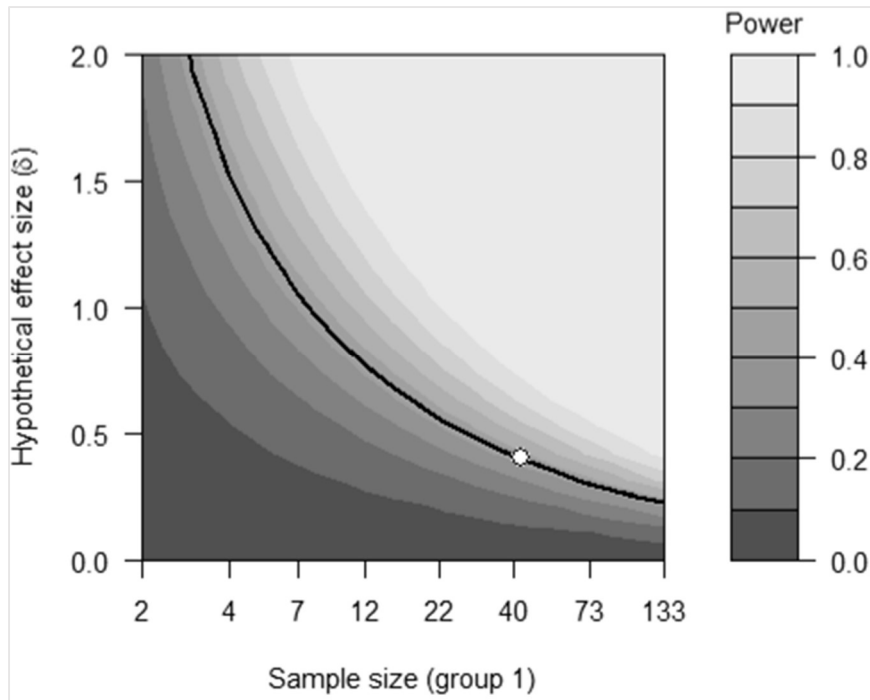


Figure 9: Courbe de puissance de test simulée pour une taille d'effet de $\delta = 0.40$, au seuil $\alpha = .05$ (JAMOV 1.1.8.0)

Ce premier graphique représente la relation entre la taille d'échantillon (ici fixée à 42 participant·es par groupe, sur la base de la taille moyenne des échantillons issus du Moyen-Orient), la puissance de test (i.e., la probabilité de détecter un effet, allant de 0 = nulle à 1 = maximale) et la taille supposée de l'effet à détecter (i.e., fixée à $\delta \pm 0.40$). Dans notre cas, la puissance de test est inférieure à 50%, et encore moins pour un effet de faible amplitude (~30%). Dit autrement, chaque étude réalisée avec un échantillon d'environ 80 participant·es a autant de chance de détecter un tel effet que d'obtenir *pile* en jetant une pièce de monnaie équilibrée (Ioannidis, 2005). Sur le long terme, mettre en place des expérimentations systématiquement en sous-puissance doit logiquement conduire à des résultats non statistiquement significatifs, en raison de la fluctuation des valeurs de p (Lakens & Etz, 2017), et ce même si l'effet existe. Voyons maintenant une conséquence logique d'expérimenter sur de tels échantillons en termes d'effets observés.

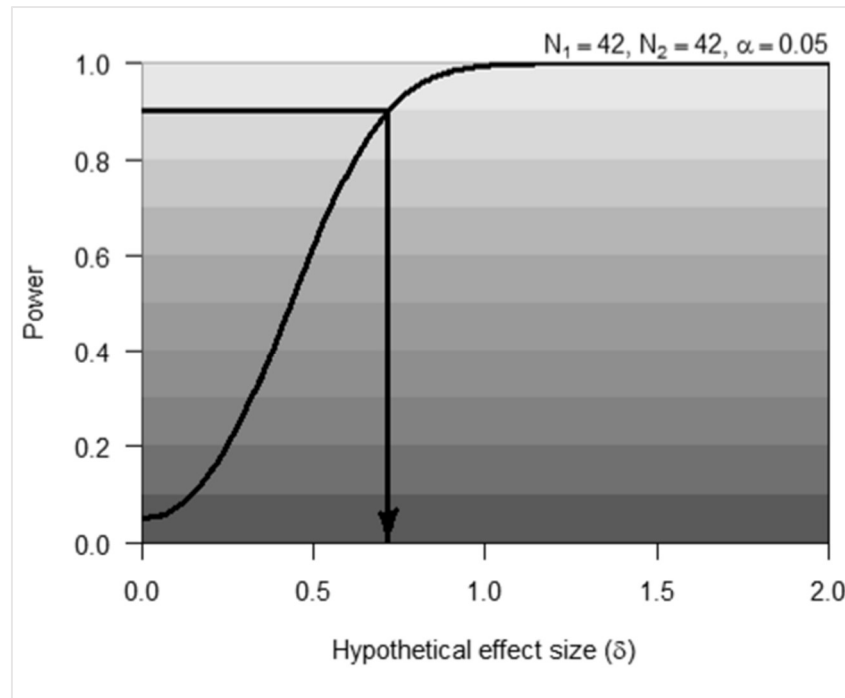


Figure 10: Analyse de sensibilité (puissance de test a posteriori) sur la base du nombre moyen de participants par condition dans la littérature Jigsaw (JAMOVI 1.1.8.0)

La Figure 10 ci-dessus met en évidence que seuls des effets « larges » (i.e., proches de $\delta = .80$) peuvent être détectés par un test statistique effectué sur un échantillon médian de 42 participant·es. Il est donc tout à fait logique que ces études rapportent mécaniquement des effets positifs d’amplitude large (e.g., $\delta = 1.06$ dans Artut & Karim, 2007 ; $\delta = 1.39$ dans Tarhan & Sesen, 2012 ; $\delta = 0.86$ dans Sahin, 2010 ; $\delta = 2.42$ dans Gömleksiz, 2007). A priori, la première conséquence d’une étude menée en sous-puissance statistique est de ne pas parvenir à détecter l’effet : dans ce cas, il est difficile de conclure à l’efficacité ou à l’inefficacité de la méthode Jigsaw. Néanmoins, dans le cas de la littérature Jigsaw, nous observons un nombre important d’études qui rapportent des effets de Jigsaw positifs et très larges, surtout entre 2010 et 2020 (Vives et al., *en préparation*). Nous pensons que l’accumulation d’études rapportant des effets positifs de d’amplitude supérieure à $\delta > 1.00$, tout en étant menées en sous-puissance statistique, doit alerter les chercheur·es.

2.3. Un probable biais de publication en faveur des résultats positifs

Dans leur article s’intéressant aux tailles d’effet en relation avec les biais de publication, Funder et Ozer (2019) écrivent : « [...] rapporter des tailles d’effet anormalement élevées dans les études menées sur des petits échantillons pourrait être un signe que leur fiabilité globale n’est pas digne de confiance » (Funder & Ozer, 2019, p. 164, traduit en Français). Sur la littérature Jigsaw récente, on peut supposer l’existence d’un biais de publication en faveur des

résultats positifs, ou un « *file-drawer effect* » (Rosenthal, 1979). : Pour reprendre la définition de Rosenthal (1979), un biais de publication existe si la probabilité qu'une étude apparaisse dans la littérature scientifique, et soit disponible pour d'autres chercheurs, dépend du résultat de l'étude en question. Lorsque les études manquantes sont systématiquement différentes - c'est-à-dire avec des résultats négatifs ou non statistiquement significatifs - de celles publiées et analysées, alors on peut établir que la littérature est biaisée.

Plusieurs indicateurs suggèrent l'existence de biais de publication dans la littérature Jigsaw. Premièrement, on devrait s'attendre à observer beaucoup plus d'effets mixtes c'est-à-dire à la fois des effets négatifs et d'autres, non statistiquement significatifs sur une littérature marquée par la sous-puissance statistique⁷. Alors que Newmann & Thompson (1987) ne rapportaient des effets positifs de Jigsaw dans seulement un article sur six, la revue de littérature de Vives et collaborateurs (*en préparation*) rapporte que Jigsaw a eu des effets positifs dans pas moins de 75% des cas (i.e., 66 études sur 88) toutes disciplines et niveaux scolaires confondus. Deuxièmement, compte tenu des tailles d'échantillons faibles, seuls des effets larges sont susceptibles d'être détectés (voir Figures 9 et 10). Or, pour reprendre les termes de Hattie (2009) : « *une taille d'effet de $d = 1.00$ indique une progression d'un écart-type sur la mesure, dans ce cas de figure, la réussite académique. Une augmentation d'un écart-type est typiquement associé à une progression dans la scolarité d'un élève de deux voire trois années de scolarisation, en augmentant le taux d'apprentissage d'environ 50% ($r = .50$) [...]* » (Hattie, 2009, p. 7). Le risque principal que pose ce biais de publication est que seuls les effets surestimés finissent par être visibles dans la littérature scientifique (Franco et al., 2014 ; Lakens & Etz, 2017).

Récemment, la synthèse de méta-analyse dirigée par John Hattie via le site visible-learning.org a intégré la méthode Jigsaw dans le top 10 des interventions académiques les plus efficaces, avec une taille d'effet estimée à $d = 1.20$. Pourtant, cette estimation ne provient que d'une seule méta-analyse de 11 études menées en Turquie entre l'année 2005-2012 (i.e., 6 articles et 5 dissertations), publiée dans une revue non référencée, dont la taille d'échantillon moyenne avoisine $N = 109$. S'il s'avère que ces résultats ont été publiés préférentiellement parce qu'ils rapportaient des effets possibles, alors l'estimation de la méthode peut être trompeuse pour les chercheur·es, mais également les enseignant·es, praticien·nes, politiques et

⁷ L'argument de la probabilité des résultats mixtes peut être décrite telle que : probabilité d'obtenir N résultats positifs d'affilée est égale à N fois la puissance statistique de l'expérimentation. Par exemple, pour 3 études réalisées avec une puissance de test de 80%, la probabilité d'obtenir 3 résultats positifs de manière consécutive est de $0.80*0.80*0.80$ soit $P = .512$.

étudiant-es, et, en conséquence, alimenter la boucle de biais de publication. Le problème posé par la présence d'un biais de publication est que l'on ne dispose virtuellement d'aucun élément permettant de se représenter le nombre d'articles dans le « *tiroir* », c'est-à-dire non publiés. On ne peut pas non plus savoir si ces articles rapportent des effets positifs, mais trop faibles pour être détectés, ou bien négatifs ou nuls (Rosenthal, 1979 ; Borenstein et al., 2010).

La sous-puissance statistique, combinée à un biais de publication en faveur des résultats positifs, peut entraîner un risque de surestimation de l'efficacité de Jigsaw. Outre ces limites, nous développons dans les paragraphes suivants une autre de limite de la littérature : les tests statistiques utilisés pour évaluer la méthode peuvent ne pas être adaptés au design expérimental déployés dans les articles récents.

3. Des analyses statistiques inadéquates pour évaluer l'efficacité de Jigsaw

3.1. Procédures statistiques et design expérimental/indépendance des données

Un second problème particulièrement récurrent dans la littérature Jigsaw repose sur l'utilisation des outils statistiques et leur interprétation. En effet, une grande part des articles empiriques utilise un design expérimental utilisant des mesures pré et post-test identiques ce qui implique l'utilisation, *a minima*, d'un test-*t* apparié qui permet de comparer les gains entre groupes expérimentaux (Borenstein et al., 2011 ; Protogerou & Hagger, 2020). Or, l'approche utilisée dans une grande partie des cas consiste 1) à effectuer un premier *t*-test indépendant sur une mesure initiale (i.e., la mesure pré-test), afin de mettre en évidence une « absence de différences » entre les groupes au début de l'expérimentation⁸ puis 2) un second *t*-test indépendant sur les mesures post-test afin de mettre en évidence une supériorité des scores obtenus en Jigsaw. D'après Robert et collaborateurs (*en préparation*), ce type de procédure concerne plus du tiers de la littérature Jigsaw. Iels rapportent également qu'une part importante des articles rédigés jusqu'à ce jour ne donnent aucune information sur les mesures utilisées pour évaluer l'efficacité de Jigsaw, ce qui pose un problème dans la transparence et la reproductibilité de telles études.

Prenons comme illustration l'étude de Sahin (2010) : dans cette étude, 80 enseignant-es en formation participent à un cours sur l'instruction de la langue turque, dans l'université Atatürk pendant l'année 2009-2010. Iels sont réparti-es dans deux conditions expérimentales,

⁸ Interprétation établie sur la base d'une *p*-value supérieure au seuil alpha de .05, alors même que dans certains cas les différences observées sont de l'ordre de $\delta \pm .20$, ce qui renvoie à une erreur de Type II (i.e., puissance de test insuffisante pour détecter des différences *a priori*).

le groupe Jigsaw II ($N = 42$) et une condition contrôle d'enseignement en classe entière ($N = 38$). Avant le début de l'expérimentation, toutes les participant·es remplissent deux mesures : une mesure d'attitudes envers les cours d'expression écrite (i.e., « Attitudes to Written Expression Scale, ATWES ») qui consiste en une échelle de 25 items de Likert à 5 points⁹, et une mesure de performance basée sur l'examen d'entrée pour les enseignant·es, composée de 25 questions (i.e., Written Expression Achievement Test, « WEAT »). Ensuite, les participant·es prennent part à l'expérimentation, qui consiste à suivre un cours de 4 heures par semaine durant 6 semaines. À la fin de ce cours, les participant·es répondent aux mêmes mesures (i.e., ATWES et WEAT) et certain·es, dans le groupe Jigsaw II, sont sélectionné·es aléatoirement pour répondre à un entretien semi-structuré sur leur appréciation de la méthode. On peut raisonnablement s'interroger sur l'intérêt d'utiliser une mesure d'évaluation d'un cours qui n'a pas encore été traité par les participant·es, ce qui n'est pas un cas isolé, ni le problème le plus important, comme nous le présenterons plus tard. Le problème critique dans ce design pré-post-test repose sur la procédure statistique utilisée pour tester l'hypothèse d'efficacité de Jigsaw II par rapport au groupe contrôle. En effet, Sahin (2011) effectue non pas un *t*-test apparié, procédure qui tient compte des gains *relatifs* entre deux groupes expérimentaux et permet une comparaison, mais bien quatre *t*-test « indépendants » : un premier comparant les scores obtenus au pré-test entre les deux groupes, un deuxième pour comparer les scores entre groupes sur les mesures du post-test, un troisième portant les différences pré-post dans la condition Jigsaw et un quatrième sur les différences pré-post dans la condition contrôle. Outre la multiplication du nombre de tests, et l'absence de correction pour les comparaisons multiples (Gelman et al., 2012), les résultats de Sahin (2010) ne permettent finalement pas de conclure sur l'efficacité de Jigsaw entre le début et la fin de l'expérimentation par rapport au groupe contrôle, puisqu'à aucun moment les gains relatifs ne sont comparés. Une solution aurait été par exemple de créer une nouvelle variable « différence pré-post » et d'effectuer un *t*-test indépendant pour obtenir une comparaison entre groupes, ou bien d'effectuer une ANOVA à mesures répétées, par exemple.

3.2. L'indépendance des données

Cette procédure pose un premier problème, puisque l'auteur conclut à la supériorité de Jigsaw II par rapport à un groupe contrôle, seulement sur la base de la comparaison des scores obtenus en post-test. Le deuxième problème repose sur le lien entre les mesures et l'utilisation

⁹ Echelle non-validée scientifiquement provenant d'un syllabus en langue turque. On ne dispose d'aucune autre information descriptive dans l'article.

du test statistique. En effet, dans cette étude, Sahin (2010) mesure les attitudes et la réussite des participant·es à deux reprises, en utilisant les mêmes échelles. Ces deux mesures ne sont donc pas indépendantes, puisqu'elles proviennent du même échantillon, et surtout du même outil de mesure. Or, l'utilisation d'un test-*t* indépendant pour évaluer l'efficacité de Jigsaw sur la seule base des mesures obtenues en post-test, ne permet pas de rendre compte de cette interdépendance. Idéalement, il faudrait disposer des corrélations pré-post et utiliser des procédures statistiques permettant de mieux prendre en considération la nature interdépendante des données. L'utilisation d'un design pré-post qui ne tient pas rigueur des relations entre mesures amène à un troisième et dernier problème : l'estimation de la taille d'effet observée dans l'article. Dans le cadre d'un *t*-test indépendant, la moyenne ainsi que l'écart-type de la mesure et le nombre de participants dans chaque condition suffisent à calculer une estimation de taille d'effet, notée sous la forme du *d* de Cohen (voir Tableau 9, chapitre 4). Avec les données provenant de l'exemple ($M_{\text{Jigsaw II}} = 23.33, SD = 1.41$; $M_{\text{Contrôle}} = 21.79, SD = 1.74$), le *d* de Cohen pourrait être estimé à $d = 0.98$ (ou $g^* = 0.97$, si une correction pour les petits échantillons est appliquée, Borenstein et al., 2010), ce qui renvoie à une taille d'effet considérée comme « très large » (Cohen, 1992). Or, dans le cas d'un test-*t* apparié, deux autres informations sont nécessaires : le score de différence entre les mesures pré et post-test et son écart-type, ainsi que la corrélation entre les mesures pré et post-test. Sans ces informations, on ne peut pas effectuer une estimation certaine de la taille d'effet, mais il reste possible de simuler une gamme de variables probables (i.e., dispersion du score et taille de la corrélation pré-post) afin d'en avoir une vision approximative (voir Borenstein et al., 2010 pour une telle démarche). Pour reprendre notre exemple, on peut estimer la différence pré-post « moyenne » en soustrayant les scores avant et après l'expérimentation dans chacun des groupes, mais nous ne disposons ni de l'écart-type, c'est-à-dire la dispersion de ces différences scores, ni de la corrélation pré-post entre les mesures. Voici le résultat d'une simulation d'estimation de taille d'effet, sur la base d'un *t*-test apparié (Tableau 7).

Corrélation (<i>r</i>)	Dispersion (<i>SD</i>)		
	Faible (1.5)	Moyenne (3)	Élevée (4.5)
Faible (.20)	2.02	1.01	0.67
Moyenne (.50)	1.60	0.80	0.53
Forte (.80)	1.01	0.51	0.33

Tableau 7: Simulations des tailles d'effet possibles pour un *t*-test apparié (Sahin, 2010)

Ce tableau nous indique dans un premier temps que les tailles d'effets peuvent varier substantiellement en fonction de la corrélation entre les mesures et de la dispersion des scores, c'est-à-dire de la précision de la mesure. Le contraste est tel que l'on peut obtenir des estimations de tailles d'effet très larges (i.e., $g^* = 2.02$), dans le cas de corrélations et de dispersion faibles, ou bien plus petites (i.e., $g^* = 0.33$) dans le cas de corrélations et de dispersion élevées. Le fait qu'aucun de ces deux éléments ne soit rapporté dans les articles empiriques réduit considérablement les chances d'obtenir une estimation fiable de l'effet de Jigsaw. Dans le cas d'un article isolé, où les données seraient insuffisantes pour effectuer de tels calculs, le problème serait mineur. Or, ce problème se retrouve dans de nombreux articles récents portant sur la méthode Jigsaw (e.g., Shaaban, 2006 ; Yoruk, 2016 ; Gomleskiz (2007) ; Basyah et al., 2018 ; Göcer, 2010 ; Hornby, 2009 ; Souvignier & Kronenberger, 2007), ce qui pose à la fois la question du sérieux méthodologique, mais également de la confiance à accorder à de tels résultats.

3.3. Un manque de transparence : le cas de l'université d'Attaturk

On peut également noter qu'une partie de la littérature Jigsaw est caractérisée par un manque de transparence sur la procédure ou les mesures utilisées. Les articles publiés au Moyen-Orient semblent particulièrement concernés par ce problème, notamment ceux menés par Doymus et collaborateurs (2007 ; 2008a ; 2008b ; 2010). Prenons les deux articles publiés par Doymus en 2008 : « *Teaching chemical bonding through jigsaw cooperative learning* », publié dans « *Research in Science & Technological Education* » ainsi que « *Teaching Chemical Equilibrium with the Jigsaw Technique* » publié dans « *Research in Science Education* ». Ces deux articles s'intéressent à l'effet de Jigsaw sur un cours de physique-chimie enseigné auprès d'étudiant·es issues de la même université, l'université d'Attaturk, en Turquie, pendant la même année scolaire 2005-2006. Aucune information concernant la composition et la sélection des étudiant·es dans les groupes n'est donnée. L'appartenance des étudiant·es à cette même université n'est d'ailleurs pas clairement explicitée : il est tout à fait possible que des étudiant·es aient participé à ces deux expérimentations, sans qu'il n'en soit fait mention dans les articles. La méthode de répartition des classes - c'est-à-dire d'aléatorisation - dans les groupes expérimentaux n'est pas non plus précisée, ce qui pose d'autant plus problème que Doymus est en charge de l'enseignement du cours dans l'article « *Teaching chemical bonding through jigsaw cooperative learning* ». Ce cas de figure n'est pas isolé puisqu'une cohorte identique, en provenance de la même université d'Attaturk, est également utilisée dans l'étude de Sahin (2011) et celle de Karacop et Doymus (2013). Sans informations sur la nature précise des

cohortes, et les périodes dans lesquelles ont eu lieu ces expérimentations, le doute subsiste quant à la possibilité que les étudiant·es aient été exposé·es à Jigsaw à plusieurs reprises et dans des disciplines différentes.

Ce scénario assez extrême permet d'effectuer un constat plus global sur l'état de la littérature concernant la méthode Jigsaw : le manque de transparence concerne plusieurs strates de la recherche menée sur cette thématique et se traduit sous une multitude de formes. Robert et collaborateurs (*en préparation*) rapportent que ce problème de transparence se traduit par des difficultés pour évaluer la façon dont la méthode Jigsaw a été effectivement implantée dans l'expérimentation. Iels observent que parmi la centaine d'articles empiriques réalisés jusqu'à ce jour, plus d'une trentaine d'articles manquent de détails dans la description de la procédure expérimentale¹⁰. Vives et collaborateurs (*en préparation*) dressent un portrait encore plus critique de la littérature, et questionnent, sur la base de ces aspects, la confiance que l'on peut accorder à ces articles : « [...] *des problèmes plus graves comme des déformations de la littérature scientifique, des citations, bibliographies et noms d'auteurs erronés, ainsi que des oublis flagrants dans le report des informations sur les mesures, procédures et populations étudiées.* » (p. 41, traduit en Français et emphase ajoutée dans le texte).

4. Tenir compte des limites méthodologiques

Dans l'état actuel de la littérature, il nous semble délicat de conclure sur l'efficacité ou la non-efficacité de la méthode Jigsaw sur les apprentissages. En effet, d'un côté, les articles rédigés entre 1970 et 2000 sont peu nombreux et souvent, ne rapportent pas suffisamment d'éléments statistiques pour être compilés dans une méta-analyse (Johnson et al., 2000) et, d'autre part, les articles plus récents rapportent des effets probablement « trop beaux pour être vrais » (Lakens et Etz, 2017), qui contrastent avec le faible soutien empirique des premiers travaux. Il est donc crucial, afin d'évaluer l'efficacité réelle de cette méthode, de disposer de revues critiques de la littérature, de méta-analyses des travaux récents et de mettre en place de nouvelles recherches réalisées avec une méthodologie appropriée et sur des échantillons larges. Dans les prochains paragraphes, nous présentons l'hypothèse de recherche principale de ce travail de thèse ainsi que les outils méthodologiques et statistiques déployés pour répondre aux questions de puissance statistique et de la fiabilité des résultats développés précédemment.

¹⁰ Sur la base des articles restants, une trentaine fait mention « d'écarts » à la méthode Jigsaw telle que conçue par Aronson et collaborateurs (1978). Cela confirme l'observation de Moskowitz et collaborateurs (1983 ;1985) concernant la flexibilité des enseignant·es dans leur application d'une telle méthode.

4.1. Définition des problématiques et de l'hypothèse de recherche

L'objectif principal de cette thèse est d'apporter des éléments empiriques pour répondre à la problématique : « *Est-ce que la méthode d'apprentissage coopératif de la classe puzzle - ou Jigsaw - permet d'améliorer l'apprentissage des élèves (mesuré par les performances académiques à des évaluations de connaissances), comparativement à un groupe contrôle sans coopération ?* ». Nous tenterons de répondre à cette question en deux étapes : Premièrement, dans le chapitre 4, nous présenterons des éléments méta-analytiques sur les articles publiés entre la période de 2000 à 2020 qui testent l'effet de Jigsaw sur les apprentissages. Nous calculerons l'effet moyen de Jigsaw sur cette variable, en comparaison à un groupe contrôle sans coopération, ainsi que la dispersion de cet effet. Les représentations graphiques (i.e., forest et funnel-plot) ainsi que plusieurs analyses de modérations (i.e., test d'hétérogénéité) permettront de tester les regroupements d'articles que nous développerons dans le chapitre correspondant. Ensuite, dans le chapitre 5, nous rapporterons les résultats issus de cinq expériences menées auprès d'élèves français scolarisés au collège. L'objectif de ces expériences est d'évaluer l'efficacité de Jigsaw par rapport à un groupe contrôle (travail individuel ou maintien des pratiques habituelles d'enseignement), à travers des interventions courtes menées par des expérimentateur·rices (expériences 1 et 2) ou tout au long d'un semestre dans des classes réelles (expériences 3A, 3B et 3C). Les résultats seront regroupés au sein d'une méta-analyse interne qui permettra la comparaison des effets de Jigsaw avec les estimateurs obtenus dans le chapitre 4 (Goh et al., 2016).

Dans l'ensemble de ces études, nous considérerons Jigsaw comme « *une manière de structurer l'interdépendance positive entre les élèves* » (Johnson & Johnson, 2009 ; Butera & Buchs, 2019). Nous avons vu dans les chapitres précédents qu'il existe plusieurs manières de structurer l'interdépendance positive (i.e., récompenses, moyens, frontières), mais également plusieurs conditions à remplir pour que l'apprentissage coopératif ait des résultats positifs sur les apprentissages. Dans ce travail de thèse, nous utiliserons la méthode Jigsaw telle que conçue par Aronson et collaborateurs (1978), c'est-à-dire une méthode qui structure particulièrement l'interdépendance de *moyens*, en allouant des ressources complémentaires aux membres du groupe.

4.2. Spécification de la « plus petite taille d'effet d'intérêt » (SESOI)

L'utilisation d'une plus petite taille d'effet d'intérêt (i.e., « *Smallest Effect Size of Interest* » ou « *SESOI* », Lakens, 2017) s'inscrit dans une volonté de dépasser le test d'hypothèse nulle afin de se focaliser sur la taille d'effet « probable » du phénomène qui nous

intéresse (Lakens et al., 2018). Cette démarche consiste à fixer, avant l'expérimentation, un seuil à partir duquel une taille d'effet sera jugée comme étant d'intérêt pratique, sur la base de la littérature existante et des ressources à disposition. L'utilisation d'une « SESOI » nous permet de comparer les résultats observés à des prédictions formulées *a priori*. Pour ce faire, une procédure statistique peut être utilisée : le test d'équivalence (i.e., « *Two One-Sided Test* » ou « *TOST* », Lakens, 2018). La procédure du test d'équivalence repose sur deux tests-*t* indépendants successifs qui comparent la distribution de données observée dans une expérimentation à celle fixée par la plus petite taille d'effet d'intérêt. Concrètement, le test permet de « *rejeter la présence d'effets au moins aussi larges que la plus petite taille d'effet d'intérêt* ». Les résultats d'un « TOST » dépendent donc à la fois de l'amplitude de l'effet recherché, mais aussi de la puissance statistique dont disposent les chercheur·es lors de l'expérimentation. Dans le cadre de cette thèse, nous utiliserons une plus petite taille d'effet d'intérêt fixée à $\delta = 0.40$, principalement motivée par le critère de « signification pratique » de Hattie (2009) ainsi que son estimation de la taille d'effet moyenne de l'apprentissage coopératif sur les apprentissages de $\delta = .41$ (voir chapitre 1). Les analyses de puissance *a priori* nous permettront d'estimer la taille d'échantillon nécessaire pour disposer d'une puissance de test satisfaisante, et le test d'équivalence nous permettra de confirmer la présence ou d'absence de tels effets sur nos expériences.

Il convient aussi de préciser que l'utilisation de « SESOI » ou d'une interprétation en termes de seuil d'intérêt pratique n'a pas pour objectif de décrédibiliser les petits effets, les résultats nuls ou non significatifs. John Cohen (1962) lui-même mettait en garde contre une interprétation des effets par convention selon sa classification. Par exemple, il serait hâtif de conclure qu'une intervention ayant un effet « faible » proche de $\delta \pm 0.20$ soit totalement inintéressante. Si l'intervention en question est peu coûteuse et non invasive, il peut s'avérer judicieux de la mettre en place même si les effets qu'elle pourrait avoir sont de faible amplitude (Slavin, 2011 ; Yeager & Walton, 2011 ; Xu et al., 2018). L'intérêt de spécifier une taille d'effet d'intérêt, de manière générale, est d'assurer que les études soient menées dans des conditions satisfaisantes de puissance de test, particulièrement lorsque la littérature est potentiellement biaisée en faveur des résultats positifs focalisés sur les *p*-value.

4.3. Présentation de la procédure de méta-analyse interne

Dans le chapitre 5, nous utiliserons une procédure de méta-analyse interne afin de synthétiser les résultats obtenus et de rapporter des indicateurs proches de la méta-analyse du chapitre 4 (i.e., taille d'effet moyenne, dispersion, tests d'hétérogénéité). D'après l'article de

Goh et collaborateurs (2016) qui décrit cette procédure, une méta-analyse interne peut être effectuée dès que les chercheur·es disposent d'au moins deux études. Bien évidemment, un nombre plus important d'études est nécessaire pour effectuer des analyses plus complexes, notamment les tests d'hétérogénéité qui permettent de comparer les effets estimés entre regroupements d'études. Les auteur·es soulignent plusieurs intérêts à utiliser cette approche. D'une part, l'inclusion de plusieurs études, dont celles rapportant des effets mixtes (i.e., positifs, négatifs et/ou nuls), permet d'améliorer la transparence et de réduire le « *file-drawer effect* ». C'est en rapportant toutes les études réalisées sur une question de recherche, et non pas seulement rapportant des résultats statistiquement significatifs, qu'une méta-analyse interne permet de valoriser les effets mixtes et d'éviter le biais de publication. D'autre part, les estimateurs obtenus via une méta-analyse interne permettent de combiner les résultats de plusieurs études, et sont donc théoriquement plus puissants et plus précis que des résultats pris indépendamment les uns des autres (Borenstein et al., 2010 ; Goh et al., 2016 ; Funder & Ozer, 2019). D'autre part, pour Borenstein et collaborateurs (2010), une méta-analyse permet de dépasser l'interprétation des résultats sur la base de leur significativité statistique, c'est-à-dire sur la base de leur *p*-value, et de s'intéresser plutôt à l'amplitude de leurs effets. Enfin, combiner plusieurs études permet d'établir si les effets observés sont homogènes, c'est-à-dire peu dispersés, ou hétérogènes, et ainsi apporter des pistes d'interprétation quant aux raisons de cette hétérogénéité à l'aide de tests statistiques.

5. Synthèse et transition

Le Tableau 8 ci-dessous reprend les différentes limites soulevées dans la littérature Jigsaw, ainsi que les outils déployés dans la thèse pour y répondre. Compte tenu des limites évoquées vis-à-vis de la puissance statistique dans les articles récents, nous avons souhaité mettre en place des expériences en mesure de nous informer sur la présence d'un effet au moins aussi élevé que $\delta = 0.40$. L'utilisation combinée de test d'équivalence et des procédures de méta-analyses permettront d'apporter des éléments supplémentaires pour qualifier les effets observés : ces analyses nous permettront d'obtenir plusieurs indicateurs (e.g., taille d'effet moyenne, intervalle de confiance, dispersion) pour mieux quantifier l'effet de Jigsaw, et éventuellement les modérateurs de son efficacité. L'interprétation de ces indicateurs, au regard de la littérature et de ses limites, servira à ancrer nos résultats dans une remise en question de l'efficacité de la méthode Jigsaw.

Limites identifiées	Réponses proposées
1. Puissance statistique	<p>Test de puissance <i>a priori</i> et analyses de sensibilité</p> <p>Utilisation d'une <i>plus petite taille d'effet d'intérêt</i> et de tests d'équivalence</p> <p>Méta-analyse (chapitre 4) et méta-analyse interne sur les expérimentations (chapitres 5)</p>
2. Indépendance des données	<p>Correction pour le coefficient de corrélation intra-classe</p> <p>Comparaison des analyses effectuées avec et sans covariables</p>

Tableau 8: Synthèse des perspectives engagées pour répondre aux limites de la littérature

Résumé du chapitre 3

- La littérature récente sur Jigsaw est marquée par un fort risque de sous-puissance statistique et pourtant, les tailles d'effet observées sont largement positives, ce qui contraste avec 1) les premières estimations de l'efficacité de Jigsaw faites avant les années 2000 et 2) les effets moyens de l'apprentissage coopératif sur les performances académiques.
-
- La sous-puissance, les analyses statistiques inadaptées et le manque de transparence tendent à suggérer la présence d'un biais de publication qui contribue à fausser les estimations de l'efficacité de Jigsaw : la publication d'études rapportant des effets positifs et larges peut contribuer à biaiser la littérature.
-
- Pour dépasser les limites de la littérature Jigsaw, nous avons choisi d'effectuer plusieurs expérimentations en se basant sur une plus petite taille d'effet d'intérêt estimée sur la base de la littérature scientifique. L'utilisation de tests d'équivalence et d'outils méta-analytiques nous permet de tester empiriquement une même hypothèse de recherche, à travers plusieurs expériences.

Chapitre 4. L'utilisation de Jigsaw en classe permet-elle de favoriser les apprentissages ? Une méta-analyse des travaux récents

1. Objectif du chapitre

L'objectif de ce chapitre sera d'examiner l'effet de Jigsaw sur les apprentissages tels que rapportés dans les articles récents publiés dans la littérature. La dernière revue cherchant à comparer l'efficacité de différentes pédagogies coopératives ayant été publiée en 2000 par Johnson et ses collègues, nous nous intéresserons aux études publiées entre 2000 et 2020. Nous présenterons plusieurs éléments méta-analytiques (i.e., estimation de taille d'effet, analyses d'hétérogénéité, forest et funnel-plot) afin de répondre à la question de recherche présentée dans le chapitre précédent. Ainsi, nous évaluerons si l'effet global de Jigsaw, comparativement à un groupe contrôle sans coopération structurée, est bien positif et d'amplitude faible à moyenne, c'est-à-dire compris entre $\delta = 0.20$ et 0.40 (Johnson et al., 2000, Hattie, 2009).

2. Méthode

2.1. Critères d'inclusion et d'exclusion

Comme mentionné précédemment, plusieurs revues ont déjà été menées sur les travaux publiés entre le moment de la création de Jigsaw, dans les années 1970, et 2000, date à laquelle la méta-analyse de Johnson et collaborateurs (2000) a été rédigée. À notre connaissance, cette méta-analyse est la dernière à avoir comparé l'efficacité de plusieurs pédagogies coopératives entre elles. Or, depuis cette méta-analyse, un nombre important d'études a été réalisé et il n'existe pas encore, à notre connaissance, de méta-analyse cherchant à compiler ces travaux. Comme pour la plupart des méta-analyses, les articles sélectionnés dans cette méta-analyse répondent à un certain nombre de critères (voir chapitres 2 et 3) notamment :

1. Être publié dans une revue de psychologie ou de sciences de l'éducation à comité de lecture, référencées dans Scopus ou Web of Science.
2. Rapporter des études suivant un design expérimental avec groupe contrôle, c'est-à-dire permettant de comparer la technique Jigsaw à une condition sans intervention coopérative structurée. Les articles de type « implémentation », c'est à dire des essais théoriques ou des articles présentant l'application de Jigsaw dans une classe ne sont pas retenus.
3. Être mené sur une population dont le niveau scolaire s'étend du primaire jusqu'aux études supérieures et dont les participant·es retenu·es ne sont pas en situation

d'apprentissage « particulières » (e.g., troubles dys-, « haut potentiel », adultes en situation de rescolarisation, etc.). Par ailleurs, l'expérimentation doit avoir été menée pendant l'année scolaire régulière.

4. Rapporter suffisamment d'éléments statistiques (i.e., moyennes, écart-type, taille d'échantillon) pour qu'une estimation de taille d'effet soit possible.

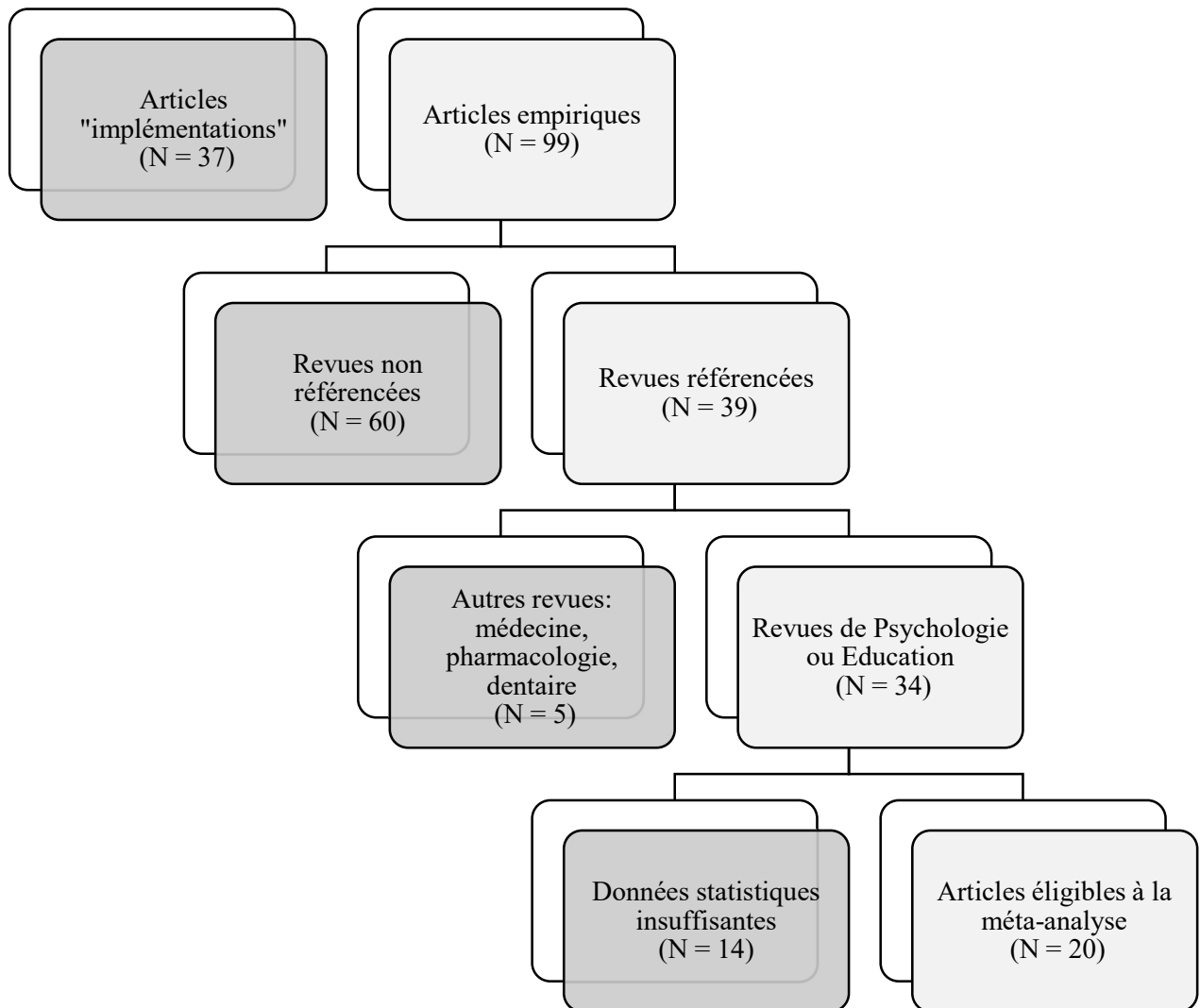


Figure 11: Tri des articles en fonction des critères d'éligibilité à la méta-analyse

La Figure 11 ci-dessous reprend les différentes étapes de sélection des articles dans la méta-analyse de ce chapitre. Sur la base de ces filtres successifs, 20 articles (i.e., environ 15% de la littérature récente) ont été sélectionnés parmi ceux publiés entre les années 2000 et 2020 (Robert et al., *en préparation*). La liste des études incluses dans la méta-analyse figure dans le Tableau 9. Dans les prochains paragraphes, nous donnerons des détails supplémentaires sur les aspects méthodologiques des articles conservés, notamment sur la catégorisation de ces derniers en vue des tests de modération.

2.2. Mesure de l'apprentissage

L'objectif de cette méta-analyse est d'estimer l'efficacité de la méthode Jigsaw sur les apprentissages, mesurée par des scores à des évaluations de connaissances, c'est-à-dire via des mesures de performance académique, mais également de tester des modérateurs de cet effet à travers les caractéristiques des études sélectionnées et sur la base de plusieurs hypothèses. Comme argumenté par Kyndt et collaborateurs (2013), une méta-analyse des travaux récents permet aussi de comparer les résultats de Jigsaw entre plusieurs périodes de publications. Dans le cadre de cette méta-analyse, on peut envisager que l'effet de Jigsaw ne diffère pas des premières observations effectuées entre 1980 et 2000 (Johnson et al., 2000 ; Newmann & Thompson, 1987 ; Sharan, 1980 ; Slavin, 1980), et que les articles récents contribuent à augmenter la précision de l'estimation de la taille d'effet. Ou bien, compte tenu des limites développées dans le chapitre 3, on peut s'attendre à ce que les estimations de taille d'effet soient différentes, et interroger l'origine de ces disparités à travers des tests d'hétérogénéité.

Tous les articles présents dans cette méta-analyse rapportent des éléments descriptifs et statistiques nécessaires pour effectuer ces tests. Par exemple, Roseth et collaborateurs (2018) utilisent plusieurs évaluations de connaissances (i.e., 3 quiz notés sur 18 et 5 autres sur 70), espacées tout au long du semestre, afin de mesurer l'efficacité de Jigsaw comparativement à groupe contrôle d'enseignement habituel. Dans une autre étude, Doymus (2007) utilise plusieurs mesures pour tester l'efficacité de Subject Jigsaw par rapport à un groupe contrôle d'enseignement habituel : un score sur des connaissances de chimie « générales » (i.e., « Chemical Achievement Test » : 20 questions à choix multiples) ainsi qu'un score sur les évaluations spécifiques aux séquences vues pendant l'expérimentation (i.e., « Phase Achievement Test » : 4 modules de 4 questions à choix multiples chacune). Enfin, Souvignier et Kronenberger (2007) mesurent l'apprentissage de leurs élèves à quatre reprises, à la fin de chaque séquence vue en cours. Les scores obtenus aux évaluations de connaissance de la fin de ces séquences sont comparés à ceux obtenus en pré-test, et entre trois conditions

expérimentales : Jigsaw, Jigsaw avec une aide métacognitive et un groupe contrôle d'enseignement habituel (i.e., individuel en classe entière). Pour cette méta-analyse, les scores obtenus sur différentes mesures ont été combinés puis moyennés, afin de disposer d'*une seule mesure de performance académique par étude*. La procédure exacte pour le calcul de ces effets est présentée dans les paragraphes suivants. Plus de la moitié des études incluses dans cette méta-analyse contiennent plusieurs évaluations de connaissances, nous avons donc systématiquement appliqué la même procédure pour regrouper et moyenner ces scores.

2.3. Groupe contrôle

Les articles éligibles à cette méta-analyse permettent une comparaison entre au moins un groupe expérimental - qui renvoie à l'utilisation d'une méthode Jigsaw - et un groupe contrôle, sans Jigsaw. Parmi les 20 articles, 18 utilisent un groupe contrôle qui consiste en une « instruction en classe entière » ou « instruction dirigée par l'enseignant-e ». L'intégralité des ces articles ne rapportent qu'une seule étude, bien certaines utilisent plusieurs mesures d'apprentissage. Dans le cas d'articles utilisant plusieurs groupes expérimentaux, tels que Karacop et Doymus (2013) qui comparent Subject Jigsaw, une technique d'enseignement assisté par ordinateur et un groupe contrôle « d'enseignement habituel », nous n'avons rapporté que la taille d'effet estimée par contraste entre Jigsaw et le groupe contrôle (voir également Souvignier & Kronenberger, 2007 ; Law, 2011 ou Moreno, 2009 pour l'utilisation de trois groupes expérimentaux). Dans les deux autres articles, l'utilisation de Jigsaw est comparée à une méthode d'enseignement « active » (Berger & Hänze, 2009) ou bien à une méthode pédagogique coopérative non structurée (Hornby, 2009) : dans le cadre de cette méta-analyse, plutôt que de nous intéresser au groupe de comparaison, nous allons tester d'autres facteurs susceptibles de modérer l'effet de Jigsaw sur les performances académiques, notamment la version ou le type de Jigsaw, la durée de l'expérimentation et la discipline étudiée.

Article	N_{total}	Revue	Localisation	Jigsaw	Discipline	Durée totale	ES (<i>var</i>)
Artut & Tarim (2007)	81	Asia-Pacific Journal of Teacher Education	Turquie	Jigsaw II	Mathématiques	36h (9 semaines)	1.58 (.06)
Basyah et al. (2018)	60	The New Educational Review	Indonésie	Jigsaw	Économie	Non précisé (14 semaines)	3.17 (.15)
Berger & Hänze (2009)	286	International Journal of Science Education	Allemagne	Jigsaw	Physique	9h (non précisé)	-0.08 (.02)
Doymus (2007)	108	Journal of Chemical Education	Turquie	Subject Jigsaw	Chimie	12h (3 semaines)	0.96 (.05)
Doymus (2008a)	68	Research in Science & Technological Education	Turquie	Subject Jigsaw	Chimie	10h (5 semaines)	1.75 (.20)
Doymus (2008b)	36	Research in Science Education	Turquie	Jigsaw	Chimie	9h (3 semaines)	1.04 (.08)
Ghaith & El-Malak (2004)	48	Educational Research and Evaluation	Liban	Jigsaw II	Lecture	15h (5 semaines)	0.28 (.13)
Gömleksiz (2007)	66	European Journal of Engineering Education	Turquie	Jigsaw II	Anglais	28h (14 semaines)	2.38 (.15)
Hänze & Berger (2007)	137	Learning and Instruction	Allemagne	Jigsaw	Physique	9h (non précisé)	-0.24 (.02)
Hornby (2009)	44	Journal of Education for Teaching	Angleterre	Jigsaw II	Education	2 h (1 seule session)	0.76 (.09)
Karacop & Doymus (2013)	115	Journal of Educational Technology	Turquie	Subject Jigsaw	Chimie	Non précisé (5 semaines)	1.09 (.09)
Law (2011)	279	Journal of Research in Reading	Chine	Jigsaw	Lecture	5h (5 semaines)	0.67 (.04)
Moreno (2009)	87	Learning and Instruction	États-Unis	Jigsaw (<i>assisté par ordinateur</i>)	Botanique	1h (1 seule session)	-0.44 (.11)

Article	N_{total}	Revue	Localisation	Jigsaw	Discipline	Durée totale	ES (<i>var</i>)
Roseth et al. (2018)	258	Journal of Educational Psychology	États-Unis	Jigsaw	Anatomie	+100h (24 semaines)	0.26 (.01)
Sahin (2011)	71	Asia Pacific Education Review	Turquie	Jigsaw III	Expression écrite	24h (6 semaines)	0.86 (.06)
Shaaban (2006)	44	Reading Psychology	Liban	Jigsaw II	Lecture	80h (8 semaines)	0.23 (.13)
Souvignier & Kronenberger (2007)	208	British Journal of Educational Psychology	Allemagne	Jigsaw	Physique	4h30 (6 semaines)	-0.48 (.02)
Tarhan & Sesen (2012)	38	Chemistry Education	Turquie	Jigsaw	Chimie	1h15 (1 seule session)	1.51 (.13)
Tarhan et al (2013)	61	Research and Practice	Turquie	Jigsaw	Physique	8h (2 semaines)	2.57 (.12)
Yoruk (2016)	63	Research in Science & Technological Education	Turquie	Jigsaw	Chimie	6h (6 semaines)	0.61 (.07)

Tableau 9: Liste des études incluses dans la méta-analyse.

2.4. Effet global et dispersion

2.4.1. Calcul de l'effet moyen et spécification du modèle

L'estimation de l'effet global ainsi que les analyses de modération permettent d'analyser la dispersion de l'effet de Jigsaw entre les études et sont effectuées via un modèle de méta-analyse à effet aléatoire. C'est le modèle le plus répandu dans la plupart des méta-analyses en psychologie, puisqu'il suppose l'existence de variabilités dans les estimations de taille d'effet entre les différentes études (Borenstein et al., 2010). C'est cette variation que nous avons mesurée à l'aide de tests de modérateurs, ou « tests d'hétérogénéité ». Nous avons également calculé un effet global de Jigsaw sur les apprentissages, sur la base des résultats rapportés dans les articles sélectionnés. Pour ce faire, nous avons estimé une taille d'effet moyenne (*ES*) ainsi que sa dispersion à l'aide du paramètre *g* de Hedge en suivant la procédure suivante : dans un premier temps, nous avons calculé un *d* de Cohen associé à chaque étude, c'est-à-dire la différence standardisée entre les moyennes des deux groupes expérimentaux (Jigsaw vs Contrôle) pondérées par un écart-type intra-étude. Puis nous avons corrigé cet indicateur *d*, ainsi que sa variance associée par un facteur *J* de « correction pour les petits échantillons » afin d'arriver au paramètre *g* de Hedge. Les formules utilisées pour le calcul d'ES sont présentées dans le Tableau 10 ci-dessous.

<i>d</i> de Cohen	$\frac{\text{Moyenne (Jigsaw)} - \text{Moyenne (Contrôle)}}{\text{Ecart - type intra étude (S)}}$
<i>S</i> (Écart-type intra-étude)	$\sqrt{\frac{(n1 - 1)SD1^2 + n2(-1)SD2^2}{n1 + n2 - 2}}$
Variance de <i>d</i>	$\frac{n1+n2}{n1*n2} + \frac{d^2}{2(n1+n2)}$
Correction <i>J</i>	$J = 1 - \frac{3}{4(df-1)} \text{ où } df = (n1+n2)-2$
<i>g</i> de Hedge	$g = d * J$
Variance de <i>g</i>	$J^2 * \text{Variance de } d$
Intervalle de confiance	$IC = g \pm (1.96 * \sqrt{\text{Variance de } g})$

Tableau 10: Formules de calcul des estimations de taille d'effet moyenne (*ES*).

2.4.2. Ajustements

Comme présenté dans le chapitre 3, la littérature portant sur Jigsaw souffre d'un certain nombre de défauts qui se retrouvent dans les considérations méta-analytiques de ce chapitre. En effet, certains paramètres nécessaires aux calculs d'*ES*, tels qu'ils auraient dû être effectués, sont indisponibles dans les articles originaux (e.g., corrélation entre les mesures dans les designs pré-post, corrélation entre les mesures dans les designs à variables dépendantes multiples). Dans notre cas, puisque ces données n'ont pas été rapportées par les auteur·es, même après une demande par mail, nous avons choisi de rapporter les *ES* à partir des scores de post-test uniquement. Cette contrainte pouvant avoir pour conséquence de surestimer les *ES* observées, les résultats de cette méta-analyse doivent être interprétés avec précaution (Maasen et al., 2020 ; Wiernik & Dahlke, 2020).

Pour le cas des mesures multiples, qui constituent une part importante de la littérature Jigsaw, nous avons estimé un coefficient de corrélation inter-mesures sur la base des recommandations de Borenstein et al. (2010). Ce coefficient de corrélation est important pour calculer le « Facteur d'Inflation de la Variance » (VIF) et ainsi corriger un autre problème que pose la non-indépendance des données : dans les études qui rapportent plusieurs mesures d'un même « *proxy* » (e.g., des mesures de performances académiques) la variance tend à diminuer quand le nombre de mesures augmente. Cela a pour conséquence de gonfler artificiellement la « précision » ou plutôt le « poids » de l'étude dans la méta-analyse, puisque ce dernier est inversement relié à la variance. Pour contrebalancer cette inflation, il est nécessaire de corriger la variance par le facteur de VIF, et ainsi attribuer un poids moins biaisé aux études qui rapportent plusieurs mesures. Sur la base des propositions de Borenstein et collaborateurs (2010), et dans une volonté d'appliquer une correction assez forte et homogène sur ces mesures, nous avons choisi d'utiliser un coefficient de corrélation inter-mesures fixé à $r = .71$ ¹¹.

2.4.3. Pondération

Une fois l'estimation de la taille d'effet et de sa variance calculée pour chaque étude, nous avons utilisé la « méthode des moments » de DerSimonian-Laird pour appliquer un poids, c'est-à-dire une importance relative à chaque étude dans la méta-analyse (Borenstein et al., 2010). Cette procédure est utile lors d'une méta-analyse à effet aléatoire puisque le poids associé à chaque étude dépend de sa précision (i.e., l'inverse de la variance, soit un indice de dispersion des scores) et de la dispersion moyenne des études entre elles (τ^2). En règle générale,

¹¹ Aucun article présenté dans cette méta-analyse ne rapporte le coefficient de corrélation entre les mesures pré-post. A titre indicatif, l'article de Göcer (2010) rapporte un coefficient proche de $r \pm .95$.

une étude sera plus précise si les échantillons sont larges (i.e, puissance statistique élevée) et surtout si les scores sont peu dispersés autour de la moyenne (i.e., faible hétérogénéité). Ce faisant, les études « précises » sont pondérées à la fois par leur propre variance interne, mais également par une variance globale entre études. En conséquence, l'*ES* dépend moins de quelques études précises qui peuvent tirer l'effet vers le haut ou le bas, mais plutôt des études dans leur ensemble.¹²

2.5. Hétérogénéité et tests de modulation

Les analyses de modulation cherchent à tester si les regroupements d'études expliquent des différences d'hétérogénéité, c'est-à-dire des différences dans la dispersion des effets de Jigsaw. Elles permettent de comparer plusieurs études selon un critère spécifique. Pour des raisons de puissance statistique, nous avons choisi d'organiser les regroupements selon deux catégories. Dans le cas de variables continues, une classification catégorielle a été effectuée à partir du score médian sur notre échantillon. Par exemple, pour le modérateur « durée de l'expérimentation » nous avons regroupé les études dans les catégories « courtes » ou « longues » sur la base de la durée médiane (en heure) dans l'échantillon, à savoir $\Delta t = 9$. Compte tenu des multiples analyses de modulation effectuées, nous avons appliqué une correction en divisant le seuil α de rejet de l'hypothèse nulle par un facteur 5, qui correspond au nombre de modérateurs testés (Ioannidis & Trikalinos, 2007 ; Streiner & Norman, 2011). Le seuil de signification statistique étant traditionnellement fixé entre $\alpha = [.10 ; .05]$ pour ce genre d'analyse, nous estimerons que les *p*-value inférieures à $\alpha = .01$ seront considérées comme statistiquement significatives. En d'autres termes, si la valeur *p* du test de modulation est inférieure au seuil de rejet α , on peut déduire que le modérateur explique une part de variance non nulle dans la dispersion des effets de Jigsaw. À l'inverse, le non-rejet de l'hypothèse nulle équivaut à H_0 : « *Le regroupement d'études n'est pas relié à la variation dans l'estimation de taille d'effet* ».

2.5.1. Type de Jigsaw

Dans un premier temps, nous comparerons les tailles d'effet observées en fonction du type de Jigsaw utilisé dans les études. En effet, sur notre échantillon d'article, la moitié utilise une procédure proche de celle développée par Aronson et collaborateurs (1978), tandis que l'autre moitié utilise une version « révisée » telle que Jigsaw II (Shaaban, 2006) ou Subject Jigsaw (Doymus, 2007) par exemple. Comme présenté dans les chapitres 2 et 3, l'efficacité des

¹² Voir van der Linden & Goldberg (2020) pour une alternative à la méthode de DerSimonian Laird.

versions « révisées » de Jigsaw comparativement à la première version ne fait pas l'objet d'un consensus même si, théoriquement, les améliorations de Jigsaw devraient améliorer son efficacité. Pour apporter des éléments empiriques à cette affirmation, nous formulons l'hypothèse H_{a1} que les versions révisées de Jigsaw devraient avoir des effets plus élevés sur les apprentissages. Dit autrement, on s'attend à ce que le regroupement des études utilisant une version « révisée » de Jigsaw rapporte des effets plus larges que ceux utilisant la version « originale ».

2.5.2. Localisation de l'article

Dans les chapitres précédents, nous avons développé l'argument selon il est possible que la localisation des études ait un impact sur l'estimation des tailles d'effets (voir chapitres 2 et 3). Pour cette raison, nous comparerons les effets observés dans les études provenant de localisations orientales (i.e., $k_{\text{Moyen-Orient}} = 14$, $k_{\text{Asie}} = 2$) avec ceux de localisations occidentales (i.e., $k_{\text{Europe}} = 4$, $k_{\text{États-Unis}} = 2$). Nous formulons l'hypothèse H_{b1} que les effets estimés via les études menées au Moyen-Orient seront plus importants que ceux estimés dans les études menées dans les autres localisations.

2.5.3. Durée de l'intervention

La plupart des articles rapportent des indications qui permettent d'évaluer la durée de l'intervention, c'est-à-dire le temps qu'ont passé les participant·es dans la condition Jigsaw ou contrôle. Sur notre échantillon, nous avons utilisé la durée médiane en heures (i.e., $\Delta t = 9$) afin de catégoriser les interventions de durée « courtes » ou « longues » et de tester une modération par la durée d'exposition à Jigsaw. Aronson et collaborateurs (1978 ; 2011) affirment que la technique Jigsaw nécessite un temps d'adaptation avant d'être utilisée de manière optimale dans la classe. Sur la base de cet argument, on peut formuler l'hypothèse H_{c1} selon laquelle les effets de Jigsaw seront plus élevés dans le cadre d'interventions « longues », comparativement à des interventions « courtes ».

2.5.4. Discipline

Pour ce qui est de la discipline - ou la nature du contenu pédagogique vue en cours - Aronson et collaborateurs (1978 ; 2011) suggèrent que les matières qui se prêtent le mieux à la méthode Jigsaw sont celles qui font l'emphase sur la narration et les compétences de l'écrit : l'histoire, la géographie et les sciences civiques et politiques sont censées être les plus adaptées, et celles liées aux mathématiques, biologies et aux langues plus difficiles à mettre en place. Cohen (1994) argumente que les disciplines littéraires favorisent la formulation de problèmes à « réponses multiples », qui bénéficient des interactions et de l'articulation de plusieurs points

de vue entre les individus, alors que les contenus scientifiques disposent le plus souvent d'une réponse unique à formuler pour résoudre l'énoncé. Cependant, les méta-analyses de Lou et collaborateurs (1996) et de Kyndt et collaborateurs (2013) rapportent des effets plus positifs de l'apprentissage coopératif sur les disciplines « scientifiques » (i.e., mathématiques, physique-chimie, sciences de l'ingénieur) que sur les autres. De plus, la majorité des recherches menées sur Jigsaw portent sur les disciplines scientifiques¹³(Robert et al., *en préparation*) et des résultats positifs de la méthode sont rapportés, quelle que soit la discipline étudiée (Vives et al., *en préparation*). Le contenu disciplinaire pourrait donc être un modérateur de l'effet de Jigsaw, mais ces éléments nous amènent à ne pas formuler d'hypothèse *a priori* sur le sens de cette modération.

2.5.5. Niveau scolaire

Enfin, nous nous intéresserons aux effets de l'âge, à travers le niveau scolaire, sur l'efficacité de Jigsaw. La majorité des articles récents étant rédigés sur les populations d'étudiant·es dans l'enseignement supérieur, nous disposons finalement d'assez peu d'éléments permettant d'effectuer des hypothèses fortes sur le sens de cette modération. Kyndt et collaborateurs (2013) observent des différences en fonction du niveau scolaire, avec des effets de l'apprentissage coopératif plus élevés dans le primaire et le tertiaire, comparativement au secondaire, mais n'expliquent pas l'origine de ces différences. Dans la littérature Jigsaw, l'effet du niveau n'est pas systématiquement testé, et l'on observe aussi bien des effets positifs, nuls et négatifs dans toutes les strates scolaires, que ce soit dès la maternelle (Akçay, 2016) jusqu'aux études supérieures (Hornby, 2009). La modération par le niveau scolaire sera donc également effectuée de manière *exploratoire*.

2.5.6. Biais de publication

Afin de caractériser les éventuels biais de publication (Fanelli, 2012 ; Ioannidis & Trikalinos, 2007) nous effectuerons un *forest-plot* cumulatif pour décrire la dispersion des *ES* en fonction de la précision des études : si les études proviennent d'un échantillon spécifique (e.g., seulement celles rapportant des effets positifs) d'une littérature, alors la méta-analyse peut refléter un tel biais. Par exemple, sur un *forest-plot*, une relation positive entre la taille d'effet rapportée dans une étude et sa variance peut être interprétée comme un signe de présence de biais méthodologique, puisque les études peu précises tendent à surestimer les *ES* (Borenstein et al., 2010 ; Ferguson & Heene, 2012 ; Kühberger et al., 2014). Un *funnel-plot* permet de

¹³ La physique-chimie, les mathématiques, la médecine les sciences du vivant couvrent plus des deux tiers des contenus pédagogiques déployés pour tester l'efficacité de la méthode Jigsaw (Robert et al., *en préparation*).

mettre en relation la dispersion des études autour de l'ES globale, en fonction de leur précision, en rapportant un coefficient *d'asymétrie*. En absence de biais, les études sont supposées être réparties de manière systématique autour de l'ES, et la présence de « pentes » d'un côté ou l'autre du funnel-plot peut suggérer, à l'inverse, une présence de biais.

3. Résultats

3.1. Effet global de Jigsaw

En prenant en compte l'ensemble des études éligibles pour cette méta-analyse, l'estimation de taille d'effet moyenne est positive, large et statistiquement significative : $g^* = 0.88$, $SE = .19$, 95% CI [0.51 ; 1.25], $Z = 4.70$, $p < .001$. Cette estimation indique qu'en moyenne, les participants en condition Jigsaw obtiennent des performances académiques largement supérieures à celles obtenues dans les conditions contrôles.

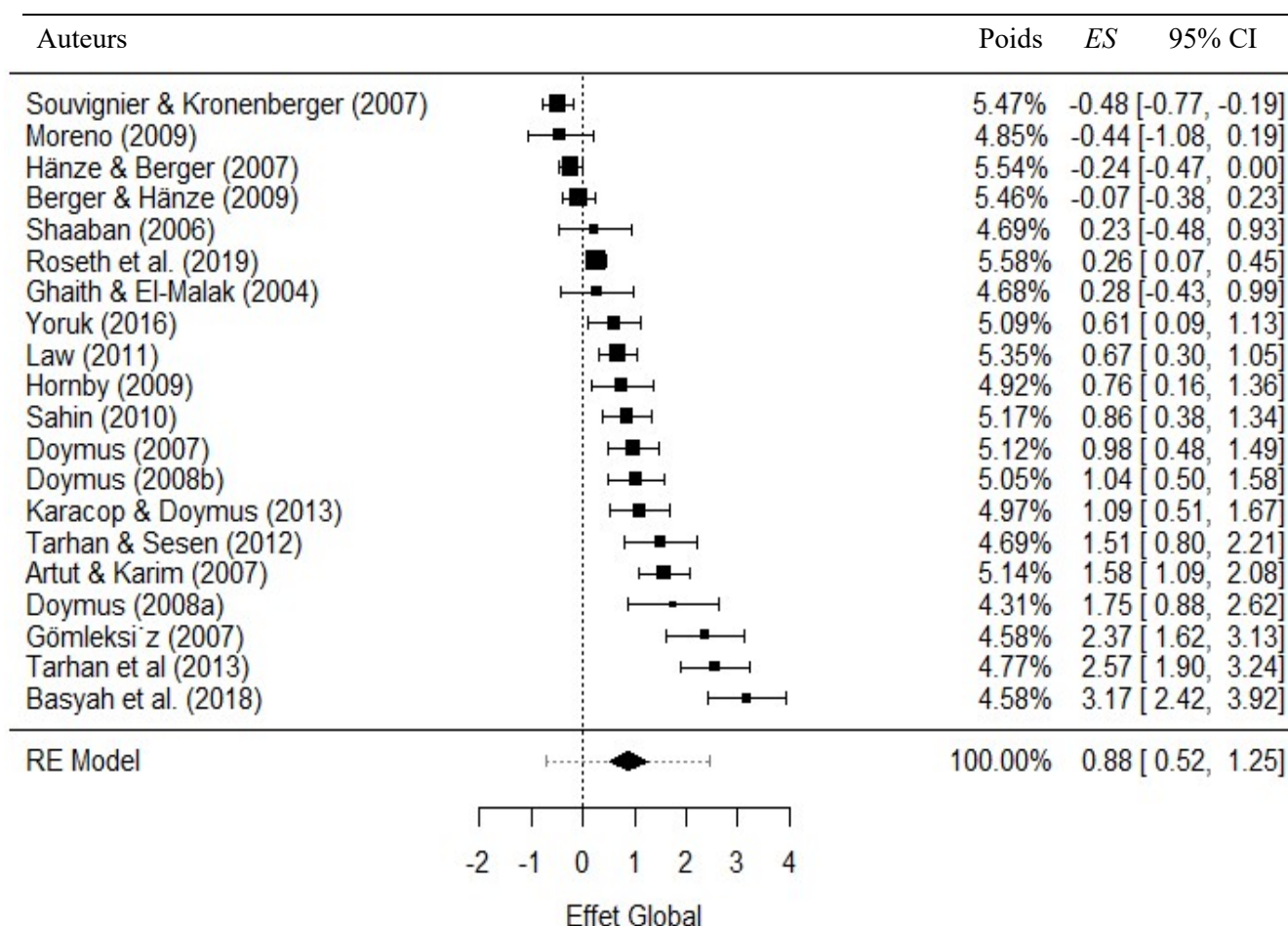


Figure 12: Forest-plot des ES observées

Le forest-plot met en évidence une hétérogénéité assez importante entre les études (Figure 12). Par exemple, on peut observer que les études précises, c'est-à-dire avec des intervalles de confiance restreints, reportent systématiquement des tailles d'effets de plus petites et centrées autour de zéro que les études moins précises. De la même manière, les études rapportant des tailles d'effet élevées semblent avoir un intervalle de confiance particulièrement large : comme mentionné précédemment, il est possible que des études peu précises surestiment l'efficacité de Jigsaw, en comparaison à des études plus précises (Funder & Ozer, 2019 ; Ferguson & Heene, 2012 ; Lakens et Etz, 2017). Cette relation est décrite de manière plus détaillée dans le funnel-plot présenté dans la Figure 13 ci-dessous, qui permet de tester l'éventuelle présence d'un biais de publication (Borenstein et al., 2010).

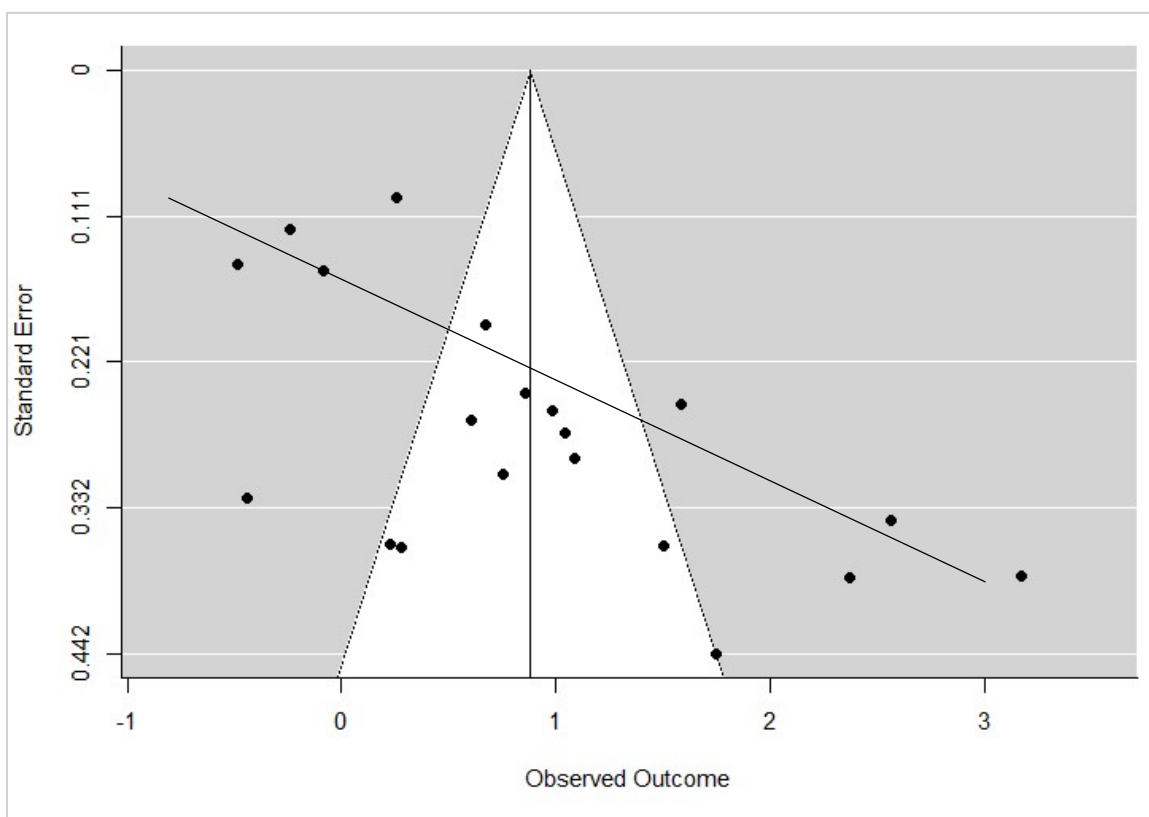


Figure 13: *Funnel-plot des ES observées.*

Le test d'asymétrie apporte un élément en faveur de l'hypothèse d'un biais de publication puisque les tailles d'effet sont inégalement réparties autour de l'ES global, $Z = 4.13$, $p < .001$. Dans le cas d'une littérature non biaisée, les tailles d'effet sont censées être réparties de manière relativement symétrique, autour de l'estimateur global, ce qui n'est pas le cas ici. On observe par exemple que les estimations plus précises – c'est-à-dire renvoyant à des erreurs standards faibles – sont majoritairement situées à gauche de l'ES global (i.e., « Observed Outcome »). À l'inverse, les études moins précises tendent à être reliées à des estimations d'ES

plus larges. Cette relation négative entre précision (i.e., variance) et taille d'effet observée se traduit par un coefficient de corrélation de Kendall de $\tau = -.43$, $p = .007$, c'est-à-dire est un signal probable de biais de publication (voir Kühberger et al., 2014 pour un argument similaire).

3.2. Dispersion de l'effet et tests de modérateurs

Les premiers éléments visuels semblent indiquer une forte hétérogénéité entre les études, avec une influence potentielle de la précision des études dans les estimations de taille d'effet de Jigsaw. La variable statistique d'hétérogénéité Q est significative et très importante, $Q(df = 19) = 265.86$, $p < .001$. La dispersion entre études est également très marquée, $\tau^2 = .62$, $SE = .30$, et une très large part de cette dispersion est due à des facteurs autres que la chance, le hasard, ou les biais d'échantillonnages, $I^2 = 92.85\%$. Le Tableau 11 ci-dessous synthétise les résultats principaux des tests d'hétérogénéité.

Variable	Q_b	p -value	Q_w	M	95%CI
Localisation	22.62	< .001*			
Orient ($n = 14$)			91.23	1.31	0.90 ; 1.72
Occident ($n = 6$)			29.44	-0.03	-0.40 ; 0.33
Durée	2.95	.053 ^{ns}			
< 9 heures ($n = 10$)			106.80	1.18	0.67 ; 1.68
> 9 heures ($n = 10$)			123.01	0.56	0.07 ; 1.04
Discipline	1.11	.218 ^{ns}			
Scientifique ($n = 13$)			175.49	0.74	0.28 ; 1.19
Autre ($n = 7$)			57.37	1.17	0.50 ; 1.83
Type de Jigsaw	< 1	.242 ^{ns}			
Jigsaw I ($n = 10$)			180.05	0.89	0.37 ; 1.40
Jigsaw révisé ($n = 10$)			49.65	0.88	0.44 ; 1.31
Population	< 1	1 ^{ns}			
Primaire/Secondaire ($n = 7$)			120.29	0.98	0.20 ; 1.75
Supérieur ($n = 13$)			143.79	0.84	0.39 ; 1.29

Note. * significative au seuil $\alpha < .01$; ^{ns} non significative.

Tableau 11: Résultats des analyses de modulation.

3.2.1. Type de Jigsaw

Jigsaw a vu apparaître différentes versions au fil des années : les premières modifications en Jigsaw II ont été proposées par Slavin (1983 ; 1989) afin de pallier les problèmes de l'absence de récompenses de groupe et de buts communs. D'autres versions ont vu le jour plus récemment, mais n'ont pas fait l'objet d'une validation empirique systématique (voir chapitre 2). Nos analyses ne mettent pas en évidence de différences dans les estimations de taille d'effet pour les études utilisant la première version de Jigsaw ($ES_{\text{Jigsaw}} = 0.89$, $SE = .26$, 95% CI [0.37 ; 1.40]) et des versions ultérieures ($ES_{\text{Jigsaw révisé}} = 0.88$, $SE = .22$, 95% CI [0.44 ; 1.31]), $Q(df = 1) < 1$, $p = .25$. Ce résultat semble indiquer que la prise en compte de la version de Jigsaw utilisée dans les études ne permet pas de mieux comprendre la dispersion des ES de Jigsaw. L'hypothèse H_{a0} est donc conservée.

3.2.2. Localisation de l'article

Le Tableau 9 donne les détails de la provenance des articles inclus dans cette méta-analyse. En tenant compte des limites développées dans le chapitre 3, nous avons formulé l'hypothèse que les articles provenant de pays orientaux pouvaient rapporter des estimations de taille d'effet plus importantes que celles provenant d'autres localisations. Les résultats indiquent que ce modérateur est statistiquement significatif, $Q(df = 1) = 22.62$, $p < .001$. Les résultats confirment l'hypothèse H_{b1} : les effets sont positifs et d'amplitude très large dans les articles provenant du Moyen-Orient ($ES_{\text{Orient}} = 1.31$, $SE = .21$, 95% CI [0.90 ; 1.72]), comparativement aux articles provenant d'autres localisations qui rapportent des effets nuls, en moyenne ($ES_{\text{Occident}} = -0.03$, $SE = .19$, 95% CI [-0.39 ; 0.33]). Cette modulation permet d'expliquer approximativement la moitié de la dispersion de l'effet de Jigsaw dans cette méta-analyse ($R^2 = .54$).

3.2.3. Durée de l'intervention

D'après les résultats de l'analyse d'hétérogénéité et notre seuil de décision fixé à $\alpha = .01$, nous ne pouvons pas rejeter l'hypothèse nulle H_{c0} : la différence d'efficacité de Jigsaw en fonction des durées courtes ($ES_{\text{courtes}} = 1.18$, $SE = .26$, 95% CI [0.67 ; 1.68]) ne diffère pas significativement de celle des durées plus longues ($ES_{\text{longues}} = 0.56$, $SE = .25$, 95% CI [0.07 ; 1.04]), $Q(df = 1) = 2.95$, $p = .053$. On peut néanmoins observer deux tendances intéressantes dans ces estimations. Premièrement, il semble que les effets de Jigsaw soient plus importants dans les durées d'expérimentation « courtes », ce qui va à l'encontre du sens de notre hypothèse H_{c1} et de l'affirmation de Aronson et Patnoe (2011) selon laquelle la coopération selon Jigsaw nécessite du temps à se mettre en place avant d'observer des gains sur les apprentissages.

Deuxièmement, même si le seuil de signification statistique n'est pas atteint, notamment suite aux corrections pour les comparaisons multiples, la différence relative entre les durées longues et courtes est assez importante, presque de l'ordre d'un facteur deux. On peut raisonnablement évoquer la forte dispersion des scores, c'est-à-dire les intervalles de confiance assez larges, et un manque de puissance statistique (i.e., nombre d'études dans chaque catégorie) pour expliquer pourquoi le modérateur « durée » n'est pas statistiquement significatif, dans cette méta-analyse.

3.2.4. Discipline

La majorité des études comprises dans la méta-analyse portent sur des contenus scientifiques en grande majorité de physique et chimie. Néanmoins, nous n'observons pas de différences statistiquement significatives entre les *ES* entre les études menées sur une discipline scientifique ($ES_{\text{scientifique}} = 0.74$, $SE = .23$, 95% CI [0.28 ; 1.19]) ou d'un autre type ($ES_{\text{autre}} = 1.17$, $SE = .34$, 95% CI [0.50 ; 1.83]), $Q(df = 1) = 1.11$, $p = .22$. Comme pour la durée, nous ne pouvons pas rejeter l'hypothèse nulle : dans cette méta-analyse, le modérateur « discipline » n'explique pas suffisamment de variance dans l'effet de Jigsaw.

3.2.5. Niveau scolaire

Dans cette méta-analyse, les différences en termes d'*ES* entre les étudiant·es dans le supérieur ($ES_{\text{supérieur}} = 0.74$, $SE = .22$, 95% [0.31 ; 1.17]) ne diffèrent pas statistiquement de celles estimées chez les populations dans le primaire ou le secondaire ($ES_{\text{Primaire/Secondaire}} = 1.24$, $SE = .39$, 95% CI [0.47 ; 2.01], $Q(df = 1) = 1.25$, $p = .19$). L'hypothèse nulle est donc conservée.

3.3. Corrélations entre modérateurs

Les analyses de modérateurs permettent de tester si les caractéristiques de certains articles expliquent une part de dispersion dans les effets rapportés dans une méta-analyse. Or, la plupart du temps ces modérateurs ne sont pas indépendants les uns des autres. Le Tableau 12 ci-dessous met en évidence que certains facteurs sont fortement corrélés entre eux. Par exemple, on observe que la corrélation entre la précision de l'étude et la taille de l'échantillon est positive et forte, $r = .70$, $p < .001$, mais également avec la localisation des articles, $r = .49$, $p = .03$. Cette dernière peut s'expliquer notamment par le fait que les articles rédigés par des équipes de chercheur·es de pays orientaux tendent à être réalisées sur des plus petits échantillons, $r = -.62$, $p < .001$. On peut également observer une corrélation entre le type de Jigsaw et la durée des expérimentations menées, puisque les études menées sur la première version de Jigsaw tendent à durer moins longtemps que celles utilisant une version révisée, $r = -.48$, $p = .05$.

	Taille d'échantillon	Type de Jigsaw	Localisation	Discipline	Durée	Précision	Niveau scolaire
Taille d'échantillon (min = 36 ; max = 440)	–	-.46* [-.03 ; -.75]	.62** [.24 ; .83]	-.19 [-.28 ; .58]	-.29 [-.66 ; .21]	.70*** [.37 ; .87]	-.38 [-.70 ; .08]
Type de Jigsaw (0 : Jigsaw I ; 1 : Jigsaw révisé)	–	–	-.41 [-.72 ; .04]	-.31 [-.66 ; .15]	-.48* [.01 ; .77]	-.19 [-.27 ; .59]	.49* [.07 ; .77]
Localisation (0 = Orient ; 1 = Occident)	–	–	–	-.04 [-.47 ; .40]	-.42 [-.74 ; .05]	.49*** [.06 ; .76]	-.38 [-.70 ; .08]
Discipline (0 = Non-scientifique ; 1 = Scientifique)	–	–	–	–	-.43 [-.75 ; .06]	.29 [-.17 ; .65]	.26 [-.20 ; .63]
Durée (0 = Courte ; 1 = Longue)	–	–	–	–	–	-.04 [-.47 ; .41]	.03 [-.42 ; .46]
Précision (se) (min = .01 ; max = .20)	–	–	–	–	–	–	-.20 [-.59 ; .26]

Note. *** $p < .001$. ** $p < .01$. * $p < .05$.

Tableau 12: Corrélations et intervalles de confiance des modérateurs de la méta-analyse

La corrélation positive entre le niveau scolaire, le type de Jigsaw ($r = .49, p = .027$) provient du fait que Subject Jigsaw est particulièrement utilisé chez les étudiant·es du supérieur, notamment en Turquie. Enfin, on remarque que ces corrélations sont marquées par une très forte dispersion, d'où les intervalles de confiance larges. D'autres recherches sont requises afin de préciser les interactions entre ces modérateurs et augmenter la précision de nos estimations.

3.4. Tests de puissance statistique appliqués à la méta-analyse

Comme pour une expérience, la question de la puissance statistique s'applique également aux méta-analyses (Borenstein et al., 2010). Dans ce cas de figure, nous avons cherché à savoir si nous disposions de suffisamment d'études, et si ces dernières étaient suffisamment précises pour détecter une taille d'effet donnée. Une analyse de sensibilité, c'est-à-dire une analyse de puissance *a posteriori*, permet d'apporter des éléments de réponses à cette question. Deux scénarii sont présentés ci-dessous pour mieux en saisir l'intérêt. Les résultats observés dans semblent indiquer deux tendances principales : une estimation de taille d'effet globale large, très supérieure à celles rapportées dans les méta-analyses précédentes (pour rappel, $d \pm .20$ pour Johnson et al., 2000), mais surtout une forte hétérogénéité entre les études, notamment en fonction de leur précision. Sur la base de notre échantillon, nous pouvons estimer le nombre *moyen* de participants par étude de $N = 110$ (soit une cinquantaine de participants par condition), la médiane sur l'échantillon étant située à $N = 74$. À partir de cette information, il est possible d'estimer la puissance de notre méta-analyse par rapport à une estimation du véritable effet de Jigsaw. En suivant le développement présenté dans les premiers chapitres de ce manuscrit, nous avançons l'argument que Jigsaw devrait avoir des effets analogues, voire légèrement inférieurs, à ceux de l'apprentissage coopératif en général, soit $\delta = 0.40$ (Hattie, 2009). Dans ce premier cas de figure, nous semblons disposer d'une puissance de test suffisante : la probabilité de détecter un effet de $\delta = 0.40$ sur la base des articles inclus dans la méta-analyse est proche de 99% (voir Figure 14). En d'autres termes, si le véritable effet de Jigsaw sur les performances académiques, en comparaison à un groupe contrôle, est bien de $\delta = 0.40$, une méta-analyse de 20 articles menés sur des échantillons d'au moins 50 participants par condition offre des conditions de test satisfaisantes pour détecter un tel effet.

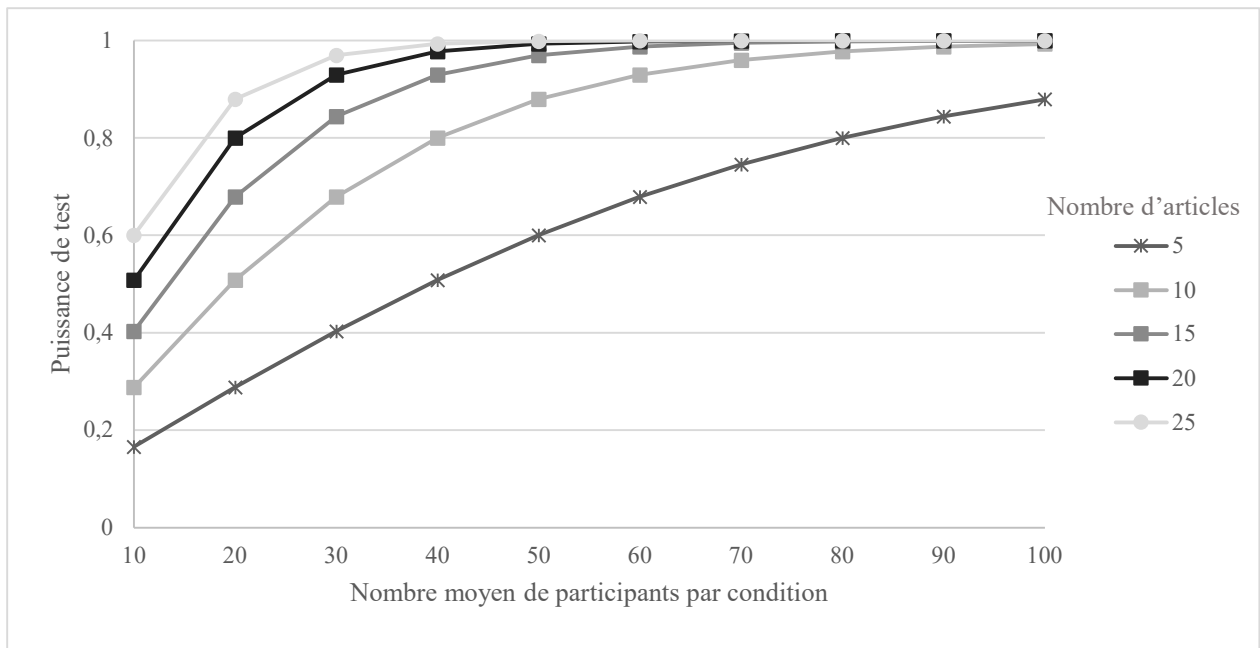


Figure 14: Analyse de puissance a posteriori pour méta-analyse (Borenstein et al., 2010)

Dans l'hypothèse du scénario plus pessimiste où le véritable effet de Jigsaw est plus proche de $\delta = 0.20$, la puissance de test de notre méta-analyse diminue à 61% de chance de détecter cet effet, sur la base du même nombre d'article. Deux possibilités sont alors envisageables pour les recherches futures : augmenter le nombre d'études à inclure dans la méta-analyse et/ou le nombre de participants par condition. Il convient de noter que ces analyses de puissance menées a posteriori ne nous informent pas sur la nature du véritable effet de Jigsaw : elles mettent seulement en évidence que la puissance statistique d'une méta-analyse et par extension, sa qualité, dépend des paramètres de puissance pour les études individuelles. Ainsi, bien que notre méta-analyse conclut sur un effet global de Jigsaw positif et très large, nous avons mis en évidence que près de la moitié des études sur lesquelles se base cette estimation sont probablement réalisées en sous-puissance et susceptibles de surestimer les effets et donc de biaiser la littérature. En d'autres termes, pour résoudre cet éventuel problème de biais de publication dans la littérature Jigsaw, il est crucial de mener des études sur des échantillons suffisamment larges et d'améliorer la rigueur et la qualité des expérimentations (e.g., utiliser des mesures répétées ou des designs appariés en rapportant les indicateurs nécessaires pour effectuer des estimations plus précises). Ces éléments sont discutés de manière plus détaillée dans les paragraphes suivants.

4. Discussion

4.1. Synthèse des principaux résultats

L'objectif principal de cette méta-analyse était de tester l'ampleur des effets de l'utilisation de Jigsaw sur les apprentissages, tels que mesurés par des tests de compréhension ou des évaluations de connaissances des contenus pédagogiques enseignés pendant les séquences travaillées en Jigsaw, comparativement à des modalités n'impliquant pas d'interdépendance positive. Dans notre méta-analyse, l'effet global de Jigsaw sur les apprentissages est positif et de large amplitude, $g^* = 0.88$, 95% CI [0.51 ; 1.25], mais surtout caractérisé par une très large dispersion et une hétérogénéité importante, $Q(df = 19) = 265.86$, $p < .001$; $I^2 = 92.85\%$. Il semble surprenant d'observer que l'effet global de Jigsaw rapporté dans cette méta-analyse soit quatre, voire six fois plus important que les dernières estimations rapportées par Johnson et collaborateurs (2000), voire même plus important que ceux rapportés classiquement dans la littérature de l'apprentissage coopératif (voir Tableau 2, chapitre 1). Néanmoins, ces comparaisons ne rendent pas compte de la diversité des études incluses dans la méta-analyse présentée dans ce chapitre. Compte tenu de l'hétérogénéité importante entre les études, c'est surtout dans la description de la dispersion des effets de Jigsaw que cette méta-analyse prend son sens.

Premièrement, les forest et funnel-plot tendent à suggérer la présence de biais de publication, via une asymétrie dans la répartition des effets autour de l'estimation globale. Dans notre méta-analyse coefficient de corrélation entre la taille d'effet et la taille de l'échantillon est de $\tau = -.43$. Dans leur recherche qui compare la relation entre taille d'effet et d'échantillon sur près de 400 articles empiriques en psychologie, Kühberger et collaborateurs (2014) observent une corrélation de $r = -.45$, soit une estimation très proche de la nôtre. Pour ces auteur·es, les estimations de taille d'effet provenant d'échantillons faibles sont plus sensibles aux erreurs d'échantillonnage, ce qui impacte leur précision et augmente la probabilité de rapporter des estimations extrêmes. Pour elleux, la corrélation négative entre ces indicateurs implique une tendance en psychologie à rapporter sélectivement les résultats positifs et surestimer les tailles d'effets provenant d'échantillons faibles. Ferguson et Heene (2012) apportent des éléments supplémentaires pour quantifier le risque de biais de publication dans les méta-analyses : pour eux, approximativement 25% des méta-analyses publiées dans les « top journals » de la psychologie présentent un risque de biais de publication, souvent peu évalué par les méta-analystes. Les auteurs rapportent aussi que la corrélation négative entre

taille d'échantillon et estimation de taille d'effet est négative dans 80% des cas, ce qui suggère que la présence de biais de publication dans les méta-analyses est très répandue.

Deuxièmement, nos analyses de modération mettent en évidence que ce sont principalement les études effectuées dans les pays orientaux (i.e., Turquie, Liban, Chine et Indonésie) qui tendent à rapporter des effets très larges et positifs, tandis que les études effectuées dans des pays occidentaux (i.e., États-Unis, Allemagne et Angleterre) rapportent des effets nuls, en moyenne. Cette modération explique environ la moitié de la dispersion de l'effet de Jigsaw, ce qui contribue à identifier la localisation des articles comme un facteur important pour comprendre l'inconsistance des résultats observés dans la littérature. Par exemple, les analyses de corrélations indiquent que la localisation est fortement liée au critère de précision de l'étude, puisque ce sont les études en provenance de pays orientaux qui sont majoritairement réalisées sur des échantillons faibles. La corrélation entre la précision de l'étude et l'ampleur de l'effet, à la lumière des limites développées dans le chapitre 3, doit alerter le lecteur sur l'existence probable de biais de publication : l'effet de Jigsaw pourrait être principalement surestimé à cause des faiblesses méthodologiques de ces études, qui tendent à gonfler l'effet global par leur nombre plus important. Cependant, d'autres interprétations peuvent être formulées pour nuancer ces interprétations.

D'une part, il est possible que certains tests d'hétérogénéité n'aient pas dépassé le seuil de signification statistique alors même que les modérateurs expliquent une vraie part de variance de l'effet de Jigsaw. En effet, le problème de sous-puissance statistique s'applique de manière similaire aux méta-analyses et à ces tests, qui tendent à être effectués dans des conditions de sous-puissance (Borenstein et al., 2010 ; Ioannidis & Trikalinos, 2007). Ainsi, certaines caractéristiques (e.g., durée de l'expérimentation, population étudiée, type de Jigsaw) peuvent tout à fait avoir un impact sur l'efficacité de Jigsaw, qui n'ont simplement pas été détectées par nos tests. Les corrélations entre modérateurs indiquent également que plusieurs d'entre eux peuvent être confondus. Par exemple, la plupart des études menées en Turquie tendent à être effectuées sur des durées moyennes (i.e., environ 6 semaines), sur des populations d'étudiant·es à l'université et sur des contenus scientifiques. Nous pensons que d'autres études sont nécessaires pour clarifier l'existence de modérateurs potentiels et dissocier les relations entre ces derniers : disposer d'un jeu de données plus important nous permettrait d'avoir des comparaisons plus fines et de mieux quantifier l'estimation de l'effet et de sa dispersion. Mener des expérimentations sur des échantillons plus larges est une première étape pour disposer d'une

puissance de test plus importante, et clarifier l'origine des variations des effets de Jigsaw dans la littérature.

D'autre part, les analyses par tests de modération et de biais de publication sont rétrospectives, et n'ont pas vocation à établir de lien de causalité : il s'agit simplement d'apporter des éléments descriptifs sur la littérature Jigsaw, afin d'en tirer des connaissances pour mettre en place les meilleures recherches possibles. Par exemple, le fait de ne pas observer de tailles d'effets plus larges dans les versions « révisées » de Jigsaw ne suffit pas à affirmer que les améliorations successives apportées à cette pédagogie n'ont aucune efficacité. Bien au contraire, c'est grâce à l'accumulation de travaux de recherche qu'il est possible d'établir que la structuration de l'interdépendance positive au sein du groupe permet de maximiser son efficacité (Johnson et al., 2000 ; Slavin, 2011 ; Topping et al., 2017). De la même manière, les signaux probables de biais de publication, observés dans les articles en provenance du Moyen-Orient, n'apportent pas une preuve conclusive de l'existence d'un tel biais : il est virtuellement impossible de le quantifier de manière précise, puisque nous ne pouvons pas disposer de l'intégralité des recherches non publiées, non référencées, ou simplement abandonnées dans le « tiroir » (Borenstein et al., 2010 ; Fanelli, 2012 ; Ferguson & Heene, 2012). D'autres explications sont également possibles, comme l'existence des différences culturelles ou de phénomènes locaux, qui pourraient expliquer nos résultats.

4.2. Un modérateur culturel ?

Dans l'ensemble, cette méta-analyse met en évidence une disparité des effets de Jigsaw en fonction de localisation des études : bien que l'hypothèse d'un biais méthodologique et de publication soit la plus plausible (voir Levine et al., 2009 pour un argument similaire) pour les études en provenance de pays orientaux, d'autres explications sont possibles. Par exemple, Kyndt et collaborateurs (2013) rapportent un effet de la « culture » dans les tests d'hétérogénéité de leur méta-analyse. Pour ces auteur·es, l'origine de cet effet provient des différences culturelles sur l'axe « collectivisme-individualisme » des participant·es : puisque les cultures orientales tendent à être plus orientées vers le collectivisme que les cultures occidentales, les effets de la coopération pourraient être plus importants, car plus marqués dans le contexte (Triandis, 1989). Dans la méta-analyse de Kyndt et collaborateurs (2013), les effets observés dans les cultures « collectivistes » sont plus élevés que ceux observés dans les cultures « individualistes ». Iels interprètent ce résultat comme une preuve que les individus provenant de ces milieux ont une préférence pour les tâches interdépendantes. À l'inverse, les individus provenant de milieux individualistes se retrouveraient en décalage - ou en *mismatch* - avec les

enjeux d'un travail en coopération, d'où une efficacité moindre. Néanmoins, l'hypothèse inverse est également envisageable. Dans les cultures occidentales majoritairement marquées par l'individualisme (Stephens et al., 2012), on pourrait s'attendre à ce que les effets de la coopération soient encore plus importants, car le référentiel « contrôle » serait beaucoup plus individualiste que le référentiel de base des cultures collectivistes, plus interdépendant. Bien que ces interprétations soient intéressantes, on peut néanmoins noter que, dans leur méta-analyse, Kyndt et collaborateurs (2013) n'explorent pas l'hypothèse alternative de la présence de biais de publication : puisque l'analyse est basée en grande partie sur les études dont nous avons précédemment identifié qu'elles étaient en sous-puissance ou présentaient d'importantes limites méthodologiques (i.e., Doymus, 2007 ; Doymus, 2008a ; Doymus, 2008b ; Gomleksi'z, 2007 ; Sahin, 2011 ; Shaaban, 2006).

Concernant la puissance statistique, il est possible que nous ne disposions pas, dans notre méta-analyse, d'un nombre suffisant d'articles rédigés en Asie, ou aux États-Unis par exemple, pour effectuer des tests de modération par la localisation de manière plus appropriée. La représentativité des localisations « orientales » vs « occidentales » n'est donc peut-être pas la plus pertinente dans l'absolu, puisqu'elle regroupe des pays qui peuvent être marqués entre autres par des différences culturelles importantes. D'autre part, la classification de Triandis (1989) selon l'axe collectivisme-individualisme a été critiquée et améliorée, notamment par Muthukrishna et collaborateurs (2018) qui proposent un modèle à deux dimensions, plutôt qu'un seul continuum. Leur modèle permet de comparer chaque pays en fonction de leur proximité avec les États-Unis d'un côté, et la Chine de l'autre, en regroupant plusieurs indicateurs de « distance culturelle » comme la distance géographique, idéologique (i.e., échelles de collectivisme-individualisme, distance au pouvoir, féminité-masculinité, évitement de l'incertitude de Hofstede, 1980) et les valeurs (i.e., harmonie, autonomie, hiérarchie, voir Scharwtz, 2006 pour le détail). Dans leur article, les auteur·es rapportent la construction d'un indicateur permettant de regrouper ces différentes échelles sous un même indicateur, le « cultural fixation index », ou CF_{ST} , utilisé pour comparer plusieurs pays aux États-Unis d'Amérique, pays le plus représenté dans la recherche en psychologie sociale, et la Chine, qui sert de point de comparaison principal dans la recherche en psychologie interculturelle (Muthukrishna et al., 2018). Le calcul du CF_{ST} permet donc de visualiser comment se distinguent plusieurs pays par rapport aux États-Unis et à la Chine, sur un graphique à deux dimensions. Par exemple, selon ces critères, la Turquie serait aussi proche des États-Unis que de la Chine ($CF_{ST} = .120$ et $.119$ respectivement), mais culturellement éloignée de l'Indonésie

($CF_{ST} = .178$ et $.167$) et du Liban ($CF_{ST} = .103$ et $.175$). Selon cette classification l'Allemagne ($CF_{ST} = .080$ et $.114$) et l'Angleterre ($CF_{ST} = .046$ et $.172$) seraient également modérément éloignées sur l'axe de distance avec la Chine. Dans une perspective interculturelle, l'utilisation d'un tel modèle pourrait être donc être extrêmement pertinente. Dans le cadre de nos recherches, une meilleure distinction entre pays permettrait une meilleure classification en termes de « localisation » : plutôt que de regrouper plusieurs pays sous le terme « pays orientaux », nous pourrions classer les localisations sur ces deux dimensions de proximité avec les États-Unis et la Chine. Si un tel modérateur culturel existe, on pourrait s'attendre à observer des différences entre les études menées en Turquie et en Indonésie par exemple : même si ces pays sont regroupés dans la catégorie « orientale » dans notre méta-analyse, ils sont assez éloignés sur l'axe chinois du « cultural fixation index ». Effectuer des comparaisons avec ce critère nous permettrait de tester les deux hypothèses de « *match* » ou « *mismatch* » entre les valeurs individualistes et collectivistes dans un contexte d'apprentissage coopératif. Cependant, pour des raisons de puissance statistique, davantage d'études sont nécessaires afin que de tels regroupements puissent être possibles.

Pour revenir à d'autres pistes d'explication concernant les disparités entre localisations, on peut aussi concevoir l'existence de « *local truth* » - ou de « phénomènes locaux » - au sein desquels de tels résultats peuvent apparaître. Sur la simple base de nos données, on ne peut réfuter l'hypothèse selon laquelle Jigsaw fonctionne très bien au Moyen-Orient, mais dans des conditions particulières (e.g., seulement auprès d'étudiant·es sur un contenu de physique-chimie tel qu'enseigné à l'université). Cependant, l'écart important entre l'efficacité de Jigsaw au Moyen-Orient et dans les autres pays soulève une question : *est-ce que les différences culturelles suffisent à expliquer un tel écart ?* Dans leur méta-analyse sur les effets de l'apprentissage coopératif sur la réussite académique, Kyndt et collaborateurs (2013) rapportent une différence dans les tailles d'effet de l'ordre de $ES = 0.38$, $se = .07$, 95% CI [0.25 ; 0.53] en faveur des cultures « orientales ». Or, dans la méta-analyse présentée dans ce chapitre, la différence est de l'ordre de $ES = 1.32$, soit près de trois fois plus. Bien que les différences entre cultures puissent certainement jouer un rôle dans les disparités de l'effet de Jigsaw, d'autres études plus précises, puissantes et transparentes doivent être déployées pour pouvoir répondre à cette question.

4.3. Conclusions intermédiaires

Les éléments développés dans ce chapitre permettent de se faire une vision plus nuancée de la littérature Jigsaw : bien que l'effet moyen de Jigsaw sur les performances académiques soit positif et de large amplitude, des disparités importantes existent entre les études et les analyses d'asymétrie suggèrent la présence d'un biais de publication. De plus, les limites évoquées dans le chapitre 3 sont illustrées par les corrélations entre modérateurs (i.e., taille d'échantillon, localisation et précision de l'étude) et les analyses d'hétérogénéité, ce qui supporte l'hypothèse du biais de publication. Les « preuves » de l'efficacité de Jigsaw, notamment les tailles d'effets anormalement larges, et les résultats de cette méta-analyse nous amènent à formuler l'argument que nous ne disposons pas à ce jour de suffisamment de preuves qui garantissent que ces effets reflètent l'efficacité de Jigsaw plus qu'un biais de publication. (Ferguson et Heene, 2012 ; Kühberger et al., 2014). Cependant, la méta-analyse n'offre pas une lecture des effets de Jigsaw en termes de causalités : pour ce faire, davantage d'expérimentations doivent être menées, dans de meilleures conditions de puissance statistique. Dans le prochain chapitre, nous présenterons les résultats de cinq expériences menées auprès d'élèves français au collège, afin de tester expérimentalement notre hypothèse de recherche.

Résumé du chapitre 4

- Une méta-analyse des articles publiés depuis les années 2000 et testant les effets de Jigsaw sur les apprentissages met en évidence un effet positif et large, marqué par une très forte hétérogénéité entre études.
-
- Les analyses d'hétérogénéité indiquent que les études provenant du Moyen-Orient rapportent des effets très larges, contrairement à celles menées en Europe ou aux États-Unis qui rapportent des effets nuls. La localisation des études explique quasiment la moitié de la dispersion des effets de Jigsaw.
-
- Les analyses d'asymétrie et les corrélations entre modérateurs suggèrent une présence de biais méthodologiques et de publication, particulièrement sur les études en provenance du Moyen-Orient.
-
- Les autres modérateurs (i.e., type de Jigsaw, discipline et population étudiée) n'expliquent pas une part de variance suffisamment importante dans la dispersion des effets de Jigsaw. Cependant, on ne peut pas conclure à l'absence d'impact de ces modérateurs sur les effets de Jigsaw : la puissance statistique des tests de modérateurs est potentiellement faible, compte tenu de la forte hétérogénéité entre les études.

Chapitre 5. Jigsaw améliore-t-il les apprentissages en science d'élèves de 6^e ? Cinq expériences et une méta-analyse interne

1. Objectif du chapitre

L'objectif de ce chapitre sera de rapporter cinq expériences testant l'efficacité de Jigsaw sur les apprentissages chez une population d'élèves de 6^e. Dans ces expériences, les participant·es ont étudié des séquences pédagogiques identiques soit avec la méthode Jigsaw soit de manière individuelle (études 1 et 2), ou en suivant un enseignement habituel (études 3A, E3B, E3C). Les deux premières études (i.e., E1 et E2), ont été menées sur une seule session d'apprentissage d'une durée de 2 heures tandis que trois autres (i.e., E3A, E3B, E3C) ont été mises en place dans des conditions de classe réelle pour une durée d'un trimestre avec des enseignant·es volontaires. Nous faisons l'hypothèse que, en comparaison avec un groupe contrôle qui consiste en un travail individuel (E1 et E2), ou un enseignement habituel (E3A, E3B, E3C), les effets de Jigsaw seront positifs et d'amplitude proche de $\delta = 0.40$.

2. Vue d'ensemble

2.1. Choix de la population

Comparativement aux États-Unis ou au Moyen-Orient, il existe relativement peu d'articles empiriques testant l'efficacité de Jigsaw sur une population française. À notre connaissance, une seule étude réalisée auprès d'élèves de lycée professionnel (Darnon et al., 2012) utilise Jigsaw afin d'améliorer le sentiment d'auto-efficacité, en mathématiques et en français, des lycéen·es de filières professionnelles. Dans cette étude, les élèves ont participé à quatre sessions de 2 heures d'aides aux devoirs (e.g., améliorer sa compréhension écrite, s'entraîner à la présentation orale) soit de manière coopérative (i.e., condition Jigsaw) soit de manière individuelle (i.e., condition contrôle). Les auteur·es s'intéressaient au sentiment d'auto-efficacité des participant·es, mesuré avant et après l'expérimentation : selon la littérature scientifique, les populations d'élèves de lycée professionnel sont particulièrement concernées par des sentiments d'auto-efficacité assez faible (Gurtner et al., 2009 cités dans Darnon et al., 2012) et pourraient bénéficier de Jigsaw pour la rehausser.

n°	N_{total}	Sujet	Durée	Groupe Contrôle	Matériel expérimental	Test d'apprentissage	Test statistique
1	252	Mathématiques	1 séance (2h)	Individuel	Crée par les expérimentateur·rices pour les besoins de l'étude	1 test : 17 QCM + 8 résolutions de problème	t -test indépendant
2	313	SVT				1 test (9 SCM, 2 questions ouvertes)	ANOVA non- paramétrique ¹⁴
3A	110	Physique- Chimie	16 séances de 1h (16h)	Enseignement habituel (TAU)	Crée par les enseignant·es, adaptés à la méthode Jigsaw en collaboration avec l'expérimentateur	4 tests (résolution de problèmes, questions ouvertes, définitions)	t -test apparié
3B	74	SVT	12 séances de 1h30 (18h)			5 tests (résolution de problèmes, questions ouvertes, définitions)	
3C	101						

Tableau 13: Caractéristiques des études présentées dans le chapitre 6.

¹⁴ Les résidus ne suivant pas une distribution normale, nous avons utilisé une ANOVA non-paramétrique (Kruskal-Willis, voir Delacre et al., 2019).

Les résultats de cette étude indiquent que Jigsaw a eu des effets positifs sur le sentiment d'auto-efficacité des élèves, comparativement à la condition individuelle où le sentiment d'auto-efficacité s'est avéré stable. Toutefois, dans cette étude, les performances n'étaient pas mesurées. Pour revenir à la littérature Jigsaw de manière plus générale, Robert et collaborateurs (*en préparation*) mettent en évidence que plus de la moitié de la recherche est effectuée sur des populations de l'enseignement supérieur, contre 17% de celle menée auprès élèves du secondaire. Sur la méta-analyse présentée dans le chapitre précédent, 5 études sur 20 (i.e., 25% de l'échantillon) étaient réalisées sur une population d'élèves du secondaire et rapportaient aussi bien des effets positifs d'amplitude variée (Law, 2011 ; Sahin, 2011 ; Shaaban, 2006 ; Tarhan et al., 2013) que des effets négatifs (Souvignier & Kronenberger, 2007) de Jigsaw sur l'apprentissage. Ce chapitre a donc pour objectif de mieux cerner l'effet que peut avoir Jigsaw sur cette population.

2.2. Calcul de puissance a priori

Nous avons effectué deux séries d'analyses de puissance *a priori* afin d'estimer le nombre de participant-es requis par condition expérimentale. Ces calculs de puissance ont été effectués en considérant une taille d'effet minimale d'intérêt de $\delta = 0.40$, au seuil $\alpha = .05$. Pour réduire le risque d'erreur de Type II, nous avons choisi de disposer d'une puissance de test de 90%. Les tailles d'échantillons requises pour un test-*t* indépendant (E1 et E2) et pour un test-*t* apparié (E3A, E3B et E3C) sont respectivement de $N_{E1\&E2} = 216$ et $N_{E3A,B,C} = 55$. Compte tenu du contexte de passation expérimentale, c'est-à-dire un contexte de classe « réelle », nous avons systématiquement retenu des échantillons plus larges que ceux prescrits par les analyses de puissance. Dans le cadre des expériences 3A, 3B et 3C, nous avons utilisé un design apparié : les élèves ont bénéficié à la fois de la procédure d'enseignement classique (i.e., « *Teaching as Usual* »), c'est-à-dire telle qu'habituellement mise en place par l'enseignant-e, ainsi que de la méthode Jigsaw (voir Tableau 14). Ce design apparié nous permet de disposer de plus de puissance statistique (Maxwell, 2004 cités dans Montoya, *preprint*). D'autre part, nous avons mesuré le coefficient de corrélation intra-classe (ICC) pour contrôler le caractère hiérarchique des données récoltées (i.e., des élèves inclus dans des classes et des établissements, voir Bressoux, 2020) : pour les cinq expérimentations, ce coefficient de corrélation est bas (*min* = .00 ; *max* = .11 ; *médiane* = .03), ce qui suggère une influence relativement restreinte des regroupements des élèves dans leur classe (Bressoux, 2020).

Notre hypothèse est la suivante :

H₁ : L'effet de Jigsaw sur les performances académiques sera positif et d'une ampleur au moins aussi élevée que $\delta = 0.40$.

Et sera comparé à l'hypothèse nulle :

H₀ : L'effet de Jigsaw sur les performances académiques n'est pas statistiquement différent de zéro ou plus petit que $\delta = 0.40$.

Pour tester cette hypothèse, nous avons systématiquement appliqué la procédure du test d'équivalence (i.e., « Two-One-Sided Tests » ; Lakens, 2018). Notre choix s'est porté sur l'utilisation de test-*t* indépendants et appariés car la procédure de tests d'équivalence n'est pas encore développée pour les analyses de variance avec covariables. Nous avons également regroupé les résultats sous forme de méta-analyse interne, afin de pouvoir rapporter un estimateur moyen et de quantifier les variations de l'effet entre les études (Goh et al., 2016). Nous avons initialement entamé une démarche de préenregistrement des trois premières études (i.e., E1, E2 et E3A) ainsi que d'autres analyses et hypothèses périphériques sur des variables socioaffectives (i.e., sentiment d'auto-efficacité, motivation intrinsèque, sentiment d'appartenance scolaire, voir Annexes D, E et F). Cependant, suite à plusieurs déviations du protocole initialement prévu (i.e., test d'un effet d'interaction de Jigsaw avec le statut socioéconomique plutôt que d'un effet principal), nous n'avons pas préenregistré les études 3B et 3C. L'intégralité des données, du matériel et des procédures est néanmoins disponible sur la page OSF du projet¹⁵. Les analyses périphériques ainsi que les statistiques descriptives des scores figurent dans la partie Annexe (voir Annexes B, L, M et N).

¹⁵ https://osf.io/4pwzy/?view_only=bfa42ad3c076490eace0fb7bc1a137d7

3. Expérience 1

3.1. Participant·es

L'expérience 1 a été menée sur 264 élèves issus de douze classes d'environ 20 élèves chacune ($M = 21.67$, $SD = 3.33$) réparties entre deux établissements. Nous avons supprimé 12 participant·es pour lesquels des données étaient manquantes (i.e., 10 n'ont pas répondu à la totalité du questionnaire et deux n'ont pas répondu au test). Sur les 252 participant·es restant·es, nous avons effectué une analyse de "Median Absolute Deviation" (MAD) pour détecter les valeurs extrêmes ou « *outliers* » ainsi que leur influence sur les données (Leys et al., 2013). Le test n'indique qu'aucun·e participant·e ne présente une variation supérieure à 2.5 fois l'écart absolu médian (critère modérément conservateur), ce qui suggère qu'il n'y avait pas de valeurs extrêmes dans l'échantillon. L'échantillon final est donc composé de 252 participant·es, 147 filles et 105 garçons ($M_{\text{âge}} = 11.14$, $SD = 0.41$, $min = 10$, $max = 13$) réparti·es de manière équilibrée dans les deux conditions expérimentales, $\chi^2_{\text{genre}} < 1$, $p = .98$.

3.2. Procédure

L'expérience s'est déroulée pendant les heures de cours, pour une durée totale de deux heures. Avant le début de l'expérimentation, les classes ont été assignées au hasard à l'une des deux conditions : « Jigsaw » ($N = 127$) ou « individuel » ($N = 125$). Trois expérimentateur·ices formé·es pour donner le cours étaient chargées de la passation expérimentale. La procédure s'est déroulée de la manière suivante : Durant les dix premières minutes, les expérimentateur·rices se sont présenté·es et ont expliqué la procédure générale du cours, notamment la présence d'une évaluation de connaissance à la fin de l'expérimentation. Les enseignant·es, bien que présents lors de l'expérimentation, n'intervenaient pas. Ensuite, les participant·es ont étudié, soit individuellement (i.e., groupe contrôle), soit en groupe Jigsaw (i.e., groupe expérimental). Dans ce dernier, les expérimentateur·rices ont aménagé la salle de classe de manière à rendre le travail de groupe possible en plaçant cinq chaises autour des tables. Ensuite, iels ont appliqué la procédure en trois étapes d'Aronson et Patnoe (2011) :

1. Tout d'abord, les participant·es sont réparti·es au hasard dans des « groupe puzzle » composés d'environ 5 individus. Ensuite, chacun d'elles reçoit une partie différente du contenu pédagogique et pendant les 10 premières minutes (*étape 1*), iels lisent le texte qui leur a été assigné et répondent individuellement à des questions d'entraînement.

2. Ensuite, les expérimentateur·rices ont formé des « groupes d'experts » (*étape 2*) en réunissant les participant·es ayant travaillé sur le même contenu au cours de l'étape 1. Ils disposent alors de 10 minutes pour discuter de leur sujet, vérifier leur compréhension et se préparer à l'expliquer aux membres de leurs groupes.
3. Pour finir, les participant·es ont rejoint leur « groupe puzzle » initial (*étape 3*) et ont dû expliquer à tour de rôle le contenu étudié aux autres membres de leur groupe, pour une durée totale de 20 minutes. Dans le groupe contrôle, les participants ont disposé d'un temps similaire (40 minutes) pour lire et étudier individuellement les trois parties des leçons (textes et questions d'entraînement).

Dans les deux conditions, les participant·es ont bénéficié d'une pause de cinq minutes suite à l'étape trois. Après la pause, ils ont répondu à une évaluation de connaissance, de manière individuelle puis à un questionnaire mesurant plusieurs variables socioaffectives et démographiques (e.g., le sentiment de compétence, le sentiment psychologique d'appartenance à l'école, voir Annexe). Enfin, l'expérimentateur·rice a expliqué le but de l'étude, fait un compte rendu aux participant·es et fourni les corrections de l'évaluation si demandées.

3.3. Matériel

3.3.1. Contenu de la leçon

La leçon porte sur différents concepts mathématiques (e.g., l'arithmétique, la proportionnalité, les diagrammes, les fractions et la géométrie). Elle est composée de plusieurs documents d'environ 550 mots qui comprennent à la fois une leçon et des questions d'entraînement.

3.3.2. Évaluation de connaissance

La performance académique est mesurée via une évaluation de connaissances - c'est-à-dire un test comprenant 17 questions à choix multiples et 8 questions de résolution de problèmes en rapport avec le contenu étudié pendant la phase d'apprentissage. Les questions portent à la fois sur les définitions et la résolution de problèmes tels que présentés dans les textes. Le score obtenu au test d'apprentissage peut aller de 0 (note la plus basse) à 20 (note la plus élevée, $M = 10.92$, $SD = 4.20$, $min = 0.80$, $max = 19.60$) et est corrigé par l'expérimentateur·rice en charge de l'étude, aveugle à la condition expérimentale.

3.3.3. Niveau scolaire initial

Le niveau scolaire initial des participant·es est mesuré par leur note trimestrielle en mathématiques. Cette note peut aller de 0 (note la plus basse) à 20 (note la plus élevée, $M = 13.85$, $SD = 3.84$, $min = 3.20$, $max = 20$) et a été obtenue avant la mise en œuvre de l'expérience par le directeur ou la directrice de l'établissement

3.4. Résultats

3.4.1. Analyses préliminaires

Les corrélations entre les mesures de niveau scolaire initial et la performance académique à l'évaluation sont présentées dans la partie Annexe (voir Annexe A). Pour vérifier d'éventuelles différences entre les groupes expérimentaux, nous avons effectué un test- t indépendant sur le niveau scolaire initial et n'avons pas observé de différences statistiquement significatives entre les groupes ($M_{différence} = 0.01$, $SE = 0,48$, $95\% \text{ CI} = [-0.97, 0.94]$), $t(250) = 0.03$, $p = .98$. Le test de normalité soutient partiellement l'hypothèse d'une distribution normale des résidus (Shapiro-Wilk = .99, $p = .01$, Kolmogorov-Smirnov = .05, $p = .60$) et le test de Levene d'égalité des variances suggère une variance homogène entre les groupes, $F(1, 250) = 1.41$, $p = .24$.

3.4.2. Effet de Jigsaw sur la performance académique (E1)

Nous avons utilisé un t -test indépendant, ainsi qu'un test d'équivalence (Lakens et al., 2018) pour comparer les scores des deux groupes sur l'évaluation de connaissances vues pendant la phase d'apprentissage¹⁶. Un second modèle statistique avec covariable a été testé afin de vérifier que l'inclusion du niveau initial dans les analyses ne faisait pas varier significativement les résultats du t -test. Les résultats de ce test figurent en pied de page dans les parties correspondantes. Contrairement à nos attentes, les participant·es de la condition « Jigsaw » ($M = 10.83$, $SD = 4.28$) n'obtiennent pas de meilleurs résultats que les participant·es de la condition « individuel » ($M = 11.01$, $SD = 4.13$), $t(250) < 1$, $p = .62$, $d = 0.04$. L'équivalence du résultat est ensuite vérifiée en suivant la procédure du "Two One-Sided Test" présentée dans le chapitre 4. Les résultats indiquent que la taille de l'effet observée peut être considérée comme statistiquement équivalente à zéro ($p_{upper} = .004$ et $p_{lower} < .001$

¹⁶ Dans les analyses incluant le covarié « niveau initial », l'effet principal de Jigsaw n'est pas statistiquement significatif $F(1, 247) = 0.56$, $p = .46$, $\eta^2 = .002$. La covariable « niveau scolaire initial » prédit significativement la performance académique, $F(1, 247) = 141.75$, $p < .001$, $\eta^2 = .37$, voir Annexe H.

respectivement). Dit autrement, l'effet de Jigsaw n'est pas statistiquement différent de zéro et n'est pas au moins aussi élevé que $\delta = 0.40$.

3.5. Discussion intermédiaire (E1)

Dans cette première expérience, nous nous attendions à ce que les participant·es en condition Jigsaw réussissent mieux l'évaluation de connaissance que ceux en condition de travail individuel. Cependant, nos résultats ne supportent pas empiriquement cette affirmation : les différences de scores observées ne sont pas statistiquement significatives, et leur amplitude est proche du zéro. Toutefois, plusieurs limitations peuvent être notées. Tout d'abord, la procédure de répartition des participant·es a eu lieu au niveau des classes entières. Même si nous avons estimé qu'en moyenne nous n'observions pas de différences de niveau scolaire initial entre les deux groupes on ne peut pas totalement exclure que l'absence de différences significatives entre les deux groupes en matière d'apprentissage soit due aux caractéristiques des classes impliquées dans chaque condition (i.e., nous obtenons un coefficient de corrélation intraclasse de .113 pour cette expérimentation). D'autre part, des discussions informelles avec les enseignant·es ont fait apparaître que certaines parties du matériel d'apprentissage utilisées dans le cadre de l'expérience avaient pu être partiellement enseignées dans certaines classes, mais pas dans d'autres. Dans l'expérimentation suivante, deux modifications ont été apportées pour corriger ces problèmes. Premièrement, l'aléatorisation a été effectuée par demi-classe, c'est-à-dire à partir de groupes créés en mélangeant les classes existantes et deuxièmement, le matériel d'apprentissage a été construit de manière à traiter un contenu qui ne constitue pas une partie du programme des élèves de sixième.

4. Expérience 2

4.1. Participant·es

Les directeurs de deux collèges - différents de ceux ayant participé à l'expérience 1 - ont accepté de participer. Nous avons recruté quatorze classes d'environ 22 élèves chacune ($M = 22.57$, $SD = 4.18$) pour un échantillon total de 319 participant·es. Nous avons supprimé six participant·es pour cause de données manquantes (i.e., trois questionnaires non remplis et trois variables démographiques non renseignées). Comme dans l'expérience 1 (E1), nous avons utilisé la procédure « MAD » pour détecter les valeurs extrêmes sur le score d'apprentissage. En suivant le même critère, le test a révélé que deux participant·es avaient des scores qui variaient de plus de 2.5 fois par rapport à la médiane absolue. Cependant, ces participant·es ont été conservé·es de l'analyse, car ils n'influençaient pas les données de manière significative (distance de Cook $< .03$). Ainsi, l'échantillon final était composé de 313 participant·es, 169 filles, 144 garçons ($M_{\text{âge}} = 11.60$; $SD = 0.56$) réparti·es de manière égale entre les deux conditions, $\chi^2_{\text{genre}} = 1.91$, $p = .17$.

4.2. Procédure

La procédure est similaire à celle de l'E1, avec quelques ajustements mineurs. Tout d'abord, un seul expérimentateur est en charge des passations, contre trois dans l'expérience 1. Ensuite, la répartition des participant·es dans les groupes expérimentaux est établie à partir de demi-classes. Ainsi la variable « classe » n'est pas confondue avec la condition expérimentale ($ICC = .013$).

4.3. Matériel

4.3.1. Contenu de la leçon

La leçon porte sur "Le cycle du sommeil", un sujet axé sur les sciences de la vie et de la Terre (SVT). Il consiste en un texte de six pages sur le sommeil, contenant plusieurs images et des questions guides. Ce texte est divisé en trois sous-sujets d'environ 400 mots chacun et provient d'un site web gratuit et reproductible, accessible à l'adresse suivante :

http://lecerveau.mcgill.ca/flash/d/d_11/d_11_p/d_11_p_cyc/d_11_p_cyc.html.

4.3.2. Évaluation de connaissances

L'apprentissage est mesuré via une évaluation de connaissances composée de neuf questions à choix multiple avec trois réponses possibles et deux questions ouvertes portant sur l'ensemble du contenu d'apprentissage, c'est-à-dire les trois sous-sections ($M = 3.99$; $SD = 1.39$; $min = 1.00$; $max = 8.50$). Les scores peuvent aller de 0 (score minimal) à 8.50 (score maximal), et sont corrigés par l'expérimentateur, aveugle aux conditions expérimentales.

4.3.3. Niveau scolaire initial

Comme pour l'expérience 1 (E1), le niveau scolaire initial est mesuré par la note moyenne trimestrielle des élèves en sciences de la vie et de la Terre ($M = 14.03$; $SD = 2.80$, $min = 5.00$, $max = 20.00$).

4.4. Résultats

4.4.1. Analyses préliminaires

Nous avons effectué les mêmes vérifications que pour l'expérimentation 1 : nous ne nous attendions pas à observer des différences significatives entre les deux conditions avant le début de l'expérience. Pour vérifier cette hypothèse, nous avons effectué un test- t indépendant sur le niveau scolaire initial. Les résultats indiquent une différence faible en faveur du groupe « Jigsaw » (M différence = 0.49, $SE = 0.32$, $95\%CI = [-1.11 ; 0.13]$, $d = 0.17$), même si elle ne dépasse le seuil de rejet $\alpha = .05$, $t(311) = 1.54$, $p = .12$. Le test de normalité rejette l'hypothèse d'une distribution normale des résidus (Shapiro-Wilk = .97, $p < .001$, Kolmogorov-Smirnov = .11, $p < .001$) et le test d'égalité des variances de Levene suggère une variance homogène entre les groupes, $F(1, 313) = 2.60$, $p = .11$. Par conséquent, nous avons testé notre hypothèse à l'aide d'une ANOVA non paramétrique (Kruskal-Willis).

4.4.2. Effet de Jigsaw sur la performance académique (E2)

Contrairement aux attentes, les participants dans la condition « Jigsaw » ($M = 3.99$, $SD = 1.44$) n'obtiennent pas de meilleurs résultats que les participants de la condition « individuel » ($M = 3.99$, $SD = 1.35$), $\chi^2 < 1$, $p = .92$ ¹⁷. Le test d'équivalence confirme que la taille de l'effet observée n'est pas statistiquement différente de zéro, $t(311) < 1$, $p = .99$ (p_{upper} et $p_{lower} < .001$).

¹⁷ L'effet principal de Jigsaw n'est pas statistiquement significatif $F(1, 308) = .29$, $p = .59$, $\eta^2 = .001$. La covariable « niveau scolaire initial » prédit significativement la performance académique, $F(1, 308) = 34.75$, $p < .001$, $\eta^2 = .101$, voir Annexe I.

4.5. Discussion intermédiaire (E2)

Dans l'expérience 2, la procédure d'aléatorisation et la nouveauté du matériel d'apprentissage ont été améliorées, comparativement à la première expérience. Le coefficient de corrélation intra-classe a diminué entre nos deux expérimentations, ce qui soutient notre démarche d'aléatorisation par demi-classe. Cependant, nous n'observons toujours pas de support en faveur de notre hypothèse de recherche : à travers ces deux premières études, les différences de moyennes observées sont proches de zéro, et l'estimation de taille d'effet loin d'atteindre notre seuil d'intérêt fixé à $\delta = 0.40$. On ne peut donc raisonnablement pas conclure, sur la base de ces deux premiers résultats, que Jigsaw a eu les effets escomptés sur les apprentissages.

Néanmoins, une limite commune à ces deux expériences est que les conditions d'expérimentations sont très différentes des conditions d'apprentissage habituelles en classe réelle : le temps consacré à l'apprentissage du contenu expérimental est relativement court et le contexte inhabituel pour des élèves de collèges. Par ailleurs, le cadre des expériences 1 et 2 peut manquer d'intérêt pédagogique ou de pertinence pour les participant·es (Topping et al., 2017). Dans la littérature portant sur l'apprentissage coopératif, la plupart des chercheurs suggèrent que la structuration du travail en groupe, en particulier avec Jigsaw, nécessite plusieurs sessions avant d'être mis en œuvre avec succès en raison de sa structure complexe (Roseth et al., 2018). Par ailleurs, le sentiment de cohésion sociale, condition *sine qua non* de la théorie de l'interdépendance sociale positive, n'a peut-être pas émergé de manière systématique dans tous les groupes. Nos analyses périphériques contribuent à soutenir cet argument, puisque nous n'observons pas de différences en faveur de Jigsaw sur les variables socioaffectives supposées être améliorées par la coopération (e.g., estime de soi, motivation intrinsèque, sentiment d'appartenance scolaire, voir Annexes C et D pour les questionnaires, et Annexe G pour une synthèse des résultats).

Si la pertinence de l'objectif commun (i.e., être responsable de l'apprentissage de l'autre) n'a pas été perçue comme particulièrement importante dans nos passations expérimentales, alors il est possible que les élèves n'aient coopéré de manière efficace pour atteindre ce but (Johnson & Johnson, 2009). Pour remédier à ces limites, les expériences 3A, 3B et 3C ont été menées dans des contextes de classe réelle et avec des enseignant·es volontaires. Dans ce scénario, le contexte expérimental est supposément plus fiable que celui d'une session courte pilotée par des expérimentateur·ices externes à l'établissement. De plus, la durée de l'expérimentation est accrue, ainsi que le nombre de mesures des apprentissages, ce qui accroît

la cohérence des résultats. Les mesures sont également directement reliées à la véritable réussite académique des élèves, puisqu'elles constituent leurs moyennes dans la discipline. Pour finir, les enseignant·es qui participent à ces expérimentations ont été accompagnés par l'expérimentateur pour adapter leur scénario pédagogique de classe selon la méthode Jigsaw.

5. Expérience 3A

5.1. Participant·es

Un professeur de physique-chimie en charge de six classes d'élèves s'est porté volontaire pour participer à l'étude 3A. L'échantillon de l'expérience est composé de ses 122 élèves ($N = 20.33$ par classe, $SD = 2.67$). L'expérience s'est déroulée sur les créneaux de cours habituels de physique-chimie à raison d'une heure par semaine, pendant 16 semaines. Douze participants ont été supprimés des analyses en raison de données manquantes (i.e., 10 variables démographiques et des absences sur plus de $\frac{3}{4}$ des évaluations) et d'une influence potentielle sur les données ($MAD > 2.50$ et distance de Cook $> .05$). L'échantillon final est composé de 110 participant·es (52 filles et 58 garçons) réparti·es de manière équilibrée dans les deux conditions, $\chi^2_{\text{genre}} < 1, p = .69$.

5.2. Procédure

Dans la première phase de l'expérience (de janvier à mars 2019, séquences 1 et 2), la moitié des classes est assignée dans la condition « Jigsaw » ($N = 53$), tandis que l'autre moitié ($N = 57$) est assignée dans la condition « TAU » (i.e., Teaching as Usual). Après deux semaines de vacances scolaires, les groupes sont contrebalancés pour les séquences 3 et 4 (de mars à mai 2019). Les participant·es qui ont d'abord étudié en condition « Jigsaw » passent en condition « TAU » et réciproquement (voir le Tableau 14 ci-dessous pour un résumé de la procédure).

		Étape 1 : Semaines 1 à 6		Étape 2 : Semaines 7 à 16	
		Évaluation 1	Évaluation 2	Évaluation 3	Évaluation 4
Classes	2, 3, 6	Jigsaw (<i>expérimental</i>)		TAU (<i>contrôle</i>)	
	1, 4, 5	TAU (<i>contrôle</i>)		Jigsaw (<i>expérimental</i>)	

Tableau 14: Distribution des classes selon le design expérimental (E3A)

L'enseignant, qui a reçu une formation sur la méthode Jigsaw (i.e., adaptation des contenus de cours, présentation de la structure des étapes), est responsable de l'enseignement des six classes. En effet, les enseignant·es de disciplines scientifiques de l'académie Clermontoise participent à des « réunions de bassin », pilotées par le rectorat et les inspecteur·rices de l'académie, dans lesquelles plusieurs formations sont dispensées. Ayant participé à ces formations, l'enseignant·e en charge de cette expérimentation avait déjà expérimenté une pédagogie coopérative basée sur l'interdépendance de ressources. Il a également mis en place, au sein d'autres classes d'élèves, une pédagogie coopérative de type « Team-Game Tournament » (Sharan, 1980 ; Slavin, 1983). Pour cette expérimentation, les contenus pédagogiques que l'enseignant a présentés à ses élèves ont été supervisés par l'expérimentateur, afin de s'assurer qu'ils correspondaient bien à la structure en trois étapes de Jigsaw (chapitre 2). Avant le début de l'expérience, chacune des classes est assignée au hasard à une condition Jigsaw ou d'enseignement habituel (i.e., « Teaching as Usual », « TAU »). Les participant·es ont étudié exactement le même matériel d'apprentissage, mais dans la condition Jigsaw, l'enseignant a suivi la procédure de Jigsaw en trois étapes telle que développée par Aronson et collaborateurs (voir chapitre 2) tandis que dans la condition « TAU », l'enseignant s'est engagé à appliquer sa pratique habituelle. Cette pratique repose essentiellement sur un enseignement « classe entière », mais également sur des travaux de groupes en binômes dans le cadre de manipulations et de travaux pratiques. À la fin de chaque séquence, les élèves remplissent une évaluation qui contribue à leur moyenne dans la matière « physique-chimie » de leur parcours scolaire.

5.3. Matériel

5.3.1. Contenu des leçons

Les leçons portent sur le thème « Matière, mouvement, énergie et information ». Les principaux sujets du thème sont présentés en Annexe. Le thème est divisé en séquences couvrant plusieurs leçons ayant une structure similaire. Tout d'abord, l'enseignant présente la séquence et les élèves rédigent des hypothèses ou prédictions sur un sujet (e.g., séquence 3 : « *Quelle quantité de sucre peut se dissoudre dans l'eau ?* »). Ensuite, iels mettent en place une expérience avec des manipulations et des observations pour tester cette prédiction (e.g., « *Mélanger 10 g de sucre avec 20 cl d'eau* »). Enfin, iels répondent à une liste de questions afin de communiquer les résultats à l'enseignant qui apporte une correction (e.g., « *Le sucre (10g) est-il entièrement dissous dans l'eau pour le groupe 3 ?* »)

En condition « Jigsaw », le contenu de la leçon et l'hypothèse associée sont organisés en plusieurs activités. Ces activités sont liées au thème principal et consistent en plusieurs exemples pratiques : pour la séquence 3 sur « la masse et la matière », les participant·es doivent apprendre la notion de « conservation de masses ». Pour ce faire, l'enseignant met en place quatre activités où les enfants doivent mélanger plusieurs contenus (e.g., du sucre, du sable, du sel ou du vinaigre) dans de l'eau, mesurer la masse avec les outils correspondants et communiquer leurs résultats sous la forme d'un test d'hypothèse. De manière générale, la condition Jigsaw suit la procédure suivante :

1. Dans le déroulement de la séance, la première étape de la procédure Jigsaw comprend un temps individuel dans des groupes créés par l'enseignant pour permettre aux participants de découvrir le sujet, lire les instructions et identifier les autres membres de leur groupe.
2. Ensuite, dans la phase de « groupes d'experts », les participants testent l'hypothèse générée en phase individuelle et mettent en œuvre une expérience en suivant des indications communiquées par l'enseignant.
3. Dans la dernière étape de la procédure Jigsaw, les participants mettent en commun leurs observations et présentent leurs résultats aux autres afin de comprendre les différences et les similitudes de leurs expériences et de reconstituer une synthèse du thème général présenté.

5.3.2. Évaluations de connaissances

À la fin de chaque séquence, les élèves ont passé des évaluations de connaissances. Ces évaluations consistent en des définitions, des exercices de résolution de problèmes des schémas à rédiger et couvrent les différentes expérimentations abordées pendant les leçons. Les deux premières évaluations correspondent à la première partie de l'expérimentation et les deux autres à la deuxième partie et sont notées, par l'enseignant de 0 (la note la plus basse) à 20 (la note la plus élevée). La moyenne de ces quatre notes constitue la note trimestrielle réelle des participant·es et le détail de chaque score est présenté dans l'Annexe B. Afin de réduire l'hétérogénéité des notes, nous avons calculé la note moyenne obtenue dans la condition « Jigsaw » (notée « xJigsaw ») et celle dans la condition d'enseignement habituel « TAU » (notée « xTAU »). En conséquence, chaque participant·e dispose d'un score moyen pour les tests d'apprentissage couverts pendant la période « Jigsaw » et la période « xTAU », ce qui nous permet d'utiliser un test-*t* apparié. Par exemple, le score moyen « xJigsaw » correspond à la moyenne des deux premiers tests pour les participant·es qui ont commencé l'expérience assignée dans cette condition. Pour ces participant·es, le score "xTAU" est calculé en agrégeant

leurs scores pour le troisième et quatrième test. Le score obtenu dans l'une des conditions se comporte donc comme une variable intra-sujet.

5.3.3. Niveau scolaire initial

La note moyenne trimestrielle des élèves en physique-chimie, pouvant aller de 0 à 20 ($M = 13.07$, $SD = 3.53$, $min = 4.38$, $max = 19.25$), constitue la mesure de niveau scolaire initial dans cette expérience

5.4. Résultats

5.4.1. Analyses préliminaires

Un t -test indépendant indique que les différences initiales entre les deux groupes ne sont pas statistiquement significatives au début de l'expérimentation, $M_{différence} = 0.18$, $95\%CI = [-1.52, 1.15]$, $SE = 0.67$, $t(108) < 1$, $p = .79$. Dit autrement, les deux regroupements de classes ne diffèrent pas statistiquement au début de l'expérimentation. Le coefficient de corrélation intra-classe est égal à $.027$ pour cette expérimentation. Le test de normalité indique une distribution normale des résidus (Shapiro-Wilk = $.99$, $p = .91$, Kolmogorov-Smirnov = $.04$, $p = .99$).

5.4.2. Effet de Jigsaw sur les performances académiques (E3A)

Dans cette expérience, les participants obtiennent deux scores d'apprentissage : Un pour les tests effectués pendant l'enseignement « Jigsaw » et un autre pour ceux couverts en « TAU ». Par conséquent, ces scores ont été utilisés comme une variable intra-sujet et notre hypothèse a été testée avec un t -test apparié. Les résultats indiquent une fois de plus que la performance aux évaluations de connaissances n'est pas plus élevée lorsque les participants sont dans la condition « Jigsaw » ($M = 10.78$, $SD = 3.58$) par rapport à la condition « TAU » ($M = 11.15$, $SD = 3.34$), $t(110) < 1$, $p = .78$, $d = -0.07$. Le test d'équivalence indique que la taille d'effet observée sur ces scores n'est pas différente de zéro¹⁸.

¹⁸ Dans les analyses incluant le covarié « niveau initial », l'effet principal de Jigsaw (en variable intra) n'est pas statistiquement significatif $F(1, 107) = .66$, $p = .42$, $\eta^2 = .006$. La covariable « niveau initial » (en variable inter-sujet) prédit significativement la performance académique, $F(1, 107) = 74.35$, $p < .001$, $\eta^2 = .410$, voir Annexe J.

6. Expérience 3B

6.1. Participant·es

L'échantillon est composé d'une enseignante et de ses 85 élèves répartis dans quatre classes ($N = 21.25$ par classe, $SD = 0.75$) qui se réunissent une fois par semaine pour un cours d'une heure et demie, pendant 16 semaines. Dix participant·es sont supprimé·es des analyses pour données manquantes (i.e., neuf variables démographiques et des absences sur plus de $\frac{3}{4}$ des tests d'apprentissage) et un participant qui présente des valeurs extrêmes ($MAD = 2.50$, distance de Cook = 0.12). L'échantillon final est composé de 74 participants (35 filles, 36 garçons et 3 non renseignés) réparti·es de manière égale dans les deux conditions, $\chi^2_{\text{genre}} < 1, p = .91$.

6.2. Procédure

Nous avons utilisé le même modèle que pour l'expérience 3A (voir Tableau 14).

6.3. Matériel

6.3.1. Contenu des leçons

Le contenu pédagogique correspond au thème « Environnement et Biodiversité ». Il est divisé en quatre séquences couvrant plusieurs aspects du thème. Comme pour l'E3A, chaque séquence est composée de plusieurs activités et une évaluation de connaissance est donnée aux participants une fois la séquence terminée. La note obtenue à cette évaluation contribue à la moyenne générale des participant·es dans la discipline.

Le cours suit globalement une structure proche de celle présentée dans l'E3A. Tout d'abord, l'enseignante présente la séquence ainsi qu'une problématique générale au niveau de toute la classe. Par exemple, dans la séquence 2, la problématique est la suivante « *Comment expliquer la répartition des espèces dans leur environnement ?* ». Sur la base de cette problématique, les participant·es doivent concevoir des hypothèses puis effectuer une activité (e.g., étudier le comportement des animaux en fonction des saisons) qui peut être réalisée en suivant la procédure « Jigsaw » ou bien l'enseignement habituel « TAU ». Dans le cas du « Jigsaw », différentes activités sont distribuées aux collégiens pour créer des « groupes Jigsaw ».

1. La première étape consiste alors à étudier individuellement un exemple précis (e.g., dans la séquence 2, un certain animal : cerf, sauterelle, serpent ou oiseau).
2. Ensuite, les enfants se réunissent en « groupes d'experts » pour répondre à un ensemble questions portant sur leurs exemples (e.g., "Qu'est-ce qui provoque un tel comportement chez le serpent en hiver ?).

3. Dans la dernière étape, les participants des « groupes puzzle » accomplissent une tâche où tous les exemples sont nécessaires (e.g., remplir un tableau avec l'animal correspondant et ses caractéristiques).

Dans la condition « TAU », cette coopération n'est pas mise en place : les élèves étudient individuellement les mêmes exemples que ceux utilisés pour constituer les « groupes Jigsaw » dans l'autre condition. Il n'y a donc pas d'interdépendance positive entre les participant·es.

6.3.2. Évaluation de connaissances

Comme pour l'expérience 3A, nous avons calculé un score moyen pour les notes obtenues aux évaluations de connaissance obtenues dans la condition « Jigsaw » et celles obtenues dans la condition « TAU ». Le détail de chaque évaluation figure dans l'Annexe B.

6.3.3. Niveau scolaire initial

L'expérience ayant eu lieu au cours du premier trimestre, nous n'avons pas disposé des notes préalables en SVT. À défaut, nous avons estimé un score de performance antérieure plus général en agréant les notes en français et en mathématiques obtenues par les participants en entrée de collège ($M = 12.49$, $SD = 3.13$, $min = 4.37$, $max = 19.32$).

6.4. Résultats

6.4.1. Analyses préliminaires

Un t -test indépendant indique que les différences initiales entre les deux groupes ne sont pas statistiquement significatives, $M_{différence} = 0.35$, 95%CI [-1.11, 1.81], $SE = 0.73$, $t(72) < 1$, $p = .63$. Le coefficient de corrélation intra-classe est assez faible ($ICC = .052$). La distribution des résidus suit une loi normale (Shapiro-Wilk = .99, $p = .84$, Kolmogorov-Smirnov = .09, $p = .53$).

6.4.2. Effet de Jigsaw sur les performances académiques (E3B)

Nous n'observons pas de différences statistiquement significatives entre les scores moyens obtenus lorsque les participant·es sont en condition « Jigsaw » ($M = 15.75$, $SD = 2.54$) ou en condition « TAU » ($M = 15.63$, $SD = 2.81$), $t(73) < 1$, $p = .35$, $d = 0.05$ ¹⁹. Le test d'équivalence confirme que la taille d'observée est équivalente à zéro et non statistiquement significative. ($p_{upper} < .001$ et $p_{lower} < .001$ respectivement).

¹⁹ Dans les analyses incluant le covarié « niveau initial », l'effet principal de « Jigsaw » (variable intra-sujet) n'est pas statistiquement significatif $F(1, 68) = .34$, $p = .57$, $\eta^2 = .005$. La covariable de niveau initial (variable inter-sujet) prédit significativement la performance académique, $F(1, 68) = 31.65$, $p < .001$, $\eta^2 = .318$, voir Annexe K.

7. Expérience 3C

7.1. Participants

L'échantillon est composé de 108 élèves et de leurs deux enseignant·es, réparti·es dans cinq classes ($N = 21.80$ par classe, $SD = 1.52$). Les cours ont eu lieu une fois par semaine pour pendant 16 semaines pour une durée d'1h30 par séance. Nous avons supprimé six participant·es des analyses en raison de données manquantes (i.e., absence sur les $\frac{3}{4}$ des évaluations) ainsi qu'un participant présentant une influence possible sur les données ($MAD = 2.70$, distance de Cook = .10). Ainsi, l'échantillon final est composé de 101 participant·es (55 filles, 44 garçons et 2 non rapportés) réparti·es de manière égale dans les deux conditions, $\chi^2_{\text{genre}} = 2.43$, $p = .12$.

7.2. Procédure

Nous avons suivi la même procédure que pour les expériences 3A et 3B avec trois différences mineures.

1. Premièrement, nous n'avons pas obtenu l'accord de l'école pour recueillir des données sur le niveau scolaire initial des enfants. Par conséquent, nous n'avons pas pu vérifier les différences entre groupes avant l'expérience.
2. Deuxièmement, deux enseignant·es ont participé à l'expérience et étaient responsables respectivement de trois et deux classes chacun·e, contre un enseignant pour l'expérimentation 3A et une enseignante pour l'expérimentation 3B.
3. Enfin, ces enseignant·es ont eu suffisamment de temps pour couvrir cinq séquences, au lieu de quatre dans les expériences 3A et 3B, ce qui nous a permis d'obtenir une cinquième mesure de performance académique.

7.3. Matériel

7.3.1. Contenu des leçons

Le matériel d'apprentissage (c'est-à-dire les activités pédagogiques, les tests d'apprentissage) ainsi que le déroulement de l'enseignement sont quasiment identiques à ceux présentés dans l'E3B : chaque séquence correspond à plusieurs activités articulées autour d'une problématique issue d'un thème. Dans la condition « Jigsaw », ces activités sont séparées en différentes sous-parties, chacune d'entre elles explorant une partie spécifique du sujet. Par exemple, dans cette expérience, la première séquence porte sur l'identification de différentes feuilles d'arbres.

1. Les « groupe puzzle » sont donc formés avec des participant·es qui étudient individuellement différents exemples de feuilles (e.g., érable, chêne, orme, bouleau) dans la première étape.
2. Ensuite, les « experts » analysent les caractéristiques spécifiques de ces exemples (e.g., la disposition des feuilles sur la branche, le contour et la netteté des feuilles, la composition des feuilles) dans un nouveau groupe.
3. Pour finir, de retour dans les « groupe puzzle », les participant·es décrivent leurs exemples en suivant les informations échangées lors de la « phase expert ».

Dans la condition d'enseignement habituel « TAU », les participant·es ont étudié des exemples identiques, mais de manière individuelle, sans structuration de l'interdépendance positive.

7.3.2. Évaluation de connaissance

Trois tests ont évalué l'apprentissage des participants sur le premier thème « Environnement et Biodiversité », et deux autres sur le thème : « De la graine à la plante ». Les tests d'apprentissage sont assez similaires à ceux utilisés dans les expériences 3A et 3B (c'est-à-dire des définitions, des exercices de résolution de problèmes et des schémas), et sont notés de 0 (la note la plus basse) à 20 (la note la plus élevée) par les enseignants responsables de leurs classes. Les scores « xJigsaw » et « xTAU » ont été calculés en suivant la même procédure que dans les expériences 3A et 3B et le détail de chaque évaluation figure dans l'Annexe B.

7.4. Résultats

7.4.1. Analyses préliminaires.

Comme nous l'avons développé ci-dessus, nous n'avons pas pu vérifier les différences initiales de résultats scolaires entre les classes. Le coefficient de corrélation intra-classe est relativement faible ($ICC = .001$). Néanmoins, le test de normalité sur les scores « xJigsaw » et « xTAU » suivent une distribution normale des résidus (Shapiro-Wilk = .99, $p = .35$, Kolmogorov-Smirnov = .07, $p = .74$).

7.4.2. Effet de Jigsaw sur les performances académiques (E3C)

Les performances académiques ne sont pas plus élevées pour les séances réalisées en « Jigsaw » ($M = 15.51$, $SD = 3.13$) par rapport aux séances réalisées en « TAU » ($M = 15.33$, $SD = 3.31$), $t(100) < 1$, $p = .29$, $d = 0.05$ ²⁰. Le test d'équivalence suggère que la taille de l'effet observée est statistiquement équivalente à zéro.

²⁰Dans les analyses incluant le covarié « niveau initial », l'effet principal de « Jigsaw » (variable intra-sujet) n'est pas statistiquement significatif $F(1, 97) = .38$, $p = .54$, $\eta^2 = .004$, voir Annexe L.

7.5. Discussion intermédiaire (E3A, E3B et E3C)

Dans l'ensemble, nous n'avons pas trouvé de soutien à l'hypothèse selon laquelle l'enseignement en Jigsaw augmente l'apprentissage par rapport à une condition d'enseignement habituel (voir le Tableau 15 pour une synthèse). Même dans un environnement de classe réelle, pendant une durée d'un trimestre et avec l'enseignant·e habituellement en charge du cours, les différences de score ($M_{\text{Jigsaw-TAU}}$) sont extrêmement faibles, d'où les estimations de taille d'effet proches de zéro. Les résultats globaux sont présentés de manière plus détaillée dans la partie suivante, à l'aide d'une méta-analyse interne (Goh et al., 2016).

Exp.	<i>N</i>	Corrélation (<i>xJigsaw;xTAU</i>)	$M_{\text{Jigsaw-TAU}}$ (<i>SE</i>)	<i>t-value</i>	<i>p-value</i>	Cohen's <i>d</i>
3A	110	-.12 ^{ns}	-0.38 (0.49)	0.77	.78	-.07
3B	74	.50 ^{***}	0.12 (0.31)	0.39	.50	.05
3C	101	.45 ^{***}	0.19 (0.34)	0.55	.29	.05

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Tableau 15: Synthèse des résultats (expériences 3A, 3B et 3C)

8. Méta-analyse interne

Nous avons effectué une méta-analyse interne afin d'obtenir un effet global de Jigsaw sur les performances académiques et de quantifier les différences entre études (Borenstein et al., 2010 ; Goh et al., 2016). Nous avons effectué nos analyses avec JAMOVI 1.1.8.0 (extension « MAJOR ») en suivant la méthode des moments de DerSimonian-Laird pour la pondération des effets (voir chapitre 4). L'analyse montre que l'effet global de Jigsaw sur l'apprentissage n'est pas statistiquement significatif et non différent de zéro, $g^* = 0.01$, 95% CI [-0.10, 0.11], $SE = 0.05$, $Z = .11$, $p = .91$. Les statistiques d'hétérogénéité entre les études (Q (df = 4) = 0.78, $p = .94$) et à l'intérieur des études ($T^2 = .00$, $SE = .01$), sont très faibles : ces résultats indiquent que les cinq expériences partagent une taille d'effet commune proche de zéro et que la variation entre les tailles d'effet est due à de l'erreur aléatoire. Ce résultat nous amène à conserver l'hypothèse nulle avec une confiance relativement importante²¹ : Jigsaw n'a pas amélioré la

²¹ La probabilité a posteriori d'obtenir cinq résultats non statistiquement significatifs avec une puissance de test de 90%, si H_1 est vraie, est inférieure à $P = 10^{-6}$. À l'inverse, si H_0 est vraie, cette probabilité est de 77%. En termes de ratio, il y a 77 000 plus de chances d'observer de tels résultats sous l'hypothèse nulle H_0 que sous l'hypothèse alternative H_1 (http://shiny.ieis.tue.nl/mixed_results_likelihood/)

performance académique des élèves, au moins dans une mesure jugée « d'intérêt pratique » fixée par le seuil de $\delta = 0.40$ (Hattie, 2009).

De plus, nous n'observons aucune différence entre les études menées sur une courte durée ($ES = -0.02$, 95% CI [-0.18, 0.15], $SE = .08$, $Z = .23$, $p = .81$) et celles menées le long d'un trimestre, dans un contexte de classe réelle ($ES = -0.04$, 95%CI [-0.17, 0.25], $SE = .10$, $Z = .38$, $p = .70$), Q_w ($df = 1$) < 1 . On peut comparer ces résultats à la méta-analyse effectuée dans le chapitre 4 puisque, dans cette dernière, les effets de la durée n'étaient pas non plus statistiquement significatifs. D'après la méta-analyse du chapitre 4, les effets de Jigsaw dépendent principalement de la localisation de l'étude plutôt que d'autres indicateurs, comme la durée, la population ou la discipline étudiée. Dans ce chapitre, nous observons une taille d'effet moyenne similaire à celle observée sur les études localisées en occident (i.e., Europe, États-Unis), c'est-à-dire une taille d'effet proche de zéro (voir Figure 15).

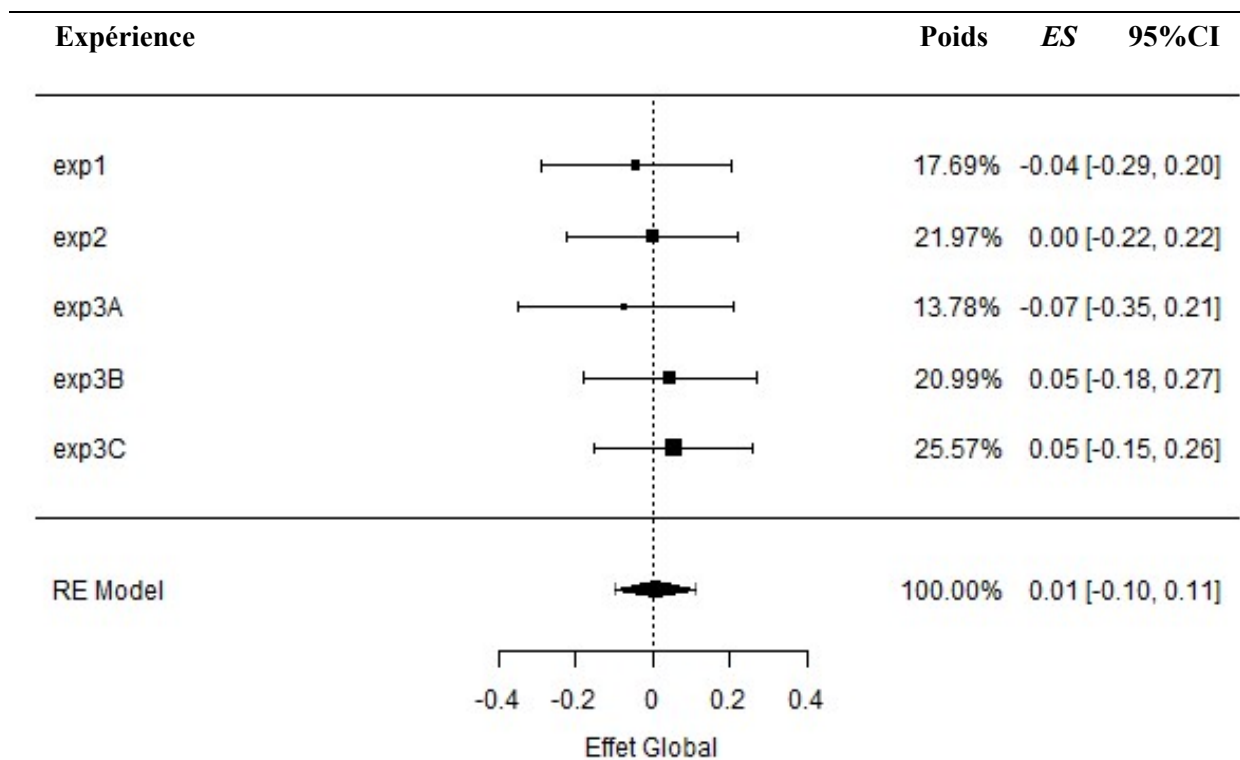


Figure 15: Forest-plot de tailles d'effet observées dans les expériences E1 à E3C

9. Discussion générale

9.1. Synthèse des principaux résultats

L'objectif de ce chapitre était de tester l'efficacité de Jigsaw sur l'apprentissage des élèves de sixième, dans le cadre de cinq expériences. Dans les expériences 1 et 2, nous avons comparé le score obtenu à une évaluation de connaissances entre des participant·es qui ont suivi la procédure Jigsaw à des groupes où les participant·es ont travaillé sur un contenu pédagogique identique, mais de manière individuelle. Dans les expériences 3A, 3B et 3C, des élèves travaillaient une partie de leur cours en Jigsaw et une autre partie dans des conditions d'enseignement habituelles. Nous avons testé l'hypothèse selon laquelle Jigsaw améliore l'apprentissage tel que mesuré par des évaluations de connaissances, c'est-à-dire des questions de compréhension sur les contenus pédagogiques. À travers ces cinq expériences, nous ne constatons aucun support empirique à cette hypothèse. Les tailles d'effet observées et l'estimation globale obtenues grâce à une méta-analyse interne des études sont systématiquement plus petites que notre plus petite taille d'effet d'intérêt (i.e., $\delta = 0.40$). La méta-analyse interne rapporte une taille d'effet moyenne de zéro, ainsi qu'une dispersion nulle. Autrement dit, la probabilité *a posteriori* de l'hypothèse d'un effet positif Jigsaw et d'ampleur proche de $\delta = 0.40$ est extrêmement faible.

9.2. Contributions et limites

Les études reportées dans ce chapitre permettent de combler l'absence relative de recherche expérimentale sur Jigsaw auprès d'élèves français du secondaire. Dans une littérature marquée par des risques de sous-puissance statistique et dont la méthodologie peut manquer de transparence (voir chapitre 3), ces cinq expériences apportent une remise en question de l'efficacité de Jigsaw (Johnson et al., 2000 ; Moskowitz et al., 1985 ; Souvignier & Kronenberger, 2007). Que ce soit sur des durées courtes et pilotées par des expérimentateur·ices (E1 et E2) ou sur des durées plus longues et dans des classes réelles, par l'intermédiaire d'enseignant·es volontaires (E3A, E3B et E3C), nous n'observons pas d'effet positif de Jigsaw sur les apprentissages. Les trois expériences menées le long d'un trimestre et mesurant l'apprentissage à plusieurs reprises constituent un apport méthodologique majeur pour la solidité de nos résultats. En effet, nous avons observé que, sur les études mises en place entre 2000 et 2020, la durée médiane des expérimentations Jigsaw était de 9 heures (réparties sur ± 6 semaines de cours, voir Tableau 9, chapitre 4). Dans nos expérimentations en classes réelles, la durée médiane peut être estimée à 16 heures de cours, réparties sur 12 à 15 semaines

d'enseignement. Cette amélioration de la durée est considérable, d'autant plus que nous avons utilisé plusieurs mesures distinctes afin d'estimer l'efficacité globale de Jigsaw sur cette période. À titre de comparaison, seules Roseth et collaborateurs (2018) utilisent une procédure similaire, c'est-à-dire une expérience longitudinale où l'apprentissage est mesuré à diverses reprises sur des évaluations de connaissances différentes. De plus, comme présentée dans les chapitres précédents, la littérature portant sur Jigsaw souffre d'un problème dans la cohérence entre la nature des données récoltées et les analyses statistiques effectuées (Protogerou & Hagger, 2020). L'utilisation de multiples mesures de performance académique, plutôt que de deux mesures identiques mises en place avant et après l'expérimentation (voir chapitre 3), nous semble être une étape importante de l'amélioration de la qualité méthodologique de ces études. Pour résumer, la durée expérimentale est dans les bornes supérieures de celles rapportées classiquement dans la littérature Jigsaw, et le problème lié à l'indépendance des données est supposément contrôlé par l'utilisation d'un test-*t* apparié et d'une méta-analyse interne.

Il est important de souligner la mise en évidence d'une forte corrélation entre le niveau scolaire initial et la réussite aux évaluations de connaissance (i.e., les performances académiques) et d'un coefficient de corrélation intra-classe à prendre en considération pour les analyses. Premièrement, les analyses de covariance (ANCOVA) effectuées à titre de contrôle méthodologique montrent que le niveau scolaire initial des élèves est fortement et positivement corrélé aux performances académiques ultérieures ($r_{\text{médian}} = .60$, voir Annexe A). Bien que les ANCOVA ne modifient pas les conclusions de nos analyses par test-*t*, elles nous informent sur le fait que nos mesures de performances sont toutes positivement associées à une autre mesure de performance, classiquement considérée comme un bon indicateur du niveau de maîtrise des concepts par les élèves. Sur le plan méthodologique, ces résultats suggèrent que le contrôle du niveau scolaire initial des participant-es pourrait permettre un gain de puissance statistique important dans les modèles statistiques (Yzerbyt et al., 2004). Or, dans la littérature Jigsaw, le contrôle du niveau initial des participant-es, ou bien du caractère hiérarchique des données, n'est pas systématiquement pris en compte. Utiliser un indicateur de niveau initial, en tenant compte du problème de l'indépendance des données évoqué dans le chapitre 3, pour gagner en puissance statistique peut être une perspective pour gagner en puissance statistique dans des recherches futures.

Deuxièmement, les tailles d'échantillon sur lesquelles nous avons effectué nos analyses statistiques ont été estimées via un calcul de puissance statistique, en amont des expérimentations (Faul et al., 2007). La puissance statistique de nos expérimentations est donc

supposément plus élevée que celles traditionnellement rapportées dans la plupart des études sur Jigsaw. Nous avons systématiquement relevé des coefficients de corrélation intra-classe (ICC) de faible amplitude à travers nos cinq expériences (Bressoux, 2020). Dans le cadre de nos analyses, le calcul de l'ICC servait surtout de contrôle méthodologique pour s'assurer des faibles variations entre les classes, réparties dans nos conditions expérimentales. La prise en compte de ce facteur peut être importante, notamment dans les études dont les données sont hiérarchisées (i.e., participant·es inclus·es dans des conditions expérimentales au niveau de classes et/ou des établissements), puisqu'elle agit sur la puissance de test associée à l'étude : lorsque l'ICC est élevé, il faut davantage de classes, c'est-à-dire de participants, pour détecter les effets attendus (Bressoux, 2020).

Nous pouvons évoquer trois limites des expériences présentées dans ce chapitre : la première étant que les disciplines que nous avons sélectionnées sont exclusivement « scientifiques » (i.e., mathématiques, sciences de la vie et de la Terre, physique-chimie), la deuxième porte sur la manière dont les performances académiques ont été évaluées dans les expérimentations 3A, 3B et 3C et la dernière concerne la nature des groupes contrôles utilisés. Par rapport au contenu disciplinaire, nous avons choisi de nous focaliser essentiellement sur les contenus scientifiques puisque ce sont sur eux que portent la majorité des recherches menées sur Jigsaw jusqu'à ce jour (Robert et al., *en préparation*). Comme présenté dans les chapitres 2 et 4, certain·es auteur·es suggèrent que la discipline étudiée peut agir comme un modérateur de l'efficacité de Jigsaw : les sciences sociales, les langues ou l'histoire-géographie permettraient davantage de flexibilité dans les contenus, tandis que les sciences « dures » seraient plus adaptées à la résolution de problèmes et aux réponses fermées (Aronson & Patnoe, 2011 ; Cohen, 1994). À ce jour, nous ne disposons malheureusement d'aucune étude qui teste spécifiquement ce modérateur. De plus, la revue de littérature sur Jigsaw menée par Vives et collaborateurs (*en préparation*) indique que les effets positifs de Jigsaw rapportés dans la littérature tendent à être indépendants des disciplines étudiées. Idéalement, il faudrait mettre en place des études sur lesquelles plusieurs disciplines seront testées, comme par exemple le français et les mathématiques, afin de comparer les effets de Jigsaw en fonction de la discipline.

Concernant la deuxième limite, les mesures que nous avons utilisées pour tester notre hypothèse de recherche possèdent plusieurs défauts : celles utilisées pour les expériences 1 et 2 n'ont pas fait l'objet de pré-test sur la population cible et ont été mises en place par l'expérimentateur en charge des passations. Il est donc possible que ces évaluations aient été inadaptées pour mesurer des différences de performance. Par exemple, dans l'expérimentation

1, des analyses exploratoires sur les différentes sous-parties composant l'évaluation de connaissances mettent en évidence des disparités entre ces dernières : en moyenne, les scores sont notamment plus élevés sur les parties relatives à la géométrie ($M = 7.50$; $SD = 2.78$) ou aux fractions ($M = 6.18$; $SD = 2.55$) que sur la proportion ($M = 3.37$; $SD = 2.96$) ou les diagrammes ($M = 4.74$; $SD = 4.43$). Contrairement à ce qui aurait pu être attendu, nous n'observons pas non plus d'effet de « compatibilité » entre la sous-partie assignée à l'élève et sa réussite ultérieure sur l'évaluation de connaissances relative à cette sous-partie (Nolan et al., 2018). De manière générale, c'est l'effet de la covariable « niveau scolaire initial » qui semble couvrir l'essentiel de la variation des scores dans l'expérience 1, et non pas la distribution de ressources complémentaires entre les groupes expérimentaux. C'est pourquoi, dans l'expérience 2, nous avons choisi de présenter un contenu pédagogique non enseigné dans le cadre du cursus scolaire des élèves de 6^e. Néanmoins, les scores observés dans cette seconde expérimentation sont quasiment similaires entre les deux conditions, ce qui pourrait potentiellement signaler la présence d'un score plafond sur notre mesure. Pour les expériences menées en contexte de classe réelle, c'est un problème d'un autre ordre qui se pose à la question des mesures : bien que ces dernières soient supposément plus adaptées que des évaluations construites par un·e expérimentateur·rice externe, elles ont été systématiquement corrigées par les enseignant·es, non-aveugles aux conditions expérimentales. Une autre source de biais peut provenir de la nature même de la notation des évaluations que pratiquent les enseignant·es (voir Butera et al., 2011). Allen (2005) met en évidence trois limites principales que posent les notations telles que pratiquées par les enseignant·es :

1. Les enseignant·es tendent à utiliser les notes pour communiquer plusieurs informations telles que les efforts fournis, la progression de l'élève et « l'équité » de la notation par rapport aux autres, plutôt que de certifier une réponse comme correcte ou fausse.
2. La notion de « performance académique », ainsi que les activités pédagogiques qui les évaluent diffèrent grandement d'un·e enseignant·e à l'autre.
3. En conséquence, les notes assignées aux élèves ne sont pas toujours basées sur un « principe solide de mesure » (Allen, 2005, p.219).

Dans leur revue de littérature sur la question des évaluations et de l'assignation de notes, Brookhart (2016) et collaborateurs développent l'argument que la note est en réalité un construit multidimensionnel qui prend en compte à la fois des critères « objectifs », qui renvoient un feedback en fonction de la justesse de la réponse, mais également des critères « subjectifs », comme les efforts fournis par l'élève, son assiduité et son comportement dans la classe. D'après

ces auteur·es, les notes assignées par les enseignant·es corrélaient de manière « modérée » avec les tests standardisés. Iels expliquent notamment que « 25% de la variation dans les notations des enseignant·es provient d'un trait mesurable par les évaluations standardisées. Les 75% restants sont attribuables à quelque chose d'autre » (Bowers, 2011 cité dans Brookhart et al., 2016). Pour nos expériences, il est possible que la subjectivité des enseignant·es ait joué un rôle dans leur notation. Puisqu'ils participaient dans l'enseignement des deux conditions expérimentales (i.e., Jigsaw et enseignement habituel), il est tout à fait probable que leurs notations aient été impactées par leur vécu au regard de l'expérimentation, ainsi que par les aspects « subjectifs » des notations des élèves sur la base de leurs efforts et de leur progression. Pour dépasser cette limite, il semble essentiel de mener des études qui garantiraient une évaluation extérieure à l'expérience.

Enfin, il peut être intéressant de noter, même si cela n'a pas fait l'objet de mesures spécifiques, que tous les enseignant·es ayant participé aux présentes expérimentations ont rapporté avoir perçu une amélioration du climat de classe après la mise en place de Jigsaw. Ces derniers ont observé que les collégien·nes étaient plus enclin·es à coopérer les un·es avec les autres, à offrir une aide spontanée à ceux qui étaient en difficulté voire même à transférer la structure en trois étapes de Jigsaw lors de travaux en groupe. Il pourrait être intéressant d'évaluer quantitativement l'évolution des comportements d'aide dans les classes, puisque cela nous permettrait d'apporter du soutien empirique à la théorie de l'interdépendance positive. Il est possible que Jigsaw affecte le climat de classe et l'émergence de comportements altruistes, mais n'affecte pas les performances académiques *per se*, comme évoqué par les enseignant·es des expériences 3A, 3B et 3C. On pourrait également évaluer dans quelles mesures les comportements altruistes et la cohésion sociale sont reliés à l'apprentissage, c'est-à-dire aux performances académiques. Si les élèves ont progressivement développé des compétences sociales et un sentiment de cohésion sociale important, ce qui se traduit par un climat de classe positif, mais que leurs scores aux évaluations de connaissances n'ont pas augmenté, c'est que le lien entre ces variables n'est pas si évident. Il est probable que Jigsaw favorise bel et bien l'émergence d'interactions facilitatrices entre les individus, mais que les évaluations de connaissances déployées dans le cadre scolaire ne captent pas cette progression. Les futures recherches pourraient utiliser des mesures de compétences sociales, ou bien s'intéresser plus précisément aux interactions entre les individus dans Jigsaw, comme le préconise l'approche cognitive (chapitre 1). On pourrait par exemple observer le nombre et la qualité des interactions

facilitatrices dans les groupes (Volet et al., 2009), ou bien mettre en place des activités mesurant spécifiquement les capacités de coopération entre individus.

L'absence de différence significative entre les deux conditions expérimentales pourrait également être due à la nature des groupes contrôle utilisés dans les présentes expériences. La plupart des recherches sur Jigsaw utilisent des groupes contrôles qui consistent en une approche « individuelle » pure (e.g., Lazarowitz et al., 1994 ; Şahin, 2011 ; Tarhan et al., 2013). Or, les conditions d'enseignement habituel utilisées comme cible de comparaison dans les expériences 3A, 3B et 3C ressemblent plus à un enseignement « hybride » qu'à un enseignement purement individuel. En effet, l'enseignement habituel mis en place dans ces groupes comprenait d'une part, une approche centrée sur l'enseignant-e qui transmet des informations à toute la classe, et d'autre part, des travaux en groupe informels. Il est donc possible que les différences entre l'apprentissage coopératif selon Jigsaw et une approche « hybride » soient beaucoup plus réduites que $\delta = 0.40$ (Johnson et al., 2000). L'usage de tels groupes contrôles nécessite de reconsidérer la notion de plus petite taille d'effet d'intérêt : on peut raisonnablement penser que les différences d'effets soient plus importantes entre un groupe expérimental individuel (ou compétitif) et coopératif, plutôt qu'entre un groupe coopératif et un groupe hybride. Pour mieux cerner les différences d'effets entre la nature des groupes contrôles, il faudrait disposer de plusieurs structurations de l'interdépendance positive : on pourrait s'attendre à ce que les différences soient plus marquées entre des groupes coopératifs vs compétitifs et individuels, et moins importantes entre des groupes coopératifs vs mixtes. L'utilisation de différents « niveaux de coopération » nous offrirait une vision plus subtile de la modulation des différences dans la performance académique des élèves par le degré de structuration de l'interdépendance positive.

9.3. Viser le « *gold standard* » des expérimentations randomisées à grande échelle

Il semble difficile de conclure à l'(in)efficacité de Jigsaw sur la base seule de ces cinq études. Bien que nos expériences aient été menées dans des conditions de puissance statistique satisfaisantes pour détecter un effet d'amplitude proche de $\delta = 0.40$, nos résultats tendent à indiquer que, sur une population de collégien·nes français·es, l'effet de Jigsaw est probablement bien plus petit que prévu par la littérature, voire inexistant ($ES = 0.01$, 95 % CI [-.10, .11]). Ces résultats contrastent avec un pan de la littérature récente qui tend à rapporter des effets larges et positifs de Jigsaw sur les apprentissages (voir chapitre 4), mais font écho à d'autres travaux comme ceux de Moskowitz et collaborateurs (1983 ; 1985) ou Souvignier et Kronenberger (2007) qui n'observent pas d'effets positifs chez les élèves de ce niveau scolaire. Nous avons proposé plusieurs pistes à explorer pour clarifier ces disparités de résultats et mieux comprendre

l'hétérogénéité des effets de Jigsaw : conduire des expériences randomisées sur des échantillons plus larges pourrait permettre de lever le doute quant à la véritable taille d'effet de Jigsaw.

Connoly et collaborateurs (2018) rapportent l'intérêt que représentent de telles expérimentations : « [...] *les essais contrôlés et randomisés ont joué un rôle central pour déterminer si des interventions académiques ont un effet discernable et mesurable sur l'apprentissage des étudiant-es et leur développement* » (p.2). Ces auteur-es mettent en évidence que la part de RCTs dans le domaine de la psychologie de l'éducation a considérablement augmenté depuis les années 1980, ce qui constitue une preuve indirecte de leur importance tout aussi bien théorique et pratique. En France, l'expérimentation ProFan, qui regroupe près d'une centaine d'établissements, soit 10 000 élèves de lycée professionnel, s'inscrit dans cette trajectoire. Ce déploiement expérimental surpasse de loin la taille moyenne des échantillons d'autres RCTs dans le domaine de la psychologie de l'éducation ($M = 1\,251$; $SD = 4\,102$, voir Connoly et al., 2018). Elle s'inscrit également dans les études en provenance d'Europe, qui représentent 16% ($N = 167$) soit la deuxième localisation la plus importante, et porte sur une population très peu étudiée à ce jour, les élèves de lycée professionnel. Le contexte dans lequel prend place cette expérimentation est celui du développement du numérique dans les activités professionnelles futures et vise à développer plusieurs activités pédagogiques dans les filières générales (i.e., mathématiques et français) et professionnelles (i.e., Accompagnement, soins et services à la personne, Commerce, Métiers de l'Électricité et de ses Environnements Connecté).

Dans l'expérimentation ProFan, certains élèves ont été amenés à collaborer sur plusieurs séquences en suivant une structure proche de Jigsaw (G1), tandis que d'autres ont été amenés à travailler dans des groupes moins structurés (G2), ou en maintenant les pratiques d'enseignement habituelles (G3) : les effets de ces différents niveaux de structuration de l'interdépendance portent sur un large panel de mesures, socioaffectives, scolaires et comportementales et sont évalués à plusieurs reprises à travers des questionnaires et des évaluations de connaissances. Ces éléments devraient nous permettre de pallier les limites des études présentées dans ce chapitre. En effet, la présence de deux conditions expérimentales, avec des degrés différents de structuration d'interdépendance positive, nous permettra de tester l'hypothèse selon laquelle cette structuration est bien reliée aux apprentissages. À cette fin, nous disposons dans ProFan de plusieurs mesures de performances académiques portant sur différents contenus pédagogiques, à la fois généraux et professionnels, qui ont été examinées par des évaluateurs externes dans le cadre de l'expérimentation. Dans ProFan, il est donc

plausible que les données récoltées soient moins sujettes à la subjectivité des enseignant·es. Outre les mesures de performances académiques, les participant·es ont également pris part à des activités en groupe, des « *Boite à Outils* », mesurant leurs compétences de coopération, de créativité et de partage de l'information. Par exemple, dans une activité de « profil caché », les participant·es devaient lire le texte d'une enquête policière et désigner la personne responsable d'un accident : chaque participant·e disposait d'une information différente qui devait être communiquée aux autres membres du groupe afin de trouver le coupable. La réussite ou l'échec à cette activité, ainsi que l'analyse des interactions ayant lieu entre les individus, par chat écrit, peut consister en une mesure comportementale des compétences de coopération entre individus. Par leur caractère longitudinal, nous pourrions à l'aide de ces données tester l'hypothèse selon laquelle la coopération selon Jigsaw permet de développer ces compétences. Enfin, l'importante taille d'échantillon de l'expérimentation nous permettra à la fois de détecter supposément inférieurs à $\delta = 0.40$, mais aussi de tester la présence de modérateurs et leurs effets éventuels sur l'efficacité de Jigsaw. Puisque nous disposons d'éléments démographiques (e.g., genre, statut socio-économique, âge) récoltés par un questionnaire déployé avant le début de l'expérimentation mais également de variables socioaffectives (e.g., estime de soi, adhésion aux buts de maîtrise et de performance, compétences verbales), nous pourrions tester l'effet de Jigsaw en fonction de caractéristiques individuelles des élèves. Cela nous permettrait de vérifier la présence de modérateurs : par exemple, on peut imaginer que Jigsaw soit plus efficace chez les individus orientés vers la maîtrise de leurs contenus (Poortvielt & Darnon, 2014), ou qui se sentent plus compétents dans leurs capacités à expliquer leur contenu (Hänze & Berger, 2007 ; Nolan et al., 2018).

Pour conclure, l'expérimentation ProFan pourrait nous permettre d'enrichir considérablement la littérature scientifique sur Jigsaw. Grâce à une variété d'échelles socioaffectives, des mesures de coopération et d'apprentissages réparties à travers plusieurs disciplines, elle offre une opportunité de dépasser les limites que nous avons rencontrées dans nos expériences. L'importante puissance de test d'une telle expérimentation sera un atout majeur pour quantifier l'hétérogénéité des résultats de Jigsaw, et mieux comprendre les disparités observées dans la littérature récente.

Résumé du chapitre 5

- À travers cinq expérimentations (dont trois en contexte de classe réelle le long d'un semestre), nous avons testé l'efficacité de Jigsaw comparativement à des groupes contrôles (individuel ou d'enseignement habituel) qui étudiaient le même contenu pédagogique.
-
- La méta-analyse interne réalisée sur ces expériences rapporte une taille d'effet relativement uniforme et proche de zéro, c'est-à-dire largement inférieure à notre plus petite taille d'effet d'intérêt : Jigsaw n'a donc pas eu les effets attendus sur nos échantillons.
-
- Nous proposons de tester sur un échantillon encore plus large, à l'aide de mesures plus fiables et variées et couvrant diverses disciplines, afin de dépasser les limites de nos expériences. L'expérimentation ProFan, qui regroupe environ 10 000 élèves répartis dans une centaine d'établissements français, devrait nous permettre de tester les différentes propositions formulées dans ce chapitre.
-

Chapitre 6. Repenser le puzzle : Synthèse, limites et perspectives

1. Objectif du chapitre

Dans ce dernier chapitre, nous rappellerons les hypothèses de recherche ainsi que les principaux résultats observés dans les chapitres empiriques précédents. Nous apporterons ensuite de nouveaux éléments permettant de discuter et réinterpréter ces résultats à la lumière des chapitres théoriques développés au début de ce manuscrit. Pour conclure, nous présenterons des pistes pour de futures recherches sur Jigsaw.

2. Résumé des chapitres précédents

Cette thèse avait pour ambition de tester l'efficacité de la « classe puzzle », ou *Jigsaw*, une pédagogie coopérative supposée améliorer l'apprentissage des individus grâce à l'interdépendance positive de ressources. Dans le premier chapitre, nous avons présenté une vue d'ensemble théorique et empirique de la littérature scientifique s'intéressant à l'apprentissage coopératif. Nous avons décrit les principales conditions nécessaires pour structurer l'apprentissage coopératif, notamment au regard de la théorie de l'interdépendance positive de Johnson et Johnson (1989 ; 2005 ; 2009), ainsi que les effets globaux de l'apprentissage coopératif sur les performances académiques. Dans le deuxième chapitre, nous nous sommes particulièrement intéressés à Jigsaw, l'une des méthodes les plus populaires dans la littérature (Aronson & Patnoe, 2011 ; Roseth et al., 2018). Nous avons soulevé l'idée que, même si cette pédagogie est caractérisée par un engouement important de la part de ses concepteur·rices et de certain·es enseignant·es, les preuves de son efficacité tendent à manquer : le nombre d'articles réalisés dans des conditions méthodologiques rigoureuses est relativement faible, et les preuves de l'efficacité de Jigsaw sur les apprentissages sont loin d'être conclusives. Comme discuté dans le troisième chapitre, de nombreuses études testant les effets de Jigsaw sur les apprentissages manquent certainement de puissance statistique, ce qui peut entraîner une surestimation de l'efficacité de l'intervention. La méta-analyse présentée dans le chapitre 4 révèle que ce problème de puissance statistique concerne particulièrement les études menées au Moyen-Orient, notamment en Turquie. Sur les pays occidentaux, notamment en Europe ou aux États-Unis où les études tendent à être réalisées avec une meilleure puissance, les estimations de l'efficacité de Jigsaw sont bien moindres, voire nulles. Dans les chapitres 5, nous avons testé expérimentalement les effets de Jigsaw à travers cinq expériences, dont trois en contexte de classe réelle mises en place le long d'un semestre. Nous n'observons aucune preuve de l'efficacité de Jigsaw par rapport à un groupe d'enseignement individuel (E1 et E2) ou avec

maintien des pratiques d'enseignement habituelles (E3A, E3B, E3C) : l'effet global est nul et ne varie pas d'une étude à l'autre. Puisque nous disposons d'une puissance de test proche de 90% pour détecter une plus petite taille d'effet fixée à $\delta = 0.40$, nous avons conclu qu'il était très peu probable que Jigsaw ait un effet positif d'intérêt pratique sur cette population.

Dans ce dernier chapitre, nous présentons les contributions principales de cette thèse ainsi que de nouveaux arguments pour discuter l'intégralité de ces résultats. D'autres modérateurs possibles des effets de Jigsaw sont développés, ainsi des pistes de réflexion axées autour des limites de nos expériences et de perspectives pour les dépasser. Enfin, nous concluons ce travail de recherche par une remise en perspective du rôle de l'apprentissage coopératif, de manière plus générale, au sein de l'institution scolaire, ainsi que plusieurs propositions pour favoriser l'implémentation de ces dispositifs.

3. Quantifier et qualifier les effets de Jigsaw sur les apprentissages

3.1. La nécessité de tester empiriquement une pratique pédagogique

Rappelons que l'étude expérimentale des effets de Jigsaw sur les apprentissages reste limitée, particulièrement dans le contexte scolaire français (chapitre 2). Dans leur revue de la littérature, Robert et collaborateurs (*en préparation*) ne rapportent qu'une seule étude empirique menée sur Jigsaw en France (Darnon et al., 2012), et celle-ci ne comprend pas de mesure d'apprentissage. De manière plus générale, les études testant les effets de Jigsaw tendent à être réalisées principalement auprès de populations de l'enseignement supérieur et sur des échantillons de plus en plus faibles (Robert et al., *en préparation* ; Vives et al., *en préparation*). Bien que la production d'études empiriques sur Jigsaw ait pour le moins explosé ces deux dernières décennies (notamment en Turquie, voir chapitre 2) ces études sont souvent menées dans des conditions méthodologiques insatisfaisantes : problème de puissance statistique, manque de transparence dans les procédures, et implémentation de piètre qualité sont autant d'indicateurs qui peuvent diminuer la confiance que l'on peut accorder à de tels résultats (Robert et al., *en préparation* ; Vives et al., *en préparation*).

Une première contribution de ce manuscrit renvoie à la revue critique et la méta-analyse effectuée sur les travaux récemment publiés dans la littérature Jigsaw (chapitres 2 à 4). Nous avons observé que les études menées dans les pays orientaux tendent à rapporter des effets très larges tandis que les études menées dans des pays occidentaux rapportent des effets proches de zéro. De manière plus détaillée, ce sont surtout les études menées en Turquie sur lesquelles les estimations de l'efficacité de Jigsaw sont très élevées, mais surtout où le risque de surestimation

de l'efficacité de la méthode Jigsaw est le plus grand. La démonstration du risque de sous-puissance statistique et des biais méthodologiques provenant de ces articles met en évidence le contraste qui existe entre nos résultats et ceux rapportés dans les recherches menées au Moyen-Orient : la seule explication culturelle, même si elle ne peut être exclue, nous semble insuffisante pour expliquer la disparité aussi importante entre nos résultats et ceux obtenus par les chercheur·es Turques. Cette contribution peut être critique, notamment en considérant les récentes mises à jour de la « méga-analyse » de Hattie (2017) qui situent les effets de Jigsaw à $\delta = 1.20$ sur la base de la méta-analyse de Batdi (2014). Une telle information, si elle s'avérait biaisée, peut mettre les chercheur·es, praticien·nes et enseignant·es sur une mauvaise piste en surestimant l'efficacité de la méthode Jigsaw. D'un point de vue scientifique, une estimation biaisée de l'efficacité de Jigsaw peut aboutir à la mise en place d'expérimentations sous-puissantes, et alimenter la surestimation de l'effet avec un biais de publication en faveur des résultats positifs (chapitre 3). Les arguments développés dans ces chapitres doivent mettre en garde les chercheur·es sur le risque de la surestimation de l'efficacité de Jigsaw.

La deuxième contribution de ce manuscrit consiste en la mise en place d'expérimentations sur des populations relativement peu étudiées jusqu'à présent : des élèves de collèges français (chapitre 4). Ces études ont été menées dans des contextes scolaires « réels » dans un milieu relativement « écologique ». Dans les expériences 3A, B3 et 3C, les cours étaient dirigés par les enseignant·es respectif·ves des élèves ayant participé aux expérimentations et les évaluations de connaissances sur lesquelles nous avons effectué nos analyses renvoient à des notes qui comptaient pour l'évaluation des élèves dans leurs études. Les résultats de ces expériences nous permettent d'apporter des arguments supplémentaires sur la remise en question de l'efficacité de Jigsaw dans les apprentissages. Pour rappel, dans leur méta-analyse qui comparait les effets de différentes pédagogies coopératives sur l'apprentissage, Johnson et collaborateurs (2000) proposaient d'employer une approche critique des interventions n'ayant pas fait l'objet de validation empirique par la recherche expérimentale (voir chapitre 1).

Pour résumer, les résultats de la méta-analyse et de nos expérimentations suggèrent que l'engouement autour de Jigsaw sur les performances académiques peut paraître disproportionné. Les expériences menées dans le chapitre 4 suggèrent que l'effet de Jigsaw sur les apprentissages des collégiens est extrêmement faible, voire inexistant. L'estimation globale de son efficacité n'est pas différente de zéro et les effets ne diffèrent pas d'une étude à l'autre, ce qui supporte l'hypothèse nulle. L'homogénéité des effets, que ce soit entre les expériences

« courtes » (E1 et E2) et les « longues » (E3A, E3B et E3C) alimente l'hypothèse selon laquelle Jigsaw n'a aucun effet positif sur cette population, indépendamment de la durée de l'intervention (voir Annexes M, N et O). Cependant, les effets de Jigsaw rapportés dans la littérature sont surtout caractérisés par leur hétérogénéité. La méta-analyse du chapitre 4 quantifie cette hétérogénéité, et rapporte que plus de la moitié de la dispersion des effets de Jigsaw provient de la localisation des articles : l'hypothèse qu'un ensemble d'articles influe fortement sur l'estimation des effets de Jigsaw est une première piste que nous avons soulevée dans ce manuscrit, bien que d'autres explications soient possibles. Afin de tester la présence d'effets potentiellement plus petits que notre seuil fixé à $\delta = 0.40$, ainsi que d'éventuels modérateurs, des études plus puissantes doivent être mises en place. L'expérimentation ProFan brièvement présentée dans le chapitre précédent devrait permettre d'atteindre cet objectif.

3.2. Des modérateurs de l'efficacité de Jigsaw ?

Bien que la modulation par la localisation semble être l'interprétation privilégiée par nos résultats, d'autres modérateurs ont sûrement des impacts non négligeables sur Jigsaw, notamment la durée de l'expérimentation, la discipline où l'âge des participant·es. Le fait que ces modérateurs n'étaient pas statistiquement significatifs dans notre méta-analyse serait une conséquence simple du manque d'études incluses, c'est-à-dire du manque de puissance de test. Dans le cadre de nos expériences, l'existence de modérateurs pourrait expliquer a posteriori pourquoi nous n'obtenons pas d'effet positif de la méthode : si Jigsaw n'a des effets positifs que dans certaines conditions, ou auprès de certaines populations, et des effets nuls voire négatifs chez d'autres, alors l'effet global de la méthode peut apparaître comme nul. En conséquence, nous pensons que d'autres études sont nécessaires pour que les comparaisons méta-analytiques gagnent en puissance de test (voir Figure 14, chapitre 4) et que certains modérateurs de l'effet de Jigsaw soient testés empiriquement. Dans les études menées auprès des élèves de collèges, les effets de Jigsaw sur les apprentissages, tels que mesurés par les performances académiques à des évaluations de connaissances, sont nuls et homogènes. Cependant, dans la méta-analyse des travaux récents publiés sur Jigsaw, plusieurs modérateurs ont été testés : même si les différences d'estimation de tailles d'effet n'étaient pas statistiquement significatives, plusieurs pistes peuvent être explorées pour traiter ces modérateurs possibles.

Par exemple, nous avons observé que la plupart des recherches mises en place après 2000 sont mises en place auprès d'étudiant·es de l'enseignement supérieur, population la plus étudiée dans la littérature Jigsaw (Robert et al., *en préparation*). On peut s'intéresser aux

différences de niveau scolaire, ou d'âge, dans l'évaluation de l'efficacité de Jigsaw : il est possible que les effets de Jigsaw soient plus marqués chez étudiant·es de l'enseignement supérieur que chez les élèves du secondaire pour des raisons développementales. Dit autrement, on pourrait s'attendre à ce que les élèves plus avancés dans la scolarité soient capables de coopérer de manière plus efficace. Pourtant, cette première explication rencontre rapidement des limites. D'une part, dans la littérature existante, des effets positifs de Jigsaw sont observés de la maternelle jusqu'aux études supérieures en passant par diverses populations de niveaux intermédiaires (Robert et al., *en préparation*). Les études *princeps* de Jigsaw portent d'ailleurs sur des populations de niveau scolaire secondaire, et non pas sur les étudiant·es du supérieur. De manière plus générale, l'argument selon lequel les compétences sociales se développent avec le temps ne résiste pas longtemps à un examen de la littérature. Buchs et Butera (2015) ont par exemple mis en évidence que les étudiant·es de psychologie, c'est-à-dire une population provenant de l'enseignement supérieur, avaient du mal à utiliser efficacement les compétences sociales pour coopérer. Dans un autre article, ces mêmes auteur·es suggèrent que la socialisation dans le contexte scolaire ne permet pas aux élèves de développer ces compétences sociales de manière optimale (Buchs & Butera, 2019). Iels argumentent que l'école étant davantage centrée vers un mode d'enseignement individualiste, les élèves ne sont pas formés à l'utilisation de compétences sociales (e.g., exprimer ses émotions ou ses sentiments, nuancer son opinion ou critiquer les idées et non la personne). Dans un contexte individualiste qui évalue et valorise la réussite individuelle des élèves, il semble peu probable que l'école fasse émerger ces compétences même après plusieurs années (Butera et al., 2006).

Un autre modérateur évoqué dans le chapitre 4 est celui de la durée des interventions : Aronson et Patnoe (2011) affirment que l'apprentissage coopératif, notamment via Jigsaw, est un changement de paradigme qui nécessite du temps pour que les élèves apprennent à coopérer entre eux. Pourtant, une fois encore, la littérature Jigsaw ne permet pas d'apporter une réponse claire à la question du temps nécessaire pour que la pédagogie fasse effet. Les enseignant·es ayant participé aux expériences présentées dans le chapitre 5 nous ont fait part de questionnement sur la durée de nos interventions : pour certain·es, un trimestre est trop court pour faire émerger les effets attendus de la méthode Jigsaw sur les apprentissages. Pourtant, comparativement au reste de la littérature scientifique, la durée des expériences menées en classe entière était relativement élevée : entre 12 à 16 heures de cours réparties sur plus de 10 semaines de cours. D'un côté, on peut formuler l'hypothèse que Jigsaw nécessite davantage de temps afin d'avoir des effets positifs sur les apprentissages (Moreno, 2009 ; Roseth et al., 2018),

mais d'un autre côté, même les études les plus longues testant la méthode n'ont rapporté aucun effet (Moskowitz et al., 1983 ; 1985). De la même manière, notre méta-analyse apporte un élément qui va à l'encontre de l'hypothèse de la nécessité d'une durée accrue : les effets de Jigsaw semblent plus élevés lorsque l'intervention est courte (i.e., inférieure à 9 semaines), plutôt que longue (i.e., supérieure à 9 semaines). En l'état, une modération par la durée ne semble pas être suffisamment établie pour pouvoir faire des préconisations claires sur le sujet et expliquer la disparité entre les résultats obtenus dans nos deux ensembles d'études. La mise en place de l'expérimentation ProFan, qui couvre l'équivalent de deux années de scolarisation pour chaque cohorte, apportera une contribution importante pour répondre à la question de la modération par la durée.

Pour finir, nos expérimentations se sont focalisées sur l'enseignement de contenus « scientifiques », c'est-à-dire de physique-chimie (expérience 3A), de mathématiques (expérience 1) et de sciences de la vie et de la terre (expériences 2, 3B et 3C). Comme pour les modérateurs présentés précédemment, plusieurs recherches rapportent des arguments et des résultats contradictoires : la méthode Jigsaw devrait être plus adaptée aux contenus littéraires (Aronson & Patnoe, 2011 ; Cohen, 1994), mais c'est sur les contenus scientifiques que portent la majorité des recherches sur lesquels les effets positifs sont les plus souvent rapportés (Robert et al., *en préparation* ; Springer et al., 1999 ; Vives et al., *en préparation*). Plutôt que de nous interroger sur la nature du contenu pédagogique en tant que tel, nous souhaitons présenter une autre piste d'explication des disparités entre ces résultats, qui porte plutôt sur la structure de la complémentarité des ressources utilisée dans Jigsaw. Deiglmayr et Schalk (2015) définissent « l'interdépendance de connaissance » comme un facteur qui prédit la réussite des individus à échanger les informations complémentaires dans la phase puzzle de Jigsaw. Les auteur-es développent l'idée que lorsque les ressources distribuées aux membres sont trop distinctes, il est plus difficile pour elleux de reconstituer un « socle de connaissances commun » et donc de mettre en commun ces ressources complémentaires de manière efficace. D'après les auteur-es, l'efficacité de Jigsaw repose sur la capacité des individus à construire des nouvelles connaissances, sur la base d'un socle de connaissances communes. Dans Jigsaw, il est possible de distribuer des ressources complémentaires qui illustrent un même « concept de base » à l'aide d'exemples différents, où bien distribuer des ressources développant des « concepts » différents. Selon elleux, les activités d'apprentissage coopératif sont plus efficaces lorsque les interactions entre individus sont « *co-constructives* », c'est-à-dire lorsque les individus vont au-delà de l'information dont iels disposent et s'engagent dans une construction de nouvelles

connaissances grâce à l'échange avec autrui. Cette co-construction de connaissances nécessite cependant que les individus engagés dans l'activité d'apprentissage possèdent un même socle de connaissance commun ; dans le cas contraire, les individus ne disposent pas des bases nécessaires pour aboutir à des nouvelles connaissances et l'apprentissage coopératif est inefficace. En conséquence, l'explication de concepts trop éloignés lors de la phase puzzle de Jigsaw ne permet pas de co-construire des connaissances. Il s'agit plutôt d'une activité d'enseignement « passive », où chaque membre effectue à tour de rôle une restitution de ses connaissances nouvellement acquises, ce qui n'a aucun bénéfice sur les apprentissages des autres membres (Chi & Wylie, 2014 ; Deiglmayr & Schalkt, 2015).

Dans nos expériences menées auprès d'élèves de collèges, bien que nous n'ayons pas observé de variations dans les effets de Jigsaw, la question de l'interdépendance des connaissances se pose : on peut formuler l'hypothèse que cette interdépendance est probablement plus faible dans les expériences 3A, 3B et 3C, comparativement aux expériences 1 et 2, puisque les séances étaient articulées autour d'une même problématique étudiée à l'aide d'exemples différents. Dans les expériences 1 et 2, chaque élève disposait d'une ressource traitant d'un « concept », et non pas d'un exemple différent, ce qui a probablement augmenté le degré de difficulté de la tâche. Pour clarifier cette question, on pourrait s'intéresser à l'analyse du contenu des séquences utilisées dans nos expériences. Suivant l'hypothèse de Deiglmayr et Schalk (2015), on pourrait s'attendre à ce que les séquences où la distribution des ressources complémentaires était effectuée de manière fortement interdépendante soient reliées à une moins bonne réussite des élèves sur les évaluations de connaissance. Dans le cadre de l'expérimentation, nous disposerions d'une gamme plus importante de séquences et de contenus disciplinaires (i.e., enseignements professionnels et généraux de mathématiques et de français), ce qui nous permettrait de tester de manière plus détaillée cette hypothèse d'interdépendance des connaissances.

Nous pourrions également nous intéresser non pas au contenu *per se* mais à la manière dont les interactions entre individus ont été structurées dans Jigsaw. Dit autrement, on pourrait questionner si l'interdépendance de ressources, qu'elles soient fortement ou faiblement interdépendantes d'un point de vue des connaissances, a suffi à créer une interdépendance positive suffisante dans les groupes. Dans le cadre de ProFan, les comparaisons entre les trois modalités de structuration de l'interdépendance positive devraient nous apporter des premiers éléments de réponse. Cependant, nous pensons que l'observation systématique et l'analyse qualitative des interactions entre individus permettraient de refléter l'émergence des

interactions facilitatrices de manière plus précise (Chi & Wylie, 2014 ; D’Mello et al., 2017 ; Volet et al., 2009).

3.3. La structure de Jigsaw : frein ou levier aux interactions facilitatrices ?

D’après la théorie de l’interdépendance positive, l’interdépendance de ressources dans Jigsaw est une des conditions majeures pour faire émerger les interactions facilitatrices, et donc un meilleur apprentissage (Johnson & Johnson, 2009). En effet, c’est lorsque les individus résolvent une tâche ensemble et atteignent leurs objectifs d’apprentissage, en mettant en commun et en intégrant leurs connaissances complémentaires, que cette interdépendance rend l’apprentissage coopératif efficace (Buchs et al., 2004 ; Johnson & Johnson, 2005). Pourtant, les controverses liées à la structure de Jigsaw, notamment l’absence de récompense de groupe et la présence de buts compétitifs, individualistes dans ses différentes étapes ne sont pas toujours résolues : il est possible que ce soit *la structure même de Jigsaw* qui freine l’émergence d’un apprentissage de meilleure qualité pour les apprenant·es (Roseth et al., 2018 ; Slavin, 1983). L’absence de récompense de groupe dans Jigsaw est une des limites soulevées par plusieurs chercheur·es s’intéressant aux pédagogies coopératives (Newmann & Thompson, 1987 ; Moskowitz et al., 1985 ; Sharan, 1980). Moskowitz et collaborateurs (1985) discutent de cette limite suite à leur expérience longitudinale menée sur plus de 400 élèves de 5^e. Les auteurs n’observent aucun effet positif de la méthode Jigsaw sur les apprentissages comparativement à des groupes contrôles sans coopération, alors que les enseignant·es étaient formés par les expérimentateur·rices au préalable et même dans les établissements où la méthode avait été implémentée de manière exemplaire. Pour ces auteur·es, c’est la structure de la récompense dans Jigsaw qui pose un problème : l’absence de produit de groupe et le fait qu’aucune récompense ne soit donnée sur la base du travail commun diminuent la responsabilité individuelle et la cohésion du groupe, deux facteurs clés dans la théorie de l’interdépendance positive (Moskowitz et al., 1983 ; 1985). Pour elleux, Jigsaw ne serait donc pas une pédagogie purement coopérative, puisque le système de récompense serait plutôt individualiste, voire compétitif, entre les élèves.

Cette affirmation a récemment été reprise par Roseth et collaborateurs (2018), qui apportent des preuves supplémentaires de la structure « *mixte* » de Jigsaw, qui combine à la fois des buts individualistes, compétitifs et coopératifs. D’après les auteur·es, la simple distribution de ressources complémentaires ne suffit pas à générer une coopération efficace et stable dans le temps : sans la présence de buts communs et d’une récompense de groupe, déterminés par la présence d’une évaluation commune d’un produit de groupe, l’interdépendance des ressources

n'est pas une condition suffisante pour faire émerger les interactions facilitatrices entre individus (Johnson & Johnson, 2005). En d'autres termes : « [...] dans Jigsaw, rendre les individus interdépendants en termes de ressources, mais indépendants en termes de résultats (i.e., évaluations de connaissances) aura pour conséquence de faire émerger des comportements coopératifs, mais également compétitifs et individualistes » (Roseth et al., 2018, p.3). Dans leur article, Roseth et collaborateurs (2018) utilisent d'ailleurs une version « révisée » de Jigsaw, qui comprend une récompense de groupe, sous la forme d'un point bonus si tous les membres réussissent leurs évaluations individuelles de manière satisfaisante (i.e., au moins 80% de réponses correctes). Cette restructuration fait directement référence aux critiques de Slavin (1983) qui constatait également le problème que posait l'absence de récompense de groupe dans Jigsaw. Malheureusement, comme pour les autres modérateurs, la littérature scientifique dont nous disposons ne nous permet pas de quantifier l'impact de la présence d'un but commun et/ou d'une récompense de groupe dans les effets de Jigsaw. La méta-analyse ne rapporte par exemple aucune différence entre les versions révisées de Jigsaw, c'est-à-dire celles avec une meilleure structuration de but et de récompense (voir Tableau 4, chapitre 2), et la version initiale telle que proposée par Aronson et collaborateurs (1978). Notre méta-analyse ne couvre cependant qu'un fragment de la littérature Jigsaw, et en conséquence, souffre certainement d'un manque de puissance ou de représentativité des différentes versions de Jigsaw pour effectuer des comparaisons satisfaisantes. On peut par exemple noter que certaines versions de Jigsaw n'ont fait l'objet d'aucune validation empirique (Holliday, 2002), tandis que d'autres ne sont utilisées que dans des contextes extrêmement spécifiques, comme Subject Jigsaw qui n'est déployé que sur des étudiants de l'enseignement supérieur sur des contenus de physique-chimie et en Turquie (Doymus, 2007, 2008a, 2008b).

L'évaluation systématique des modifications apportées à Jigsaw paraît difficile, puisqu'il semble exister autant de versions différentes que d'articles empiriques qui testent leur efficacité. Moskowitz et collaborateurs (1983 ; 1985) rapportent par exemple que même en ayant formé les enseignant·es participant à leur expérience de la même manière, des différences d'implémentation de Jigsaw s'observaient d'un établissement à l'autre. Dans notre cas de figure, nous n'avons pas pu observer ce qu'il se passait réellement à l'intérieur des classes qui participaient à nos expériences. Non seulement les groupes contrôles de nos expériences n'étaient pas purement individualistes (e.g., travail en groupe pour le partage des équipements et des paillasse), mais nous ne pouvons pas non plus garantir avec certitude la qualité de l'implémentation de Jigsaw dans ces classes. Comme pour la proposition précédente,

l'observation des classes peut être une solution pour pallier cet obstacle : à l'aide d'une grille d'évaluation de la qualité d'implémentation de Jigsaw, qui assurerait nous pourrions évaluer à quel le respect de la méthode telle que décrite par Aronson et ses collègues (1978 ; 1997) est relié aux effets sur les apprentissages (voir Robert et al., *en préparation*, pour la description d'un indice d'implémentation de Jigsaw). On pourrait également, comme c'est le cas pour l'expérimentation ProFan, établir un indice d'implémentation de Jigsaw sur la base d'entretiens avec les personnes chargées de vérifier le déroulement de l'expérimentation au sein des établissements. Ou bien, plutôt que de chercher à contrôler les différences et les variations inévitables dans l'implémentation de pédagogies coopératives, on pourrait se focaliser sur ce qu'il se passe concrètement dans les groupes, c'est-à-dire aux processus (voir chapitre 1). En d'autres termes, nous ne chercherions plus à comparer les variations entre les versions de Jigsaw, mais plutôt à comprendre son fonctionnement global, sous-tendu par les interactions structurées par l'interdépendance de ressources entre les membres. Dans le prochain paragraphe, nous développons les avantages de l'approche par élaboration cognitive, centrée sur les processus cognitifs, comparativement à l'approche par cohésion sociale, davantage orientée vers les résultats, dans l'étude des interactions facilitatrices entre les individus.

4. Perspectives et ouvertures

4.1. Jigsaw et l'élaboration cognitive

Comme l'évoquent Roseth et collaborateurs (2018), nous pensons qu'il existe une nécessité de « *repenser la méthode Jigsaw* », c'est-à-dire s'interroger plus spécifiquement sur son fonctionnement et les raisons qui pourraient expliquer l'hétérogénéité des résultats que nous observons dans la littérature. Ce manuscrit entame un processus de description et de qualification de cette hétérogénéité, en contrastant la présence d'études peu rigoureuses et rapportant des effets très larges avec des études plus précises, mais qui rapportent des résultats bien plus modestes. Bien que la présence d'un biais de publication en faveur des résultats positifs soit une première piste d'explication, notamment développée dans les chapitres 3 et 4, il semble illusoire de clore le débat sur cette seule base. Nous pensons que l'examen attentif de l'efficacité de cette méthode ne pourra s'effectuer qu'en étudiant de manière approfondie les mécanismes en jeu lors de l'interaction entre les individus. L'argument que nous souhaitons développer ici est que la focalisation sur les effets, ou les produits ou « *outcomes* », ne permet pas de refléter ce qu'il se passe réellement au sein des groupes. Nous pensons qu'il est nécessaire d'intégrer plus fortement la perspective théorique de l'élaboration cognitive pour

mieux comprendre comment les interactions facilitatrices, telles que décrites par la théorie de l'interdépendance positive, émergent dans Jigsaw. Jusqu'à présent, très peu d'études menées dans la littérature Jigsaw s'intéressent aux processus ayant lieu lors de l'apprentissage coopératif, et comparent principalement des produits (e.g., les scores obtenus à des évaluations de connaissances) entre les groupes expérimentaux pour tester l'efficacité de la méthode (Berger & Hänze, 2009 ; Hänze & Berger, 2007 ; Souvignier & Kronenber, 2007, Roseth et al., 2018). Si d'après la théorie de l'interdépendance positive ce sont les interactions facilitatrices qui permettent l'émergence des effets positifs sur les apprentissages, nous ne disposons pas d'assez d'éléments empiriques permettant de confirmer ou d'infirmer cette relation dans le cadre de Jigsaw.

Chi et Wylie (2014) offrent un modèle théorique qui pourrait être pertinent pour observer, quantifier et prédire ces interactions : le modèle ICAP (i.e., « Interactif, Constructif, Actif, Passif »). D'après ces auteur·es, l'apprentissage coopératif permet d'aboutir au niveau le plus élevé d'élaboration cognitive : l'interaction, ou la co-construction de connaissances (voir aussi Deiglmayr & Schalk, 2015). Les auteur·es définissent la co-construction de connaissances comme un échange mutuel d'idées entre individus, qui résulte en la création d'une idée nouvelle qui n'aurait pas pu voir le jour sans l'interaction entre ces individus. Ce principe est proche de la conception de la tâche « purement coopérative » de Slavin (1983 ; 2016), qui ne peut être résolue que grâce à l'interaction entre les individus. Cet échange d'idée implique un degré d'engagement important dans la tâche, ainsi que la mise en place d'un traitement cognitif dit « profond ». Selon le modèle ICAP, ce niveau d'engagement peut être observé par un certain nombre de comportements cognitifs dits « de haut niveau », par exemple, défendre un argument, expliquer une position, demander une justification, clarifier et/ou développer une idée. Pour Chi et Wylie (2014), un échange n'est interactif que dans la mesure où les individus s'échangent et apportent des contributions substantielles. À l'inverse, un échange « passif » d'informations se caractérise par une élaboration faible et un traitement cognitif superficiel : c'est une des critiques que l'on peut formuler à l'égard du fonctionnement de Jigsaw, notamment dans le cas d'une forte interdépendance de connaissances (Deiglmayr & Schalk, 2015). Plutôt que d'aboutir à une co-construction de connaissances, l'échange d'informations dans la phase puzzle peut être interprété comme étant passif : chaque individu effectue, à tour de rôle, une courte présentation à la manière d'un cours habituel à ses camarades. Outre le problème lié au manque d'élaboration cognitive d'une telle pratique, d'autres chercheur·es ont également relevé la difficulté que représente un tel exercice pour des

élèves. Dans Jigsaw, chaque « expert·e » maîtrise supposément sa sous-partie, mais n'est exposé·e au contenu des autres « expert·e » que pendant un faible laps de temps. En conséquence, la combinaison d'un apprentissage « passif » piloté par des pairs peu compétents et sur une période de temps plus restreinte que celle d'un enseignement classique a de grandes chances de ne pas aboutir à un quelconque bénéfice d'élaboration cognitive (Chi & Wylie, 2014 ; Cohen, 1994 ; Nokes-Malach et al., 2015 ; Pai et al., 2015).

Comme évoqué précédemment, une limite de la littérature repose sur l'utilisation de la mesure pour évaluer l'apprentissage : d'après elleux, il est possible que les évaluations de connaissances ne soient pas suffisamment sensibles aux variations du degré de l'engagement (i.e., passif, actif, constructif ou actif) dans les activités d'apprentissage, puisque ces évaluations ne nécessitent qu'un rappel d'informations qui peut être répondu par une réception « passive » des informations (Chi & Wylie, 2014). Cette piste d'interprétation peut être mise en relation avec les résultats observés dans les expérimentations du chapitre 5 chez les élèves de collèges : peut-être qu'une des raisons pour laquelle nous n'observons pas d'effets de Jigsaw provient du fait que la mesure (i.e., les évaluations de connaissance) ne permet pas de refléter les apprentissages qui ont lieu dans Jigsaw. La plupart des enseignant·es ayant participé à ces expérimentations nous ont rapporté cette limite à travers des échanges informels. Dans le cadre de l'expérimentation 3A, par exemple, l'enseignant en charge des classes ayant participé nous a évoqué le cas de figure d'élèves en difficulté sur les évaluations écrites, alors même qu'ils mettaient en place tous les comportements traduisant un niveau de compétence satisfaisant sur les séquences étudiées. Dans ce cas de figure, leur « mauvaise performance » ne correspond donc pas à la réalité de leurs compétences et de leurs apprentissages. Les enseignant·es des expérimentations 3B et 3C font aussi mention de cette rupture entre l'évaluation censée traduire l'efficacité de Jigsaw et l'émergence de comportements prosociaux et cognitifs au sein de leur classe (e.g., élèves ayant des facilités qui ont pris en charge des élèves en plus grande difficulté, répartition des tâches et des contenus à la manière de Jigsaw lors de travail en coopération non structurée). Pour ces enseignant·es, ces comportements facilitateurs ont émergé grâce à Jigsaw et se sont maintenus voire généralisés à la suite de l'expérimentation, mais n'ont pas été « captés » par les évaluations de connaissances déployées dans nos expériences. Bien que ces informations n'aient qu'une faible valeur « qualitative », puisqu'obtenues hors de notre protocole expérimental, elles apportent un soutien indirect à la théorie de l'interdépendance sociale.

D'après ces observations, Jigsaw pourrait permettre de faire émerger des relations positives entre individus, et en conséquence un climat favorable aux apprentissages. Cependant, même si ces comportements semblent bel et bien être apparus, nous n'avons pas observé de gains sur les apprentissages tels que nous les avons mesurés. Plutôt que de nous intéresser aux notes obtenues à des évaluations de connaissances, on pourrait tester l'efficacité de Jigsaw dans le développement des compétences sociales : par exemple, dans l'expérimentation ProFan, les participant·es ont rempli plusieurs activités permettant de mesurer leur capacité à coopérer, échanger des informations et faire preuve de créativité. Disposer à la fois des évaluations de connaissances, d'informations sur le niveau de structuration de l'interdépendance positive et de mesures « comportementales » de coopération devraient nous permettre de dépasser la limite de nos expériences. De manière plus générale, on pourrait sortir du cadre des performances académiques se reposer sur la *raison d'être* historique de Jigsaw, à savoir le développement des compétences sociales et la réduction des préjugés inter-ethniques (Aronson et al., 1978 ; Aronson & Bridgeman, 1979).

4.2. Retourner aux fondamentaux?

Historiquement, Jigsaw avait pour vocation à réduire les préjugés inter-ethniques et améliorer l'inclusion des minorités Mexicaines et Afro-Américaines dans les écoles américaines, peu après l'annonce politique de leur déségrégation. La mise en place des premières « classes puzzle » est donc une réponse à l'émergence de problèmes sociétaux inédits, et constitue pour Aronson et collaborateurs (1978) en une alternative pratique à la résolution de conflits entre individus. Pourtant, un glissement s'est rapidement opéré dans la littérature scientifique, puisqu'une grande partie des articles réalisés jusqu'à présent mesurent les apprentissages, tels que mesurés par des scores de performances académiques, et non pas la réduction des préjugés inter-ethniques ou le développement des compétences sociales (Robert et al., *en préparation*). D'après ces auteur·es, les attitudes envers les pair·es ($N = 14$, soit 8%) et les compétences sociales ou académiques ($N = 4$, soit 2%) ne représentent qu'une faible portion de la littérature rédigée jusqu'à ce jour. Pourtant, la théorie de l'interdépendance positive formule des hypothèses relativement fortes sur ces variables : on devrait s'attendre à ce que les attitudes envers les pairs s'améliorent et que les compétences sociales soient plus développées une fois que la coopération est mise en place, comparativement à des situations dans lesquelles il n'y a pas de coopération entre les élèves (Johnson & Johnson, 1989 ; 2002, 2009). Une refocalisation sur ces variables permettrait d'apporter plus de preuves empiriques pour valider ou infirmer la théorie de l'interdépendance positive et ainsi remettre à jour l'état

des connaissances sur les effets de Jigsaw sur les variables sociales (Vives et al., *en préparation*). Dans la littérature Jigsaw, très peu d'études se sont finalement intéressées à la réduction des préjugés inter-ethniques *per se* et les résultats de ces études ne permettent pas de conclure à l'efficacité de Jigsaw sur ces variables (voir Bratt, 2008 ; Desforges et al., 1991 ; Walker & Crogan, 1998, chapitre 2). Bratt (2008) note dans son article après n'avoir trouvé aucun support empirique à l'hypothèse de réduction des préjugés inter-ethniques chez ses participant·es : « À en juger par la littérature existante, les preuves de l'efficacité de Jigsaw pour réduire les préjugés ne sont aucunement documentées » (Bratt, 2008, p. 416). Certaines études récentes suggèrent pourtant que l'apprentissage coopératif peut avoir des effets positifs sur la réduction des inégalités scolaires entre élèves, notamment celles liées au statut socio-économique (SSE) des élèves, c'est-à-dire leur origine sociale (Dietrichson et al., 2017 ; Spitzer & Aronson, 2015). En effet, de nombreux travaux soulignent l'importance de la prise en compte du SES dans le développement des comportements, des attitudes et des pensées chez les individus, notamment en milieu scolaire (Goudeau & Croizet, 2016 ; Jury et al., 2017 ; Kraus et al., 2009 ; Piff et al., 2018 ; Stephens et al., 2014). Les inégalités dues au SSE sont largement répertoriées dans la littérature et impactent la réussite des élèves à tous les niveaux scolaires de manière assez importante ($r = .27$, 95% CI [.23 ; .30] dans Sirin, 2005). En France, les écarts dus aux différences de statut sont systématiquement observés dans les rapports PISA de l'OCDE, et plusieurs chercheur·es mettent en évidence l'impact négatif du statut socio-économique dans la réussite des élèves (OCDE, 2018).

Dans la méta-analyse de Dietrichson et collaborateurs (2017), l'apprentissage coopératif est présenté comme un outil efficace pour réduire les écarts de réussite dus au SSE ($\delta = 0.22$, 95 % CI [0.10 ; 0.34]). Pour ces auteur·es, l'apprentissage coopératif peut avoir pour effet de contrebalancer les domaines dans lesquels les élèves issus de bas SSE sont déficitaires, comparativement aux élèves issus de milieux sociaux plus favorisés. Parmi ces domaines, les auteur·es mentionnent le développement cognitif et social, ainsi que le support pédagogique et motivationnel accru. En d'autres termes, la perspective d'utiliser l'apprentissage coopératif pour réduire les inégalités de SSE s'inscrit parfaitement dans la théorie de l'interdépendance positive : si les interactions entre pair·es sont structurées de telle manière à améliorer le soutien social et cognitif, alors les effets positifs de l'apprentissage coopératif peuvent apparaître. Dans nos expériences, le statut socio-économique des élèves n'a pas modéré l'effet de Jigsaw : nous n'avons observé aucune réduction de l'écart dû au SSE entre les élèves suite à la mise en place de la méthode (voir Annexes H, I, J et K). Pourtant, l'hypothèse d'une modération par le statut

reste possible : la perspective du « *cultural mismatch* » suggère que les étudiant·es de bas SES peuvent profiter d'un contexte de coopération, davantage orienté vers les valeurs d'entraide et de communauté (Stephens et al., 2012, 2014 ; Utz, 2004) ». Développée par Nicole Stephens et ses collègues dans le cadre de leurs travaux sur les étudiant·es de « première génération » (i.e., ceux dont les parents n'ont pas fait d'études supérieures) en contexte universitaire, l'hypothèse du « *cultural mismatch* » propose que les étudiant·es de « première génération » souffrent d'une déconnexion entre leur sois « *interdépendants* », orientés vers autrui et la communauté qu'ils développent en se socialisant dans des milieux défavorisés, et les normes « *d'indépendance* » centrée sur la performance et l'individualisme promues à l'université. Cette rupture aurait pour effet de diminuer le sentiment d'appartenance des étudiant·es de première génération, leur bien-être, leur intégration sociale ainsi que leurs performances académiques (Jury et al., 2017 ; Phillips et al., 2016 ; Rubin, 2012 ; Stephens et al., 2012). Dans leur recherche, les chercheur·es sont parvenu·es à réduire cet écart de statut en l'aide d'une simple lettre de bienvenue présentant l'université comme un milieu valorisant l'interdépendance, à travers les interactions avec autrui et le travail d'équipe (Stephens et al., 2012, étude 3). Dans ce cas de figure, les étudiants étaient en « *match* » avec leur contexte. Or, quand la lettre présentait l'université comme un contexte compétitif où la réussite individuelle était la plus valorisée, les chercheurs observaient un écart de réussite dû au statut.

On peut donc supposer que la méthode Jigsaw pourrait avoir des effets positifs chez les individus de bas statut car elle met en place un cadre d'apprentissage coopératif qui repose sur l'interdépendance positive entre les apprenants. Dans une étude récente, Dittman et collaborateurs (2020) apportent un soutien empirique à cette hypothèse. À travers plusieurs expériences ainsi qu'une analyse d'archives, les auteur·es mettent en évidence que les étudiant·es issus de milieux défavorisés réussissent mieux lorsqu'ils travaillent en groupe qu'individuellement. Pour ces auteur·es, l'évaluation individuelle n'est pas neutre, d'un point de vue de la classe sociale, puisqu'elle est congruente avec les valeurs individualistes et le soi indépendant promu à l'université. De la même manière, Dittman et collaborateurs (2020) apportent des éléments qui suggèrent que des évaluations collectives, ainsi que le simple fait de travailler en groupe, peut permettre aux élèves d'origines sociales défavorisées de mieux réussir que dans un contexte individualiste. Cependant, une réserve doit être formulée sur cette hypothèse : si Jigsaw a pour effet d'accroître non seulement la coopération, mais aussi le sentiment de compétition et les efforts individualistes (Roseth et al., 2018), cela pourrait avoir un effet pervers sur la réussite des étudiant·es de bas SSE. L'exacerbation des comparaisons

sociales à l'intérieur du groupe coopératif peut avoir pour conséquence de mettre ces élèves en difficulté (Croizet et al., 2004 ; Dijkstra et al., 2008 ; Monteil & Huguet, 1999 ; Nokes-Malach et al., 2015), particulièrement s'ils sont de bas SSE (Crouzevialle & Darnon, 2019 ; Jury et al., 2017 ; Smeding et al., 2013).

À la vue de ces éléments, il apparaît que les preuves empiriques attestant l'efficacité de Jigsaw sont insuffisantes, que ce soit sur les performances académiques ou bien sur la réduction des préjugés inter-ethniques ou de statut socioéconomique. Le nombre d'articles, et pour certains, leur qualité, ne permet pas, à ce jour d'avoir une idée claire des effets de Jigsaw sur ces variables, ni des conditions dans lesquelles ces effets se produisent ou non. Il nous semble important d'évoquer, à ce propos, un véritable questionnement sur la pertinence et la faisabilité de Jigsaw en tant qu'objet d'étude.

4.3. Mener des études à grande échelle pour détecter les effets de Jigsaw

Au regard de ces éléments, nous pouvons affirmer que l'effet probable de Jigsaw, au moins sur les apprentissages, est très certainement plus petit que notre plus petite taille d'effet d'intérêt, mais également très en dessous du seuil de signification pratique des interventions en psychologie de l'éducation (Gall et al., 1996, cités dans Kyndt et al., 2013). Les éléments statistiques que nous avons développés tout au long de ce manuscrit suggèrent qu'on ne peut pas répondre à la question de l'efficacité de Jigsaw sans prendre en considération la question de l'amplitude de ses effets, et la capacité des recherches futures à les détecter. Une première question relative à la puissance statistique est donc celle des ressources nécessaires pour piloter des expérimentations s'intéressant à de tels effets (voir Tableau 18 ci-dessous).

		Puissance statistique (1-β)		
Taille d'effet	Test effectué	.80	.90	.95
δ = .20	<i>t</i> -test indépendant	788	1054	1302
	<i>t</i> -test apparié	155	215	271
δ = .10	<i>t</i> -test indépendant	3142	4206	5200
	<i>t</i> -test apparié	620	858	1083

Tableau 16: Estimation du nombre total de participant-es en fonction de la puissance statistique du test et de la taille d'effet à détecter (*G*Power*, Faul et al., 2007)

Le Tableau 16 donne plusieurs informations : dans un premier temps, on peut noter le nombre très important de participant·es requis·es dans le cas d'un t -test indépendant : au moins 788 pour une puissance statistique de 80% et une taille d'effet de $\delta = 0.20$, au seuil $\alpha = .05$, et trois plus pour une taille d'effet de $\delta = 0.10$. La mise en place de telles expérimentations nécessite *a minima* des ressources extrêmement importantes, et on pourrait rapporter que sur la littérature Jigsaw, même la taille maximale de l'échantillon rapportée dans une étude ($N = 770$ dans Theobald et al., 2017) est potentiellement trop faible pour détecter un tel effet. Pour rappel, les tailles d'échantillon médianes sont bien plus petites (voir Figure 7, chapitre 2). Outre les difficultés évidentes quant à la récolte d'une telle quantité de données, l'autre observation que l'on peut tirer de ce tableau est la supériorité, en termes de puissance statistique, de l'utilisation d'un modèle apparié, comparativement au modèle du t -test indépendant, qui tend à réduire le nombre de sujets par condition par un facteur proche de cinq. Il va sans dire que nous préconisons l'utilisation d'une telle procédure pour la mise en place d'éventuelles recherches (voir Montoya, 2020 pour une argumentation similaire). Ce tableau suggère implicitement l'idée selon laquelle seuls des essais à grande échelle peuvent disposer d'une puissance de test satisfaisante pour détecter de tels effets : l'expérimentation ProFan devrait offrir des conditions de puissance statistiques favorables à des tests statistiques puissants. Pourtant, le déploiement de telles expérimentations présente aussi des risques. Dans leur article, Lortie-Forgues et Inglis (2019) mettent en évidence que le déploiement d'expérimentations longitudinales à essais randomisés et contrôlés ne permet pourtant pas d'évaluer correctement l'efficacité d'interventions académiques supposées améliorer l'apprentissage : à travers une synthèse de 141 RCTs, soit un échantillon de plus d'un million d'élèves du secondaire, les auteurs observent une taille d'effet extrêmement faible caractérisée par une forte dispersion ($ES = 0.06$, $se = 0.30$). Pour eux, bien que les RCTs disposent *a priori* d'une bonne validité externe, puisqu'ils sont réalisés dans des contextes scolaires véritables, ils ne disposent pas d'une puissance de test suffisante à cause du « bruit expérimental », c'est-à-dire des variations importantes dans les mesures effectuées. Dans l'expérimentation ProFan, plusieurs outils ont été déployés pour quantifier ce bruit, notamment grâce à des grilles d'observation à remplir par les enseignant·es à la fin des séquences pédagogiques et des entretiens semi-directifs menés auprès des responsables du suivi du déroulement de l'expérimentation. Dans ProFan, la mise en place de filtres permettant de mesurer la qualité de l'implémentation de Jigsaw, ainsi que le suivi de l'expérimentation dans les classes devrait permettre de contrôler une partie de ce bruit.

Pour résumer, Lortie-Forgues et Inglis (2019) argumentent que la plupart des RCTs sont inefficaces et non-informatifs car réalisés dans des conditions qui sous-estiment ce bruit expérimental. Les auteurs formulent trois explications possibles à ce problème :

1. La littérature scientifique sur laquelle se basent les chercheur·es pour implémenter les RCTs est biaisée, et donc non crédible.
2. Les interventions sont efficaces, mais les mesures ne permettent pas de détecter les effets d'intérêt.
3. La plupart des interventions déployées sont mal implémentées

Dans le cadre de ce manuscrit de thèse, nous avons principalement avancé l'argument qu'une partie de la littérature scientifique sur Jigsaw est probablement biaisée et surestime les effets de la méthode, c'est-à-dire le premier point soulevé par les auteurs. Nous avons également soulevé l'idée que nos mesures d'apprentissages n'ont pas été adaptées pour capter les effets de Jigsaw. Il est donc tout à fait possible que ces explications soient pertinentes au regard des résultats développés dans cette thèse, même si les échantillons concernés sont bien plus faibles que ceux observés dans des RCTs. En attendant de disposer de preuves empiriques plus solides, on peut finalement regrouper ces réflexions en reconsidérant la signification pratique de Jigsaw. Quand bien même son « effet » existerait et serait de l'ordre de $\delta \pm [0.10 ; 0.20]$, la question de sa pertinence en tant qu'intervention académique se pose alors : *pourquoi ne pas mettre en place une autre intervention avec des chances de réussites supérieures, ou moins d'ambiguïté quant aux résultats observés dans les recherches ?*

On pourrait par exemple proposer des interventions académiques basées sur un nombre plus important de recherches et rapportant des tailles d'effet plus importantes, comme celle du « *peer tutoring* » (Rohrbeck et al., 2003 ; Topping et al., 2017). Le tutorat par les pairs s'inscrit parfaitement dans la perspective socioconstructiviste de la théorie de l'interdépendance positive et de l'élaboration cognitive : les individus sont responsables de leurs apprentissages et de celui de leurs pair·es, et les interactions sont d'autant plus faciles à mesurer puisque les groupes sont restreints à deux individus. Comme présenté dans le chapitre 1, Rohrbeck et collaborateurs (2003) rapportent des effets positifs d'amplitude modérée sur les apprentissages ($d = 0.33$, 95% CI [0.29 ; 0.37]) sur des populations d'élève du secondaire. Sur cette même population, Ginsburg-Block et collaborateurs (2008) observent des effets analogues sur les compétences sociales ($k = 30$; $d = 0.28$, 95% CI [0.25 ; 0.32]), le concept de soi ($k = 15$; $d = 0.18$, 95% CI [0.14 ; 0.23]) et le comportement prosocial ($k = 12$; $d = 0.45$, 95% CI [0.39 ; 0.50]). Dans leur méta-analyse des interventions académiques cherchant à réduire les inégalités dues au SSE,

Dietrichson et collaborateurs (2017) rapportent que le tutorat par les pairs est non seulement l'intervention la plus efficace, mais qu'elle est aussi réalisée sur le plus grand nombre d'études ($k = 36$; $\delta = 0.36$, 95% CI [0.26 ; 0.45]). Pour finir, le tutorat par les pairs figure également parmi les interventions prometteuses et recommandées par Hattie (2009), avec un effet moyen de $\delta = 0.55$, $SE = .103$, obtenu via une synthèse de 14 méta-analyses ($k_{\text{études}} = 767$, $N_{\text{total}} = 2676$), ce qui alimente l'idée selon laquelle les effets du « *peer tutoring* » existent et jouissent d'une pertinence pratique selon les critères standards en psychologie de l'éducation (voir Gall et al., 1996 dans Springer et al., 1999). De manière générale, nous préconisons l'utilisation d'interventions académiques dont les effets sont validés par un nombre important de recherches et qui ne sont pas « trop beaux pour être vrais », c'est-à-dire dont l'amplitude est relativement proche de l'effet médian observé dans la discipline de la psychologie de l'éducation (voir Dunlosky et al., 2013 pour une revue des interventions visant à améliorer la réussite académique des étudiant·es). Jigsaw ne représente qu'une manière parmi d'autres de structurer l'interdépendance positive : les effets positifs de la coopération, à travers l'interdépendance positive, restent validés par un corpus scientifique extrêmement important (Hattie, 2009 ; Johnson & Johnson, 2009).

Les arguments développés dans cette partie ont apporté différentes informations pour reconsidérer 1) les approches théoriques, outils et mesures utilisés pour persévérer dans cette ligne de recherche, 2) l'utilisation du « *proxy* » des performances académiques comme variable dépendante d'intérêt et 3) la pertinence de l'étude de Jigsaw et de ses effets, compte tenu des ressources importantes à déployer et de l'existence d'autres interventions académiques supposément plus fiables et puissantes. Outre ces considérations, nous pensons qu'il est important de conclure ce manuscrit en questionnant l'influence du contexte dans lequel des chercheur·es pourraient implémenter Jigsaw, ou toute autre intervention académique cherchant à promouvoir la coopération. Bien que nous n'obtenions pas d'effet positif de Jigsaw sur les apprentissages, les preuves de l'efficacité de l'apprentissage coopératif, selon la théorie de l'interdépendance positive, restent solides et préconisées pour le développement des apprenants (Hattie, 2009 ; 2012 ; Johnson & Johnson, 2009 ; Slavin, 2016). Dans cette dernière partie, nous développons l'idée selon laquelle l'évaluation de l'efficacité d'une pédagogie coopérative visant à améliorer les apprentissages est *de facto* mise à mal par le contexte scolaire.

5. Améliorer les apprentissages avec la coopération ?

5.1. L'obsession des performances académiques

Depuis les années 2000, l'évaluation des performances des apprenant·es dans le système éducatif Français s'effectue de plus en plus à l'échelle internationale, notamment via les rapports PISA de l'OCDE (2004 ; 2018) qui comparent les résultats d'apprentissage des étudiant·es en lecture, mathématique et en sciences parmi plusieurs pays. Ces résultats ont gagné une importance exponentielle, et servent aujourd'hui dans certains pays à piloter des réformes éducatives. L'objectif de ces rapports est de : « *fournir des études et des preuves comparables au niveau international sur les compétences jugées importantes pour la vie adulte, dans le cadre de la mondialisation et les connaissances émergentes la société* » (OCDE, 2004 ; cité dans Breakspear, 2014). D'après Breakspear (2014), l'importance de ces résultats entraîne un certain nombre de tensions à l'intérieur et entre les pays évalués : la visibilité accrue des pays qui réussissent et/ou échouent exacerbe l'attention des médias, et en conséquence, le rôle des évaluations PISA dans le débat des réformes éducatives. Selon lui, les évaluations PISA ne reflètent pas la complexité des phénomènes d'apprentissage ni la « *performance* » d'un système éducatif, et devraient être davantage réinterprétées au sein du contexte idéologique duquel elles émergent. L'auteur évoque la dérive inquiétante de la confusion entre l'indicateur et le phénomène étudié, c'est-à-dire lorsque la performance aux évaluations PISA devient synonyme de performance du système éducatif. Pour le sociologue de l'éducation Patrick Rayou (2017), le système éducatif français, s'inscrit depuis les années 2000 dans une perspective de compétition avec les autres pays de l'Europe, notamment via « *une logique économique de préparation des ressources humaines qualifiées pour faire face aux défis contemporains, un discours de la qualité visant l'efficience et l'efficacité, une rhétorique de la citoyenneté diversifiant et élargissant les appartenances[...].* » (Rayou, 2017, p.33). De manière plus générale, l'École rencontre des difficultés à promouvoir la coopération : d'après le récent rapport PISA de l'OCDE (2018), les élèves français figurent parmi ceux qui coopèrent le moins dans les classes (45% en moyenne contre 62% pour la moyenne de l'OCDE). De fait, la volonté de promouvoir la coopération semble être en décalage avec la focalisation sur l'évaluation de la réussite individuelle de ses apprenant·es.

Pour Butera et collaborateurs (2006), les systèmes d'éducation tendent à être organisés autour des valeurs de la compétition : les notes, classements et mentions sont autant d'outils permettant de rendre saillante la fonction de sélection et de classification entre les élèves, et avoir des répercussions négatives sur le fonctionnement cognitif et social des individus. Dans

un tel contexte de promotion de la réussite individuelle, il paraît surprenant d’y articuler la coopération : les pratiques pédagogiques coopératives risquent d’entrer en confrontation, ou en « *mismatch* » avec la culture individualiste dans laquelle les élèves sont traditionnellement socialisés (Butera & Buchs, 2019 ; Stephens et al., 2012). Ce *mismatch* est particulièrement flagrant dans les pratiques des enseignant·es au sein de l’institution scolaire : Darnon et collaborateurs (2009) illustrent par exemple le fait que, les enseignant·es incitent et valorisent dans leurs discours la « formation » des élèves, notamment en faisant la promotion des buts orientés vers la maîtrise. Ces buts renvoient à la motivation d’apprendre, d’acquérir une connaissance et de la maîtriser. À l’inverse, la « sélection » des individus s’effectue avec des évaluations individuelles et orientent les élèves vers les buts de performance, qui renvoient à la motivation de réussir mieux que les autres (Elliot & Harackiewicz, 1996). Pour ces auteur·es, c’est la double fonction de sélection et de formation des étudiant·es à l’université qui fait émerger ce *mismatch*. L’évaluation individuelle et les performances académiques (i.e., sous forme d’évaluation « sommative ») seraient donc un outil de sélection et de classification des élèves entre eux, favorisant la compétition et les efforts individualistes (Darnon et al., 2011). Or, dans nos expériences, la « réussite » des participant·es aux évaluations de connaissances était systématiquement mesurée par leurs performances à une évaluation sommative individuelle. Outre le problème de structuration de l’interdépendance positive que pose l’absence d’évaluation de groupe (Slavin, 1983), on peut se demander si cette focalisation sur les performances individuelles n’a pas pour conséquence de contrebalancer les effets positifs de la coopération selon Jigsaw (Roseth et al., 2018). Dit autrement, il y aurait une *incompatibilité* entre la mesure de l’apprentissage selon les critères de réussite individuelle, et la coopération telle que mise en place dans Jigsaw (Butera et al., 2011 ; Dittman et al., 2020).

Pour résumer, on peut formuler deux hypothèses pour situer *a posteriori* le « *mismatch* » entre individualisme et coopération : il peut apparaître dans la structure de Jigsaw, puisqu’elle autorise l’émergence de comportements individualistes et compétitifs, mais également dans le principe même de coopérer alors que la fonction de sélection de l’institution scolaire se base sur la réussite individuelle (Stephens et al., 2012 ; Dittman et al., 2020). Pour résoudre ces conflits, on pourrait tenter d’améliorer la structure de Jigsaw, en suivant les recommandations de Slavin (1980 ; 1983) et de Holliday (2002) et garantir à la fois une évaluation de groupe pour les individus (i.e., interdépendance positive de récompense) ainsi qu’une emphase sur la maîtrise de leurs apprentissages (Guskey, 2007 ; Poortvliet & Darnon, 2014). L’évaluation de groupe permettrait de valoriser la production commune, et éventuellement de créer un

« *match* » avec les valeurs relatives au soi interdépendant (Dittman et al., 2020), ce qui pourrait atténuer la perception de la fonction de sélection de l'école (Croizet et al., 2017). Malgré cette apparente incompatibilité, la plupart des enseignant·es perçoivent l'utilité et sont favorables à l'implémentation des pédagogies coopératives (Slavin, 2011). Plusieurs recherches montrent d'ailleurs qu'il est possible de favoriser l'implémentation de la coopération à l'école, notamment en étudiant les challenges que rencontrent les enseignant·es (Baloche & Brody, 2017) en améliorant l'articulation entre la recherche et la pratique (Burkhardt & Schoenfeld, 2003), mais également en tenant compte des idéologies qui peuvent faciliter ou freiner leur mise en place (Darnon et al., 2018).

5.2. Implémenter efficacement un changement

Plusieurs auteur·es affirment que la mise en place de recherches et d'interventions dans les milieux scolaires doivent passer par une meilleure articulation entre les résultats de travaux scientifiques et leur mise en pratique dans les classes. Pour Baloche et Brody (2017), les chercheur·es ont une responsabilité vis-à-vis de la diffusion des résultats empiriques à l'adresse du corps enseignant : pour elleux, une meilleure formation des enseignant·es aux méthodes coopératives permettrait de dépasser certaines contraintes perçues (e.g., difficulté de mise en place, manque de temps et d'expérience) qui freinent leur application en classe. Au niveau plus local, des incitations à la coopération émergent et rassemblent plusieurs enseignant·es et chercheur·es en science de l'éducation. Dans leur ouvrage « *Osez les pédagogies coopératives au collège et au lycée* », Caron et collaborateurs (2020) mobilisent par exemple les enseignant·es aux classes coopératives. D'autres initiatives comme celle du Collectif « Didactique pour Enseigner » appellent au rapprochement entre la recherche et le « terrain » : « [... une science de la pratique d'enseignement doit reposer selon nous sur la coopération entre professeurs et chercheurs dans la conception, la mise en oeuvre et l'amélioration des dispositifs d'enseignement. Il s'agit donc de penser des ingénieries de la coopération [...] » (Collectif « Didactique pour Enseigner », 2020, p. 174). Pour Burkhardt et Schoenfeld (2003), la mise en place de recherches efficaces sur le terrain doit passer par une meilleure articulation entre les acteur·rices de terrain et les scientifiques : tester l'efficacité d'interventions académiques ne peut donc pas s'effectuer sans une coordination entre chercheur·es et enseignant·es sur le terrain., une volonté également affirmée récemment par le conseil scientifique de l'Éducation Nationale.

Un des apports de ce manuscrit de thèse repose sur la remise en question, au niveau scientifique, de la crédibilité des effets de Jigsaw rapportés dans la littérature : nous avons apporté plusieurs éléments théoriques et empiriques afin de mettre en évidence la disparité entre l'affirmation de son efficacité par ses concepteur·rices et les preuves empiriques qui constituent la littérature scientifique. Nous pensons toutefois que la mise en place de pédagogies coopératives, en général, peut être une plus-value pour le développement d'autres compétences que celles mesurées par les évaluations de connaissances ou les notes sommatives (Guskey, 2007 ; Johnson & Johnson, 2009 ; Noddings, 1995). De manière plus spécifique, la remise en question de « l'outil » Jigsaw est une étape nécessaire pour que les futures recherches soient effectuées dans des conditions méthodologiques plus rigoureuses (Johnson et al., 2000 ; Lakens & Etz, 2017). De cette manière, la qualité des preuves empiriques provenant de la recherche orientée vers les effets sera améliorée, et pourra être communiquée plus efficacement aux acteur·rices du terrain.

6. Conclusion

L'objectif de ce travail de thèse était d'évaluer l'efficacité de Jigsaw, une pédagogie coopérative qui structure l'interdépendance positive à l'aide de ressources complémentaires, sur l'apprentissage. À travers un examen de la littérature scientifique et de ses limites, appuyé par une méta-analyse et plusieurs expériences menées dans des collèges nous ne pouvons pas conclure à l'efficacité de cette méthode sur l'apprentissage des élèves de 6^e, tels que nous les avons mesurés. L'expérimentation ProFan permettra d'apporter davantage de preuves empiriques sur une autre population, celle d'élèves de lycées professionnels, et de spécifier effets de la méthode en testant différents modérateurs de son efficacité.

Nous pensons cependant qu'il est crucial de reconsidérer les preuves de l'efficacité Jigsaw, que ce soit d'un point de vue aussi bien théorique qu'empirique. D'un point de vue théorique, dans un premier temps, car la théorie de l'interdépendance positive ne permet pas de clarifier l'hétérogénéité des résultats que nous observons entre nos expériences et le reste de la littérature Jigsaw. Si Jigsaw est bel et bien « *la moins efficace des pédagogies coopératives* » (Newmann & Thompson, 1987), vouée à l'échec de par sa structure combinant efforts compétitifs, individualistes et coopératifs (Moskowitz et al., 1985 ; Roseth et al., 2018 ; Slavin, 1983), on peut être surpris de voir une multiplication des recherches récentes rapportant des effets larges et positifs de cette méthode (Vives et al., *en préparation*). La reconsidération de la qualité empirique, dans un second temps, nous amène à mieux comprendre cette disparité : la quantité d'articles testant l'efficacité de Jigsaw de manière expérimentale reste relativement

faible et la qualité de ces articles est bien souvent contestable (Robert et al., *en préparation*). Nous avons formulé l'argument que se baser sur ces données pourrait aboutir à un risque considérable de surestimer l'efficacité de la méthode : notre méta-analyse et les expériences menées auprès d'élèves de collèges tendent plutôt à suggérer que la méthode Jigsaw n'entraîne pas des gains « d'intérêt pratique » (Funder & Ozer, 2019 ; Hattie, 2009). Les preuves empiriques provenant de l'expérimentation ProFan permettront de quantifier les effets de Jigsaw de manière plus détaillée.

La contribution principale de ce manuscrit repose donc sur la quantification des effets et de Jigsaw, globalement bien plus faibles que ce qui avait été prévu, et de l'hétérogénéité entre nos études et le reste de la littérature scientifique. Ce travail de thèse permet également de questionner la pertinence pratique de Jigsaw comparativement à d'autres pédagogies ayant fait l'objet d'une validation empirique plus rigoureuse. Enfin, les éléments théoriques et méthodologiques développés dans ce manuscrit soulignent le défi que représente l'évaluation de dispositifs pédagogiques coopératifs. Le contexte scolaire et la fonction que porte l'École, les idéologies associées aux performances individuelles et leur articulation autour des origines des élèves sont autant de paramètres à prendre en considération dans les recherches futures. Nous pensons que le rapprochement entre recherche et terrain, combinées à l'utilisation d'outils méthodologiques puissants et la reconsidération théorique et empirique de Jigsaw permettront d'apporter d'autres précieuses contributions à la littérature scientifique.

Bibliographie

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 78(5), 218-223.

<https://doi.org/10.3200/TCHS.78.5.218-223>

Allport, G. (1954), *The nature of prejudice*. Basic Books.

Arenas, R., & Meirieu, P. (2020). Pour une école de l'entraide et de la coopération. Libération.fr.

Retrieved from https://www.liberation.fr/debats/2020/06/24/pour-une-ecole-de-l-entraide-et-de-la-cooperation_1792136

Aronson, E., Blaney, N., Stephan, C., Sikes, J., & Snapp, M. (1978). *The jigsaw classroom*. Sage.

Aronson, E., & Bridgeman, D. (1979). Jigsaw groups and the desegregated classroom: In pursuit of common goals. *Personality and Social Psychology Bulletin*, 5(4), 438-446.

<https://doi.org/10.1177/014616727900500405>

Aronson, E., & Patnoe, S. (2011). *Cooperation in the classroom: The jigsaw method*. Printer & Martin Limited.

Artut, P. D., & Tarim, K. (2007). The effectiveness of Jigsaw II on prospective elementary school teachers. *Asia-Pacific Journal of Teacher Education*, 35(2), 129-141.

<https://doi.org/10.1080/13598660701268551>

Baloche, L., & Brody, C. M. (2017). Cooperative learning: exploring challenges, crafting innovations. *Journal of Education for Teaching*, 43(3), 274-283. <https://doi.org/10.1080/02607476.2017.1319513>

Baer, M., Leenders, R. T. A. J., Oldham, G. R., & Vadera, A. K. (2010). Win or lose the battle for creativity: The power and perils of intergroup competition. *Academy of Management Journal*, 53(4), 827-845. <https://doi.org/10.5465/amj.2010.52814611>

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Basyah, N. A., Muslem, A., & Usman, B. (2018). The effectiveness of using the jigsaw model to improve students' economics teaching-learning achievement. *The New Educational Review*, 30.

https://www.educationalrev.us.edu.pl/dok/volumes/tner_1_2018b.pdf#page=30

Batdi, V. (2014). JİGSAW TEKNİĞİNİN ÖĞRENCİLERİN AKADEMİK BAŞARILARINA ETKİSİNİN META-ANALİZ YÖNTEMİYLE İNCELENMESİ. [Une méta-analyse des effets de Jigsaw sur le développement des performances académiques]. *EKEV Akademi Dergisi*, 58(58), 699-714

Blaney, N. T., Stephan, C., Rosenfield, D., Aronson, E., & Sikes, J. (1977). Interdependence in the classroom: A field study. *Journal of Educational Psychology*, 69(2), 121.

<https://psycnet.apa.org/doi/10.1037/0022-0663.69.2.121>

- Berger, R., & Hänze, M. (2009). Comparison of two small-group learning methods in 12th-grade physics classes focusing on intrinsic motivation and academic performance. *International Journal of Science Education*, 31(11), 1511-1527. <https://doi.org/10.1080/09500690802116289>
- Berger, R., & Hänze, M. (2015). Impact of expert teaching quality on novice academic performance in the Jigsaw cooperative learning method. *International Journal of Science Education*, 37(2), 294-320. <https://doi.org/10.1080/09500693.2014.985757>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (Eds.). (2010). *Introduction to meta-analysis*. Wiley.
- Bratt, C. (2008). The Jigsaw classroom under test: No effect on intergroup relations evident. *Journal of Community & Applied Social Psychology*, 18(5), 403-419. <https://doi.org/10.1002/casp.946>
- Bressoux, P. (2020). Using multilevel models is not just a matter of statistical adjustment: Illustrations in the educational field. *L'Année psychologique*, 120(1), 5-38. <https://doi.org/10.3917/anpsy1.201.0005>
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., ... & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803-848. <https://doi.org/10.3102/0034654316672069>
- Brown, R. (1986). *Social Psychology* (2nd ed.). Free Press.
- Bridgeman, D. L. (1981). Enhanced role taking through cooperative interdependence: A field study. *Child Development*, 52(4), 1231-1238. <https://doi.org/10.2307/1129511>
- Buchs, C., Butera, F., & Mugny, G. (2004). Resource interdependence, student interactions and performance in cooperative learning. *Educational Psychology*, 24(3), 291-314. <https://doi.org/10.1080/0144341042000211661>
- Buchs, C., & Butera, F. (2009). Is a partner's competence threatening during dyadic cooperative work? It depends on resource interdependence. *European Journal of Psychology of Education*, 24(2), 145-154. <https://doi.org/10.1007/BF03173007>
- Buchs, C., Gilles, I., Antonietti, J.-P., & Butera, F. (2016). Why students need to be prepared to cooperate: A cooperative nudge in statistics learning at university. *Educational Psychology*, 36(5), 956-974. <https://doi.org/10.1080/01443410.2015.1075963>
- Buchs, C., Wiederkehr, V., Filippou, D., Sommet, N., & Darnon, C. (2015). Structured cooperative learning as a means for improving average achievers' mathematical learning in fractions. *Teaching Innovations*, 28(3), 15-35. <https://doi.org/10.5937/inovacije1503015B>

- Buchs, C., & Butera, F. (2015). Cooperative learning and social skills development. In Gillies R. M. (Ed.), *Collaborative Learning: Developments in Research and Practice*. Nova Science. <https://archive-ouverte.unige.ch/unige:84406>
- Burkhardt, H., & Schoenfeld, A. H. (2003). Improving educational research: Toward a more useful, more influential, and better-funded enterprise. *Educational Researcher*, 32(9), 3-14. <https://doi.org/10.3102/0013189X032009003>
- Butera, F., Buchs, C., & Darnon, C. (2011). *L'évaluation, une menace ?* Puf.
- Butera, F., Darnon, C., Buchs, C., & Muller, D. (2006). Les méfaits de la compétition: comparaison sociale et focalisation dans l'apprentissage. In Joule R.-V. & Hugué (Eds.), *Bilans et Perspectives en Psychologie Sociale* (pp. 15-44). Presses Universitaires de Grenoble. <https://archive-ouverte.unige.ch/unige:27270>
- Butera, F., & Buchs, C. (2019). Social interdependence and the promotion of cooperative learning. In K. Sassenberg & M. L. W. Vliek (Eds.), *Social Psychology in Action* (pp. 111-127). Springer International Publishing. https://doi.org/10.1007/978-3-030-13788-5_8
- Chambers, C. (2019). The registered reports revolution Lessons in cultural reform. *Significance*, 16(4), 23-27. <https://doi.org/10.1111/j.1740-9713.2019.01299.x>
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology*, 65(3), 145-153. <https://doi.org/10.1037/h0045186>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155. <https://psycnet.apa.org/doi/10.1037/0033-2909.112.1.155>
- Cohen, E.G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, 64(1), 1-35. <https://doi.org/10.3102/00346543064001001>
- Collectif "Didactique pour Enseigner" (2020). *Enseigner, ça s'apprend*. Retz.
- Connolly, P., Keenan, C., & Urbanska, K. (2018). The trials of evidence-based practice in education: A systematic review of randomised controlled trials in education research 1980-2016. *Educational Research*, 60(3), 276-291. <https://doi.org/10.1080/00131881.2018.1493353>
- Crone, T. S., & Portillo, M. C. (2013). Jigsaw variations and attitudes about learning and the self in cognitive psychology. *Teaching of Psychology*, 40(3), 246-251. <https://doi.org/10.1177/0098628313487451>
- Croizet, J.-C., Després, G., Gauzins, M.-E., Hugué, P., Leyens, J.-P., & Méot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin*, 30(6), 721-731. <https://doi.org/10.1177/0146167204263961>

- Croizet, J.-C., Goudeau, S., Marot, M., & Millet, M. (2017). How do educational contexts contribute to the social class achievement gap: documenting symbolic violence from a social psychological point of view. *Current Opinion in Psychology*, 18, 105-110. <https://doi.org/10.1016/j.copsyc.2017.08.025>
- Darnon, C., Smedig, A., Toczek-Capelle, M. & Souchal, C. (2011). Chapitre 11 : L'évaluation comme outil de formation et/ou de sélection. In Butera F. (Eds.), *L'évaluation, une menace* (pp. 115-125). Presses Universitaires de France. <https://doi.org/10.3917/puf.darno.2011.01.0115>
- Darnon, C., Buchs, C., & Desbar, D. (2012). The Jigsaw technique and self-efficacy of vocational training students: A practice report. *European Journal of Psychology of Education*, 27(3), 439-449. <https://doi.org/10.1007/s10212-011-0091-4>
- Darnon, C., Smeding, A., & Redersdorff, S. (2018). Belief in school meritocracy as an ideological barrier to the promotion of equality. *European Journal of Social Psychology*, 48(4), 523-534. <https://doi.org/10.1002/ejsp.2347>
- Darnon, C., Wiederkehr, V., Dompnier, B., & Martinot, D. (2018). 'Where there is a will, there is a way': Belief in school meritocracy and the social-class achievement gap. *British Journal of Social Psychology*, 57(1), 250-262. <https://doi.org/10.1111/bjso.12214>
- Deci, E. L., & Ryan, R. M. (1985). *Motivation and self-determination in human behavior*. Plenum Publishing Co.
- Deiglmayr, A., & Schalk, L. (2015). Weak versus strong knowledge interdependence: A comparison of two rationales for distributing information among learners in collaborative learning settings. *Learning and Instruction*, 40, 69-78. <https://10.1016/j.learninstruc.2015.08.003>
- Delacre, M., Leys, C., Mora, Y. L., & Lakens, D. (2019). Taking parametric assumptions seriously: Arguments for the use of Welch's *F*-test instead of the classical *F*-test in One-Way ANOVA. *International Review of Social Psychology*, 32(1), 13. <https://doi.org/10.5334/irsp.198>
- Deutsch, M. (1949). A theory of co-operation and competition. *Human Relations*, 2(2), 129-152. <https://doi.org/10.1177/001872674900200204>
- Desforges, D. M., Lord, C. G., Ramsey, S. L., Mason, J. A., Van Leeuwen, M. D., West, S. C., & Lepper, M. R. (1991). Effects of structured cooperative contact on changing negative attitudes toward stigmatized social groups. *Journal of Personality and Social Psychology*, 60(4), 531. <https://doi.org/10.1037/0022-3514.60.4.531>
- DeVries, D. L., Edwards, K. J., & Slavin, R. E. (1978). Biracial learning teams and race relations in the classroom: Four field experiments using Teams-Games-Tournament. *Journal of Educational Psychology*, 70(3), 356-362. <https://psycnet.apa.org/doi/10.1037/0022-0663.70.3.356>
- Dewey, J. (1916). *Democracy and education*. Macmillan.

- Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A.-M. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-analysis. *Review of Educational Research*, 87(2), 243-282. <https://doi.org/10.3102/0034654316687036>
- Di Stasio, M. R., Savage, R., & Burgos, G. (2016). Social comparison, competition and teacher-student relationships in junior high school classrooms predicts bullying and victimization. *Journal of Adolescence*, 53, 207-216. <https://doi.org/10.1016/j.adolescence.2016.10.002>
- Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A. P., & van der Zee, Y. G. (2008). Social comparison in the classroom: A Review. *Review of Educational Research*, 78(4), 828-879. <https://doi.org/10.3102/0034654308321210>
- Dittmann, A. G., Stephens, N. M., & Townsend, S. S. M. (2020). Achievement is not class-neutral: Working together benefits people from working-class contexts. *Journal of Personality and Social Psychology*, 119(3), 517-539. <https://doi.org/10.1037/pspa0000194>
- D'Mello, S., Dieterle, E., & Duckworth, A. (2017). Advanced, Analytic, Automated (AAA) measurement of engagement during learning. *Educational Psychologist*, 52(2), 104-123. <https://doi.org/10.1080/00461520.2017.1281747>
- Dori, Y. J., Yeroslavski, O., & Lazarowitz, R. (1995). *The effects of teaching the cell topic using the Jigsaw method on students' achievement and learning activity*. [Paper presentation]. Annual National Association for Research in Science Teaching 68th conference, San Francisco, CA.
- Doymus, K. (2007). Effects of a cooperative learning strategy on teaching and learning phases of matter and one-component phase diagrams. *Journal of Chemical Education*, 84(11), 1857-1860. <https://doi.org/10.1021/ed084p1857>
- Doymus, K. (2008a). Teaching chemical bonding through jigsaw cooperative learning. *Research in Science & Technological Education*, 26(1), 47-57. <https://doi.org/10.1080/02635140701847470>
- Doymus, K. (2008b). Teaching chemical equilibrium with the jigsaw technique. *Research in science Education*, 38(2), 249-260. <https://doi.org/10.1007/s11165-007-9047-8>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4-58. <https://doi.org/10.1177/1529100612453266>
- Enquête PériODique sur l'Enseignement. (2018). *EPODE: Enquête Périodique sur l'Enseignement*. <https://www.education.gouv.fr/epode-enquete-periodique-sur-l-enseignement-2018-6881>

- Faul, F., Erdfelder, E., Lang, A. et al. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39, 175-191 (2007).
<https://doi.org/10.3758/BF03193146>
- Fanelli, D. (2010). “Positive” results increase down the hierarchy of the sciences. *PLoS ONE*, 5(4), e10068. <https://doi.org/10.1371/journal.pone.0010068>
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3), 891-904. <https://doi.org/10.1007/s11192-011-0494-7>
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502-1505. <https://doi.org/10.1126/science.1255484>
- Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science’s aversion to the null. *Perspectives on Psychological Science*, 7(6), 555-561. <https://doi.org/10.1177/1745691612459059>
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156-168. <https://doi.org/10.1177/2515245919847202>
- Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction*. Longman Publishing.
- Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, 5(2), 189-211. <https://doi.org/10.1080/19345747.2011.618213>
- Ghaith, G., & El-Malak, M. A. (2004). Effect of Jigsaw II on literal and higher order EFL reading comprehension. *Educational Research and Evaluation*, 10(2), 105-115. <https://doi.org/10.1076/edre.10.2.105.27906>
- Gillies, R. M. (2003). Structuring cooperative group work in classrooms. *International Journal of Educational Research*, 39(1-2), 35-49. [https://doi.org/10.1016/S0883-0355\(03\)00072-7](https://doi.org/10.1016/S0883-0355(03)00072-7)
- Gillies, R. M., & Ashman, A. F. (2003). *Co-operative learning: The social and intellectual outcomes of learning in groups*. Routledge Falmer.
- Gilbert, S. (2001, March 27). *A CONVERSATION WITH/Elliot Aronson; No One Left to Hate: Averting Columbines*. The New York Times. <https://www.nytimes.com/2001/03/27/health/a-conversation-with-elliott-aronson-no-one-left-to-hate-averting-columbines.html>
- Göcer, A. (2010). A comparative research on the effectivity of cooperative learning method and Jigsaw technique on teaching literary genres. *Educational Research and Reviews*, 5(8), 439-445.

- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, 10(10), 535-549. <https://doi.org/10.1111/spc3.12267>
- Gömlüksü, M. N. (2007). Effectiveness of cooperative learning (Jigsaw II) method in teaching English as a foreign language to engineering students. *European Journal of Engineering Education*, 32(5), 613-625. <https://doi.org/10.1080/03043790701433343>
- Goodman, S. (2008). A dirty dozen: Twelve p-value misconceptions. *Seminars in Hematology*, 45(3), 135-140. <https://doi.org/10.1053/j.seminhematol.2008.04.003>
- Goudeau, S., & Croizet, J.-C. (2017). Hidden advantages and disadvantages of social class: How classroom settings reproduce social inequality by staging unfair comparison. *Psychological Science*, 28(2), 162-170. <https://doi.org/10.1177/0956797616676600>
- Gurtner, J.-L., Mauroux, L., Hitz, A., de Rocha, B., & Schumacher, J. (2009). *Learning in multiple contexts: Intra and transcontextual effects on the evolution of the learner's motivation*. [Paper presentation]. European Association for Research on Learning and Instruction biennial conference, Amsterdam.
- Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle p-value generates irreproducible results. *Nature Methods*, 12(3), 179-185. <https://doi.org/10.1038/nmeth.3288>
- Hänze, M., & Berger, R. (2007). Cooperative learning, motivational effects, and student characteristics: An experimental study comparing cooperative learning and direct instruction in 12th grade physics classes. *Learning and Instruction*, 17(1), 29-41. <https://doi.org/10.1016/j.learninstruc.2006.11.004>
- Harackiewicz, J. M., Barron, K. E., & Elliot, A. J. (1998). Rethinking achievement goals: When are they adaptive for college students and why? *Educational Psychologist*, 33(1), 1-21. https://doi.org/10.1207/s15326985ep3301_1
- Hattie, J. (2009). *Visible learning: a synthesis of 800+ meta-analyses on achievement*. Abingdon.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.
- Hattie, J. (2015). The applicability of Visible Learning to higher education. *Scholarship of Teaching and Learning in Psychology*, 1(1), 79. <https://doi.org/10.1037/stl0000021>
- Hedges, L. V., & Olkin, I. (1980). Vote-counting methods in research synthesis. *Psychological bulletin*, 88(2), 359. <https://doi.org/10.1037/0033-2909.88.2.359>
- Hofstede, G. H. (1980). *Culture's Consequences: International Differences in Work-Related Values* (integral ed.). Sage.

- Holliday, D. C. (2002). *Jigsaw IV: Using student/teacher concerns to improve Jigsaw III*. (ED465687). ERIC. <https://eric.ed.gov/?id=ED465687>
- Hooper, S., & Hannafin, M. J. (1988). Cooperative CBI: The effects of heterogeneous versus homogeneous grouping on the learning of progressively complex concepts. *Journal of Educational Computing Research*, 4(4), 413-424. <https://doi.org/10.2190%2FT26C-3FTH-RNYP-TV30>
- Hornby, G. (2009). The effectiveness of cooperative learning with trainee teachers. *Journal of Education for Teaching*, 35(2), 161-168. <https://doi.org/10.1080/02607470902771045>
- Huang, T. C., Huang, Y. M., & Yu, F. Y. (2011). Cooperative weblog learning in higher education: Its facilitating effects on social interaction, time lag, and cognitive load. *Journal of Educational Technology & Society*, 14(1), 95-106. <https://www.jstor.org/stable/jeductechsoci.14.1.95>
- Huang, Y. M., Liao, Y. W., Huang, S. H., & Chen, H. C. (2014). A Jigsaw-based cooperative learning approach to improve learning outcomes for mobile situated learning. *Journal of Educational Technology & Society*, 17(1), 128-140. <https://www.jstor.org/stable/jeductechsoci.17.1.128>
- Huguet, P., Charbonnier, E., & Monteil, J.M. (1999). Productivity loss in performance groups: People who see themselves as average do not engage in social loafing. *Group Dynamics: Theory, Research, and Practice*, 3, 118-131. <https://doi.org/10.1037/1089-2699.3.2.118>
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Ioannidis, J. P., & Trikalinos, T. A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *Cmaj*, 176(8), 1091-1096. <https://doi.org/10.1503/cmaj.060410>
- Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research*. Interaction Book Company.
- Johnson, D. W., Johnson, R. T., & Taylor, B. (1993). Impact of cooperative and individualistic learning on high-ability students' achievement, self-esteem, and social acceptance. *The Journal of Social Psychology*, 133(6), 839-844. <https://doi.org/10.1080/00224545.1993.9713946>
- Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). *Cooperative learning methods: A meta-analysis*. Researchgate. https://www.researchgate.net/publication/220040324_Cooperative_learning_methods_A_meta-analysis
- Johnson, D. W., & Johnson, R. T. (2002). Cooperative learning and social interdependence theory. In *Theory and research on small groups* (pp. 9-35). Springer.
- Johnson, D. W., & Johnson, R. T. (2005). New developments in social interdependence theory. *Genetic, Social, and General Psychology Monographs*, 131(4), 285-358. <https://doi.org/10.3200/MONO.131.4.285-358>

- Johnson, D. W., Johnson, R. T., & Smith, K. (2007). The state of cooperative learning in postsecondary and professional settings. *Educational Psychology Review*, 19(1), 15-29. <https://doi.org/10.1007/s10648-006-9038-8>
- Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, 38(5), 365-379. <https://doi.org/10.3102/0013189X09339057>
- Jury, M., Smeding, A., Stephens, N. M., Nelson, J. E., Aelenei, C., & Darnon, C. (2017). The experience of low-SES students in higher education: Psychological barriers to success and interventions to reduce social-class inequality. *Journal of Social Issues*, 73(1), 23-41. <https://doi.org/10.1111/josi.12202>
- Kao, G. Y. M. (2013). Enhancing the quality of peer review by reducing student “free riding”: Peer assessment with positive interdependence. *British Journal of Educational Technology*, 44(1), 112-124. <https://doi.org/10.1111/j.1467-8535.2011.01278.x>
- Karacop, A., & Doymus, K. (2013). Effects of Jigsaw cooperative learning and animation techniques on students’ understanding of chemical bonding and their conceptions of the particulate nature of matter. *Journal of Science Education and Technology*, 22(2), 186-203. <https://doi.org/10.1007/s10956-012-9385-9>
- Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, 65(4), 681-706. <https://doi.org/10.1037/0022-3514.65.4.681>
- Kitayama, S., Markus, H. R., Matsumoto, H., & Norasakkunkit, V. (1997). Individual and collective processes in the construction of the self: Self-enhancement in the United States and self-criticism in Japan. *Journal of personality and social psychology*, 72(6), 1245-1267. <https://doi.org/10.1037/0022-3514.72.6.1245>
- Koc, Y., Doymuş, K., Karaçöp, A., & Şimşek, Ü. (2010). The effects of two Cooperative learning strategies on the teaching and learning of the topics of chemical kinetics. *Journal of Turkish Science Education (TUSED)*, 7(2), 52-.65. <http://www.pegem.net/dosyalar/dokuman/124768-2011090212498-4.pdf>
- Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample Size. *PLOS ONE*, 9(9), e105825. <https://doi.org/10.1371/journal.pone.0105825>
- Kyndt, E., Raes, E., Lismont, B., Timmers, F., Cascallar, E., & Dochy, F. (2013). A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educational Research Review*, 10, 133-149. <https://doi.org/10.1016/j.edurev.2013.02.002>

- Lakens, D., & Evers, E. R. (2014). Sailing from the seas of chaos into the corridor of stability: Practical recommendations to increase the informational value of studies. *Perspectives on psychological science*, 9(3), 278-292. <https://doi.org/10.1177%2F1745691614528520>
- Lakens, D. (2017). Equivalence tests: A practical primer for *t*-tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355-362. <https://doi.org/10.1177/1948550617697177>
- Lakens, D., & Etz, A. J. (2017). Too true to be bad: When sets of studies with significant and nonsignificant findings are probably true. *Social Psychological and Personality Science*, 8(8), 875-881. <https://doi.org/10.1177%2F1948550617693058>
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259-269. <https://doi.org/10.1177/2515245918770963>
- Lazarowitz, R., Hertz-Lazarowitz, R., & Baird, J. H. (1994). Learning science in a cooperative setting: Academic achievement and affective outcomes. *Journal of Research in Science Teaching*, 31(10), 1121-1131. <https://doi.org/10.1002/tea.3660311006>
- Law, Y.-K. (2011). The effects of cooperative learning on enhancing Hong Kong fifth graders' achievement goals, autonomous motivation and reading proficiency: Effects of cooperative learning. *Journal of Research in Reading*, 34(4), 402-425. <https://doi.org/10.1111/j.1467-9817.2010.01445.x>
- Leek, J. T., & Peng, R. D. (2015). Statistics: p-values are just the tip of the iceberg. *Nature*, 520(7549), 612-612. <https://doi.org/10.1038/520612a>
- Le Monde. (2018, September 26). *J'utilise les "groupes puzzles" pour que mes élèves élaborent le cours tous ensemble* ». Le Monde. https://www.lemonde.fr/campus/article/2018/09/26/j-utilise-les-groupes-puzzles-pour-que-mes-eleves-elaborent-le-cours-tous-ensemble_5360618_4401467.html
- Monteil, J.M., & Huguet, P. (1999). *Social context and cognitive performance: Towards a social psychology of cognition*. Psychology Press.
- Lewin, K. (1936). A dynamic theory of personality: Selected papers. *The Journal of Nervous and Mental Disease*, 84(5), 612-613.
- Lewin, K. (1942). Field theory and learning. In N. B. Henry (Ed.), *The forty-first yearbook of the National Society for the Study of Education: Part 2, The psychology of learning* (p. 215-242). University of Chicago Press. <https://doi.org/10.1037/11335-006>
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764-766. <https://doi.org/10.1016/j.jesp.2013.03.013>

- Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *International Review of Social Psychology*, 32(1). <https://doi.org/10.5334/irsp.289>
- Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher*, 48(3), 158-166. <https://doi.org/10.3102/0013189X19832850>
- Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, 66(4), 423-458. <https://doi.org/10.3102/00346543066004423>
- Lucker, G. W., Rosenfield, D., Sikes, J., & Aronson, E. (1976). Performance in the interdependent classroom: A field study. *American Educational Research Journal*, 13(2), 115-123. <https://doi.org/10.3102/00028312013002115>
- Maassen, E., van Assen, M. A. L. M., Nuijten, M. B., Olsson-Collentine, A., & Wicherts, J. M. (2020). Reproducibility of individual effect sizes in meta-analyses in psychology. *PLOS ONE*, 15(5), e0233107. <https://doi.org/10.1371/journal.pone.0233107>
- Mattingly, R. M., & VanSickle, R. L. (1991). Cooperative learning and achievement in social studies: Jigsaw II. *Social Education*, 55(6), 392-395. <https://eric.ed.gov/?id=ED348267>
- Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, 9(2), 147-163. <https://psycnet.apa.org/doi/10.1037/1082-989X.9.2.147>
- Montoya, A. (2020). *The Power of Design: Impact of Experimental Design Outweighs Impact of Inferential Methods on Statistical Power to Detect Indirect Effects*. PsyArXiv. <https://osf.io/3njw6/>
- Moreno, R. (2009). Constructing knowledge with an agent-based instructional program: A comparison of cooperative and individual meaning making. *Learning and Instruction*, 19(5), 433-444. <https://doi.org/10.1016/j.learninstruc.2009.02.018>
- Muthukrishna, M., Bell, A., Henrich, J., Curtin, C., Gedranovich, A., McInerney, J., & Thue, B. (2018). *Beyond WEIRD psychology: Measuring and mapping scales of cultural and psychological distance*. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.3259613>
- Moskowitz, J. M., Malvin, J. H., Schaeffer, G. A., & Schaps, E. (1983). Evaluation of a cooperative learning strategy. *American Educational Research Journal*, 20(4), 687-696. <https://doi.org/10.2307/1162822>

- Moskowitz, J. M., Malvin, J. H., Schaeffer, G. A., & Schaps, E. (1985). Evaluation of Jigsaw, a cooperative learning technique. *Contemporary Educational Psychology*, *10*(2), 104-112. [https://doi.org/10.1016/0361-476X\(85\)90011-6](https://doi.org/10.1016/0361-476X(85)90011-6)
- Nebel, S., Schneider, S., Beege, M., Kolda, F., Mackiewicz, V., & Rey, G. D. (2017). You cannot do this alone! Increasing task interdependence in cooperative educational videogames to encourage collaboration. *Educational Technology Research and Development*, *65*(4), 993-1014. <https://doi.org/10.1007/s11423-017-9511-8>
- Newmann, F. M., & Thompson, J. A. (1987). *Effects of cooperative learning on achievement in secondary schools: A summary of research* (ED288853). ERIC. <https://eric.ed.gov/?id=ED288853>
- Noddings, N. (1995). *Philosophy of education*. Boulder, CO.
- Nokes-Malach, T. J., Richey, J. E., & Gadgil, S. (2015). When is it better to learn together? Insights from research on collaborative learning. *Educational Psychology Review*, *27*(4), 645-656. <https://doi.org/10.1007/s10648-015-9312-8>
- Nolan, J. M., Hanley, B. G., DiVietri, T. P., & Harvey, N. A. (2018). She who teaches learns: Performance benefits of a jigsaw activity in a college classroom. *Scholarship of Teaching and Learning in Psychology*, *4*(2), 93-104. <https://doi.org/10.1037/stl0000110>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600-2606. <https://doi.org/10.1073/pnas.1708274114>
- Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., van 't Veer, A. E., & Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences*, *23*(10), 815-818. <https://doi.org/10.1016/j.tics.2019.07.009>
- OECD. (2018). *Equity in education: Breaking down barriers to social mobility*. OECD. <https://doi.org/10.1787/9789264073234-en>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, *128*(1), 3-72. <https://doi.org/10.1037/0033-2909.128.1.3>
- Pai, H.-H., Sears, D. A., & Maeda, Y. (2015). Effects of small-group learning on transfer: A meta-analysis. *Educational Psychology Review*, *27*(1), 79-102. <https://doi.org/10.1007/s10648-014-9260-8>

- Parsazadeh, N., Ali, R., & Rezaei, M. (2018). A framework for cooperative and interactive mobile learning to improve online information evaluation skills. *Computers & Education*, *120*, 75-89. <https://doi.org/10.1016/j.compedu.2018.01.010>
- Pek, J., & Flora, D. B. (2018). Reporting effect sizes in original psychological research: A discussion and tutorial. *Psychological Methods*, *23*(2), 208-225. <http://dx.doi.org/10.1037/met0000126>
- Piaget, J. (1926). *The language and thought of the child*. New York.
- Pianta, R. C., Belsky, J., Houts, R., & Morrison, F. (2007). Opportunities to learn in America's elementary classrooms. *Science*, *315*(5820), 1795-1796. <https://doi.org/10.1126/science.1139719>
- Protogerou, C., & Hagger, M. S. (2020). A checklist to assess the quality of survey studies in psychology. *Methods in Psychology*, *3*, 100031. <https://doi.org/10.1016/j.metip.2020.100031>
- Puzio, K., & Colby, G. T. (2013). Cooperative learning and literacy: A meta-analytic review. *Journal of Research on Educational Effectiveness*, *6*(4), 339-360. <https://doi.org/10.1080/19345747.2013.775683>
- Qin, Z., Johnson, D. W., & Johnson, R. T. (1995). Cooperative versus competitive efforts and problem solving. *Review of Educational Research*, *65*(2), 129-143. <https://doi.org/10.3102%2F00346543065002129>
- Rayou, P. (2017). *Sociologie de l'éducation*. Que sais-je.
- Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, *95*(2), 240-257. <https://doi.org/10.1037/0022-0663.95.2.240>
- Robert, A., Vives E., Poletti, C., Demolliens, M., Régner I., Huguet P. & Consortium ProFan. *La classe puzzle : un examen critique* [Manuscript in preparation]. Laboratoire de Psychologie Sociale et Cognitive, Université Clermont Auvergne.
- Roseth, C. J., Lee, Y., & Saltarelli, W. A. (2018). Reconsidering Jigsaw social psychology: Longitudinal effects on social interdependence, sociocognitive conflict regulation, motivation, and achievement. *Journal of Educational Psychology*. *111*(1), 149-169. <https://doi.org/10.1037/edu0000257>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638-641. <https://psycnet.apa.org/doi/10.1037/0033-2909.86.3.638>

- Sahin, A. (2010). Effects of Jigsaw II technique on academic achievement and attitudes to written expression course. *Educational Research and Reviews*, 5(12), 777-787.
<https://doi.org/10.5897/ERR.9000131>
- Şahin, A. (2011). Effects of Jigsaw III technique on achievement in written expression. *Asia Pacific Education Review*, 12(3), 427-435. <https://doi.org/10.1007/s12564-010-9135-8>
- Scheel, A. M. (2019). *Positive result rates in psychology: Registered Reports compared to the conventional literature*. ZPID. <https://doi.org/10.23668/psycharchives.2390>
- Schwartz, S. H. (2006). A theory of cultural value orientations: Explication and applications. *Comparative sociology*, 5(2), 137-182. <https://doi.org/10.1163/156913306778667357>
- Shaaban, K. (2006). An initial study of the effects of cooperative learning on reading comprehension, vocabulary acquisition, and motivation to read. *Reading Psychology*, 27(5), 377-403.
<https://doi.org.10.1080/02702710600846613>
- Sharan, S. (1980). Cooperative learning in small groups: Recent methods and effects on achievement, attitudes, and ethnic relations. *Review of Educational Research*, 50(2), 241-271.
<https://doi.org/10.3102/00346543050002241>
- Sherif, M. (1961). *Conformity-Deviation, Norms, and Group Relations*. In I. A. Berg & B. M. Bass (Eds.), *Conformity and deviation* (p. 159-198). Harper and Brothers. <https://doi.org/10.1037/11122-006>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11), 1359-1366. <https://doi.org/10.1177/0956797611417632>
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2), 534-547. <https://doi.org/10.1037/a0033242>
- Slavin, R. E. (1980). Cooperative learning in teams: State of the art. *Educational Psychologist*, 15(2), 93-111. <https://doi.org/10.1080/00461528009529219>
- Slavin, R. E., & Karweit, N. L. (1981). Cognitive and affective outcomes of an intensive student team learning experience. *Journal of Experimental Education*, 50(1), 29-35.
<https://doi.org/10.1080/00220973.1981.11011797>
- Slavin, R. E. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin*, 94(3), 429-445. <https://doi.org/10.1037/0033-2909.94.3.429>
- Slavin, R. E. (1985). Cooperative Learning: Applying contact theory in desegregated schools. *Journal of Social Issues*, 41(3), 45-62. <https://doi.org/10.1111/j.1540-4560.1985.tb01128.x>

- Slavin, R. E. (2011). Instruction based on cooperative learning. In Mayer A. E. & Alexander P. A. (Eds.), *Handbook of Research on Learning and Instruction* (2nd edition., pp. 388-405). Routledge.
- Slavin, R. E. (2013). Cooperative Learning and Student Achievement. In R. E. Slavin (Ed.), *School and Classroom Organization* (0 ed., pp. 129-158). Routledge.
<https://doi.org/10.4324/9780203056950-5>
- Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society open science*, 3(9), 1-17. <https://doi.org/10.1098/rsos.160384>
- Sotola, L. K., & Crede, M. (2020). Regarding class quizzes: A meta-analytic synthesis of studies on the relationship between frequent low-stakes testing and class performance. *Educational Psychology Review*, 1-20. <https://doi.org/10.1007/s10648-020-09563-9>
- Souvignier, E., & Kronenberger, J. (2007). Cooperative learning in third graders' Jigsaw groups for mathematics and science with and without questioning training. *British Journal of Educational Psychology*, 77(4), 755-771. <https://doi.org/10.1348/000709906X173297>
- Spitzer, B., & Aronson, J. (2015). Minding and mending the gap: Social psychological interventions to reduce educational disparities. *British Journal of Educational Psychology*, 85(1), 1-18.
<https://doi.org/10.1111/bjep.12067>
- Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, 69(1), 21-51. <https://doi.org/10.3102/00346543069001021>
- Stahl, R. J., & VanSickle, R. L. (1992). *Cooperative Learning in Social Studies Classroom, An Introduction to Social Study* (ED361243). ERIC. <https://eric.ed.gov/?id=ED361243>
- Stahl, R. J. (1994). *Cooperative Learning in Social Studies: Hand Book for Teacher*. USA Kane.
- Stanne, M. B., Johnson, D. W., & Johnson, R. T. (1999). Does competition enhance or inhibit motor performance: A meta-analysis. *Psychological Bulletin*, 125(1), 133-154. <https://doi.org/10.1037/0033-2909.125.1.133>
- Stephens, N. M., Fryberg, S. A., Markus, H. R., Johnson, C. S., & Covarrubias, R. (2012). Unseen disadvantage: How American universities' focus on independence undermines the academic performance of first-generation college students. *Journal of Personality and Social Psychology*, 102(6), 1178-1197. <https://doi.org/10.1037/a0027143>
- Stephens, N. M., Townsend, S. S. M., Markus, H. R., & Phillips, L. T. (2012). A cultural mismatch: Independent cultural norms produce greater increases in cortisol and more negative emotions among first-generation college students. *Journal of Experimental Social Psychology*, 48(6), 1389-1393.
<https://doi.org/10.1016/j.jesp.2012.07.008>

- Stephens, N. M., Markus, H. R., & Phillips, L. T. (2014). Social class culture cycles: How three gateway contexts shape selves and fuel inequality. *Annual Review of Psychology*, 65(1), 611-634. <https://doi.org/10.1146/annurev-psych-010213-115143>
- Streiner, D. L., & Norman, G. R. (2011). Correction for multiple testing: is there a resolution? *Chest*, 140(1), 16-18. <https://doi.org/10.1378/chest.11-0523>
- Tarhan, L., & Sesen, B. A. (2012). Jigsaw cooperative learning: acid-base theories. *Chemistry Education Research and Practice*, 13(3), 307-313. <https://doi.org/10.1039/C2RP90004A>
- Tarhan, L., Ayyıldız, Y., Ogunc, A., & Sesen, B. A. (2013). A Jigsaw cooperative learning application in elementary science and technology lessons: Physical and chemical changes. *Research in Science & Technological Education*, 31(2), 184-203. <https://doi.org/10.1080/02635143.2013.811404>
- Theobald, E. J., Eddy, S. L., Grunspan, D. Z., Wiggins, B. L., & Crowe, A. J. (2017). Student perception of group dynamics predicts individual performance: Comfort and equity matter. *PLoS One*, 12(7), e0181336. <https://doi.org/10.1371/journal.pone.0181336>
- Topping, K., Buchs, C., Duran, D., & Van Keer, H. (2017). *Effective peer learning: From principles to practical implementation*. Taylor & Francis.
- Tressoldi, P. E. (2012). Replication unreliability in psychology: elusive phenomena or “elusive” statistical power? *Frontiers in Psychology*, 3(218), 1-5. <https://doi.org/10.3389/fpsyg.2012.00218>
- Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychological review*, 96(3), 506-520. <https://doi.org/10.1037/0033-295X.96.3.506>
- Utz, S. (2004). Self-construal and cooperation: Is the interdependent self more cooperative than the independent self? *Self and Identity*, 3(3), 177-190. <https://doi.org/10.1080/13576500444000001>
- van der Linden, S., & Goldberg, M. H. (2020). Alternative meta-analysis of behavioral interventions to promote action on climate change yields different conclusions. *Nature Communications*, 11(3915), 1-2. <https://doi.org/10.1038/s41467-020-17613-7>
- van 't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology: A discussion and suggested template. *Journal of Experimental Social Psychology*, 67, 2-12. <https://doi.org/10.1016/j.jesp.2016.03.004>
- Vives E., Poletti, C., Régner I. & Consortium ProFan. *Learning with Jigsaw: A review gathering all the pieces of the puzzle, 40 years later* [Manuscript in preparation]. Laboratoire de Psychologie Sociale, Université Aix-Marseille.
- Volet, S., Summers, M., & Thurman, J. (2009). High-level co-regulation in collaborative learning: How does it emerge and how is it sustained? *Learning and Instruction*, 19(2), 128-143. <https://doi.org/10.1016/j.learninstruc.2008.03.001>

- Voyles, E. C., Bailey, S. F., & Durik, A. M. (2015). New pieces of the jigsaw classroom: Increasing accountability to reduce social loafing in student group projects. *The New School Psychology Bulletin*, 13(1), 11-20. <http://www.nspb.net/index.php/nspb/article/view/264>
- Vygotsky, L. S. (1978). *Mind in society*. Harvard University Press.
- Walker, I., & Crogan, M. (1998). Academic performance, prejudice, and the Jigsaw classroom: New pieces to the puzzle. *Journal of Community & Applied Social Psychology*, 8(6), 381-393. [https://doi.org/10.1002/\(SICI\)1099-1298\(199811/12\)8:6<381::AID-CASP457>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1099-1298(199811/12)8:6<381::AID-CASP457>3.0.CO;2-6)
- Wiernik, B. M., & Dahlke, J. A. (2020). Obtaining unbiased results in meta-analysis: The importance of correcting for statistical artifacts. *Advances in Methods and Practices in Psychological Science*, 3(1), 94-123. <https://doi.org/10.1177/2515245919885611>
- Wittrock, M. C. (1992). Generative learning processes of the brain. *Educational Psychologist*, 27(4), 531-541. https://doi.org/10.1207/s15326985ep2704_8
- Xu, D., Solanki, S., McPartlan, P., & Sato, B. (2018). EASEing students into college: The impact of multidimensional support for underprepared students. *Educational Researcher*, 47(7), 435-450. <https://doi.org/10.3102/0013189X18778559>
- Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research*, 81(2), 267-301. <https://doi.org/10.3102/0034654311405999>
- Yoruk, A. (2016). Effect of Jigsaw method on students' chemistry laboratory achievement. *International Journal of Educational Sciences*, 15(3), 377-381. <https://doi.org/10.1080/09751122.2016.11890547>
- Yzerbyt, V. Y., Muller, D., & Judd, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of Experimental Social Psychology*, 40(3), 424-431. <https://doi.org/10.1016/j.jesp.2003.10.001>
- Ziegler, S. (1981). The effectiveness of cooperative learning teams for increasing cross-ethnic friendship: Additional evidence. *Human Organization*, 40(3), 264-268. <https://www.jstor.org/stable/44126039>
- Zimmerman, B. J., & Schunk, D. H. (Eds.). (2002). *Educational psychology: A century of contributions*. Routledge.

Table des annexes

Annexe A: Questionnaire à destination des élèves (E1).....	iv
Annexe B: Questionnaire à destination des élèves (E2).....	vi
Annexe C: Questionnaire à destination des élèves (E3).....	vii
Annexe D: Contenu des séquences pédagogiques (E3A, E3B, E3C)	viii
Annexe E: Corrélations entre le niveau scolaire initial (moyennes scolaires) et les performances aux tests (E1 à E3B).....	ix
Annexe F: Scores moyens et écart-type des tests (E3A, E3B, E3C).....	ix
Annexe G: Effets de Jigsaw sur les variables des questionnaires auto-rapportés	x
Annexe H: Résultats de l'ANCOVA (modèle complet) pour l'expérience 1	xi
Annexe I: Décomposition de l'interaction entre genre et condition (E1)	xii
Annexe J: Résultats de l'ANCOVA (modèle complet) pour l'expérience 2	xiii
Annexe K: Résultats de l'ANOVA à mesures répétées pour l'étude 3A.....	xiv
Annexe L: Résultats de l'ANOVA à mesures répétées pour l'étude 3B.....	xv
Annexe M: Décomposition de l'interaction entre genre et ordre (E3B).....	xvi
Annexe N: Résultats de l'ANOVA à mesures répétées pour l'étude 3C	xvii
Annexe O: Scores aux évaluations de connaissances en fonction de la condition (E3A)....	xviii
Annexe P: Scores aux évaluations de connaissances en fonction de la condition (E3B)	xviii
Annexe Q: Scores aux évaluations de connaissances en fonction de la condition (E3C).....	xix



QUESTIONNAIRE A DESTINATION DES ÉLÈVES



	<i>Pas du tout d'accord</i>				<i>Tout à fait d'accord</i>		
	1	2	3	4	5	6	7
Je me sens aussi compétent(e) que les autres élèves de mon âge en mathématiques.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'oublie souvent ce que j'apprends en mathématiques.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je suis plutôt lent(e) pour faire mon travail en mathématiques.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'ai l'impression de très bien réussir mon travail en mathématiques.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'ai beaucoup de mal à trouver les bonnes réponses en mathématiques.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Voici une nouvelle série de questions. Comme tout à l'heure, tu dois répondre en indiquant à quel point tu es d'accord avec les phrases suivantes.

	<i>Pas du tout comme moi</i>			<i>Exactement comme moi</i>	
	1	2	3	4	5
J'évite de donner une mauvaise image des autres.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Si deux ami(e)s se battent, je cherche un moyen d'arranger la situation.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quand je travaille en groupe, je fais ma part du boulot.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'écoute les idées des autres.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je contrôle ma colère lors d'une dispute avec un(e) ami(e).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'arrive à discuter d'un problème avec un(e) ami(e) sans empirer les choses.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je respecte les règles dans un parc, une salle de spectacle ou un événement sportif.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Je respecte le point de vue des autres, même si je ne suis pas d'accord avec eux.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
---	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

	1	2	3	4	5
Je me sens vraiment faire partie du collège X.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dans mon collège, j'ai l'impression que lorsque je fais quelque chose de bien, les gens le remarquent.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C'est difficile pour les gens comme moi d'être accepté ici.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Les autres élèves de ce collège prennent mes opinions au sérieux.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
La plupart des professeurs de ce collège s'intéressent à moi.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quelques fois, je me sens comme si je n'étais pas à ma place ici.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dans ce collège, il y a au moins un professeur ou un adulte avec qui je peux parler si j'ai un problème.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dans ce collège, les gens sont sympas avec moi.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dans ce collège, les professeurs ne sont pas intéressés par les gens comme moi.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je participe à plusieurs activités au collège Joliot-Curie.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je suis traité(e) avec autant de respect que les autres élèves.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je me sens différent(e) de la plupart des autres élèves dans ce collège.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dans ce collège je peux réellement être moi-même.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dans ce collège, les gens me respectent.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Les gens ici savent que je peux faire du bon travail.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je me sens fier(e) de faire partie du collège Joliot-Curie.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Les autres élèves de ce collège m'aiment comme je suis.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'aime travailler en groupe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je préfère travailler en groupe plutôt que seul.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je comprends mieux quelque chose quand je dois l'expliquer aux autres.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Dans un groupe, c'est toujours mieux que tout le monde s'aide.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quand je travaille en groupe, je me sens nerveux à l'idée de devoir communiquer avec les autres.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Pas du tout d'accord

Tout à fait d'accord

1 2 3 4 5

J'ai du mal à comprendre le principe de travailler en groupe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je suis satisfait quand je m'implique entièrement dans la réussite de mon groupe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je me sens rarement relaxé au sein d'un groupe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je pense que ce cours a été intéressant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'ai vraiment apprécié ce cours.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je pense que ce cours a été une perte de temps.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je trouve que ce cours a été amusant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'ai trouvé ce cours ennuyeux.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'ai été content d'avoir suivi ce cours.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je n'ai pas du tout aimé ce cours.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Avoir beaucoup d'imagination est important pour moi.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'aime bien être direct avec les gens que je rencontre.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Être unique et différent des autres de diverses manières me plaît.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je serais prêt à sacrifier mon propre intérêt pour le bénéfice des autres.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'ai souvent le sentiment que mes relations avec les autres sont plus importantes que mes réussites personnelles.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mon bonheur dépend de celui de mon entourage.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je resterais dans un groupe s'ils ont besoin de moi, même si je ne suis pas heureux (se) avec ce groupe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Imagine maintenant que l'échelle ci-dessous représente la position des gens dans notre société. Au sommet de l'échelle se trouvent les gens les plus riches et ceux qui ont les meilleurs emplois. Au bas de l'échelle se trouvent les gens ayant le moins d'argent et des emplois inintéressants voire pas d'emploi du tout. Sur cet axe il t'est demandé de mettre une croix à l'endroit où tu penses que tes parents se trouvent (Une seule croix).



J'ai travaillé individuellement sur la séquence :

- Arithmétique Géométrie Diagrammes Fractions Proportionnalité

Je souhaite avoir la correction des exercices :

Oui

Non

Tu es :

Un garçon

Une fille

Merci d'indiquer :

Age :

Classe :

NOM :

Prénom :

Annexe A: Questionnaire à destination des élèves (E1)

Pour chacune des affirmations ci-dessous, indique à quel point tu es d'accord en cochant la case correspondante.

	1	2	3	4	5
	Pas du tout				Tout à fait
	d'accord				d'accord
Je pense que ce cours a été intéressant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'ai vraiment apprécié ce cours.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je pense que ce cours a été une perte de temps.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je trouve que ce cours a été amusant.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'ai trouvé ce cours ennuyeux.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'ai été content d'avoir suivi ce cours.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je n'ai pas du tout aimé ce cours.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'aime travailler en groupe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je préfère travailler en groupe plutôt que seul.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je comprends mieux quelque chose quand je dois l'expliquer aux autres.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dans un groupe, c'est toujours mieux que tout le monde s'aide.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quand je travaille en groupe, je me sens nerveux à l'idée de devoir communiquer avec les autres.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'ai du mal à comprendre le principe de travailler en groupe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je suis satisfait quand je m'implique entièrement dans la réussite de mon groupe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je me sens rarement relaxé au sein d'un groupe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'évite de donner une mauvaise image des autres.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Si deux ami(e)s se battent, je cherche un moyen d'arranger la situation.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quand je travaille en groupe, je fais ma part du boulot.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'écoute les idées des autres.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je contrôle ma colère lors d'une dispute avec un(e) ami(e).	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'arrive à discuter d'un problème avec un(e) ami(e) sans empirer les choses.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Je respecte les règles dans un parc, une salle de spectacle ou un événement sportif.	Pas du tout d'accord	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Tout à fait d'accord	<input type="checkbox"/>
--	----------------------	--------------------------	--------------------------	--------------------------	----------------------	--------------------------

	1	2	3	4	5
Je respecte le point de vue des autres, même si je ne suis pas d'accord avec eux.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je me sens vraiment faire partie du collège de X	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dans mon collège, j'ai l'impression que lorsque je fais quelque chose de bien, les gens le remarquent.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Les autres élèves de ce collège prennent mes opinions au sérieux.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
La plupart des professeurs de ce collège s'intéressent à moi.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quelques fois, je me sens comme si je n'étais pas à ma place ici.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dans ce collège, les gens sont sympas avec moi.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je suis traité(e) avec autant de respect que les autres élèves.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je me sens différent(e) de la plupart des autres élèves dans ce collège.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dans ce collège je peux réellement être moi-même.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dans ce collège, les gens me respectent.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Les gens ici savent que je peux faire du bon travail.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je me sens fier(e) de faire partie du collège de X.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Les autres élèves de ce collège m'aiment comme je suis.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je suis plus intelligent que ceux de mon âge.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je fais facilement le travail qui m'est demandé.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je me souviens facilement des choses.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je trouve souvent les réponses en classe.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je me fais facilement des amis.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je suis apprécié par la plupart des autres élèves de mon âge.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
À l'école, je ne fais rien qui pourrait m'attirer des problèmes.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je ne fais pas de choses que je ne dois pas faire.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je m'apprécie en tant que personne.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je suis souvent mécontent de moi	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je pense que je fais bien les choses.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Annexe B: Questionnaire à destination des élèves (E2)

Pas du tout d'accord

Tout à fait d'accord

	1	2	3	4	5
Je pense que ces séances ont été intéressantes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'ai vraiment apprécié ces séances.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je pense que ces séances ont été une perte de temps.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je trouve que ces séances ont été amusantes.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'ai trouvé ces séances ennuyeuses.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
J'ai été content d'avoir suivi ces séances.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je n'ai pas du tout aimé ces séances.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

NOM : PRENOM :

CLASSE :

Tu es :

Un garçon :

Une fille :

Annexe C: Questionnaire à destination des élèves (E3)

Etude	Discipline	Thème	Séquence				
			1	2	3	4	5
3A	Physique-Chimie	Matière, mouvement, énergie et information	Matériaux et électricité	Préparer de l'eau salé	Mélange de masses et de matières	Qu'est ce qui change lors d'un mélange	/
3B	SVT	Biodiversité et milieu	Décrire son environnement	Répartition des espèces dans leurs milieux	Diversité des espèces	Classification des espèces	/
3C	SVT	Biodiversité et milieu	Décrire son environnement	Biodiversité	Classification des espèces	Les besoins de la plante et de la graine	

Annexe D: Contenu des séquences pédagogiques (E3A, E3B, E3C)

Étude	N	r (Pearson)	95% CI	
			LL	UL
1	252	.62***	.54	.69
2	313	.32***	.21	.41
3A	110	.64***	.52	.74
3B	74	.57***	.39	.70

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. CI = intervalle de confiance ; LL = borne basse ; UL = borne haute

Annexe E: Corrélations entre le niveau scolaire initial (moyennes scolaires) et les performances aux tests (E1 à E3B)

Étude	Test				
	1	2	3	4	5
3A	9.26 (4.44)	9.99 (3.66)	13.12 (3.51)	12.85 (2.55)	/
3B	16.46 (3.36)	14.63 (2.45)	15.54 (2.81)	15.82 (3.49)	/
3C	15.36 (3.05)	18.34 (2.36)	15.61 (3.48)	13.95 (4.64)	14.86 (3.55)

Note. Score allant de 0 (score minimal) à 20(score maximal). Ecart-type entre parenthèse.

Annexe F: Scores moyens et écart-type des tests (E3A, E3B, E3C)

Variable	Résultat de l'étude		
	1	2	3
Compétence perçue en mathématiques	H ₀ non rejetée, tendance à un effet inverse à H ₁ : $p = .106$, $\eta^2_p = .010$	/	/
Compétence sociale perçue	H ₀ non rejetée, effet inverse à H ₁ : $p = .049$, $\eta^2_p = .016$	H ₀ non rejetée, $p = .356$, $\eta^2_p = .003$	/
Sentiment d'appartenance scolaire	H ₀ non rejetée, effet inverse à H ₁ : $p = .014$, $\eta^2_p = .014$	H ₀ non rejetée, $p = .929$, $\eta^2_p = .000$	/
Attitudes envers le travail de groupe	H ₀ non rejetée, tendance à un effet inverse : $p = .132$, $\eta^2_p = .016$	H ₀ non rejetée, $p = .301$, $\eta^2_p = .004$	/
Soi indépendant	H ₀ non rejetée, $p = .440$, $\eta^2_p = .002$	/	/
Soi interdépendant	H ₀ non rejetée, $p = .140$, $\eta^2_p = .009$	/	/
Motivation Intrinsèque	H ₀ non rejetée, effet tendanciel : $p = .082$, $\eta^2_p = .013$	H ₀ non rejetée, $p = .864$, $\eta^2_p = .000$	H ₀ non rejetée, $p = .864$, $\eta^2_p = .000$
Sentiment d'efficacité personnelle (scolaire)	/	H ₀ non rejetée, $p = .214$, $\eta^2_p = .005$	/
Sentiment d'efficacité personnelle (social)	/	H ₀ non rejetée, $p = .814$, $\eta^2_p = .000$	/

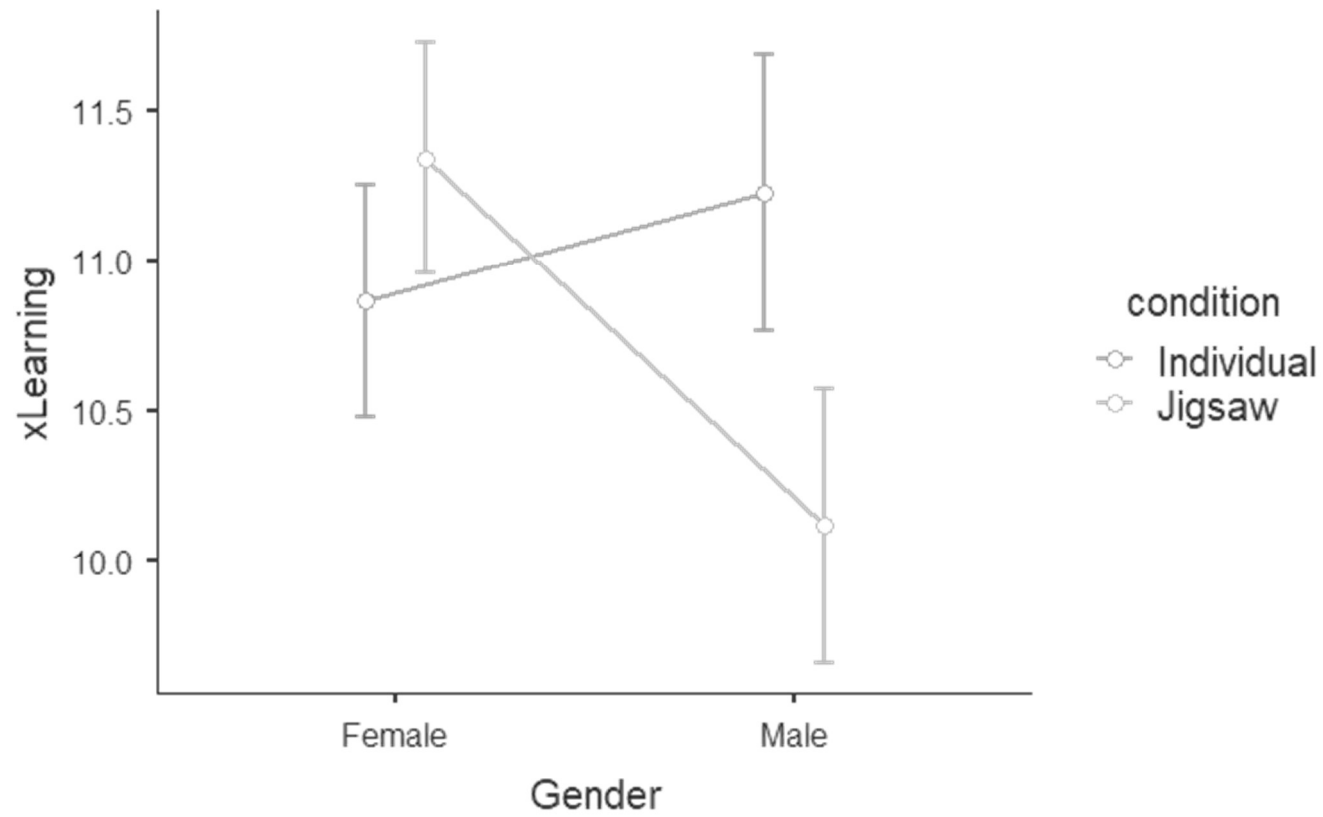
Note. H₁ = Jigsaw a un effet positif sur la variable auto-rapportée. H₀ : Jigsaw n'a pas d'effet positif sur la variable auto-rapportée.

Annexe G: Effets de Jigsaw sur les variables des questionnaires auto-rapportés

ANOVA Omnibus tests	SS	df	F	p	n²p
Model	1819.2504	8	21.16678	< .001	0.4107
SES(0;1)	40.5198	1	3.77155	0.053	0.0153
Gender(0;1)	5.9889	1	0.55744	0.456	0.0023
Condition (0;1)	10.3569	1	0.96401	0.327	0.0040
Niveau_initial(cov)	1176.4302	1	109.50108	< .001	0.3106
SES * Gender	0.6280	1	0.05845	0.809	0.0002
SES * Condition	30.7605	1	2.86316	0.092	0.0116
Gender * Condition	53.1161	1	4.94400	0.027	0.0199
SES01 * Gender * Condition	1.6269	1	0.15143	0.698	0.0006
Résidus	2610.6823	243			
Total	4429.9327	251			

Note : **SES** : statut socioéconomique, classification socioprofessionnelle de INSEE (0 = bas ; 1 = haut) ; **Gender** : genre (0 = fille, 1 = garçon) ; **Condition** (0 = Individuel ; 1 = Jigsaw)

Annexe H: Résultats de l'ANCOVA (modèle complet) pour l'expérience 1



Annexe I: Décomposition de l'interaction entre genre et condition (E1)

Model	SS	Mean Square	F	p	n²p
SES(0;1)	20.478767	20.478767	11.982784	< .001	0.0343
Gender(0;1)	0.449732	0.449732	0.263153	0.608	0.0008
Condition (0;1)	0.124004	0.124004	0.072559	0.788	0.0002
Niveau_initial(cov)	53.120888	53.120888	31.082738	< .001	0.0890
SES * Gender	2.055506	2.055506	1.202742	0.274	0.0034
SES * Condition	0.029326	0.029326	0.017160	0.896	0.0000
Gender * Condition	0.002695	0.002695	0.001577	0.968	0.0000
SES * Gender * Condition	1.018381	1.018381	0.595888	0.441	0.0017
résidus	519.540782	1.709016			

Note : **SES** : statut socioéconomique (0 = bas ; 1 = haut) ; **Gender** : genre (0 = fille, 1 = garçon) ; **Condition** (0 = Individuel ; 1 = Jigsaw)

Annexe J: Résultats de l'ANCOVA (modèle complet) pour l'expérience 2

<i>Within Subject Effects</i>	SS	df	Mean Square	F	p	η^2p
Continuous evaluation	823.06	3	274.354	31.5186	< .001	0.2777
Continuous evaluation * order	48.26	3	16.086	1.8480	0.139	0.0220
Continuous evaluation * niveau_initial(cov)	363.89	3	121.298	13.9351	< .001	0.1453
Continuous evaluation * gender	62.58	3	20.860	2.3965	0.069	0.0284
Continuous evaluation * ses	27.88	3	9.295	1.0678	0.363	0.0129
Continuous evaluation * order * gender	87.33	3	29.111	3.3443	0.020	0.0392
Continuous evaluation * order * ses	21.99	3	7.329	0.8420	0.472	0.0102
Continuous evaluation * gender * ses	36.72	3	12.239	1.4061	0.242	0.0169
Continuous evaluation * order * gender * ses	14.87	3	4.958	0.5695	0.636	0.0069
Residual	2141.31	246	8.705			

Note : **order** (0 = controle-jigsaw ; 1 = jigsaw-contrôle) ; **Gender** : genre (0 = fille, 1 = garçon) ; **ses** (0 = bas ; 1 = haut) ; **continuous evaluation** = time of measurement

Between Subjects Effects

	SS	df	Mean Square	F	p	η^2p
order	29.723139	1	29.723139	2.92176	0.091	0.0344
gender	6.923357	1	6.923357	0.68056	0.412	0.0082
ses	0.001702	1	0.001702	1.673e-4	0.990	0.0000
order * gender	1.232851	1	1.232851	0.12119	0.729	0.0015
order * ses1	5.207629	1	5.207629	0.51191	0.476	0.0062
gender * ses1	0.945909	1	0.945909	0.09298	0.761	0.0011
order * gender * ses1	75.148312	1	75.148312	7.38702	0.008	0.0826
niveau_initial(cov)	796.780184	1	796.780184	78.32291	< .001	0.4885
Residual	834.187252	82	10.173015			

Annexe K: Résultats de l'ANOVA à mesures répétées pour l'étude 3A

Within Subjects Effects

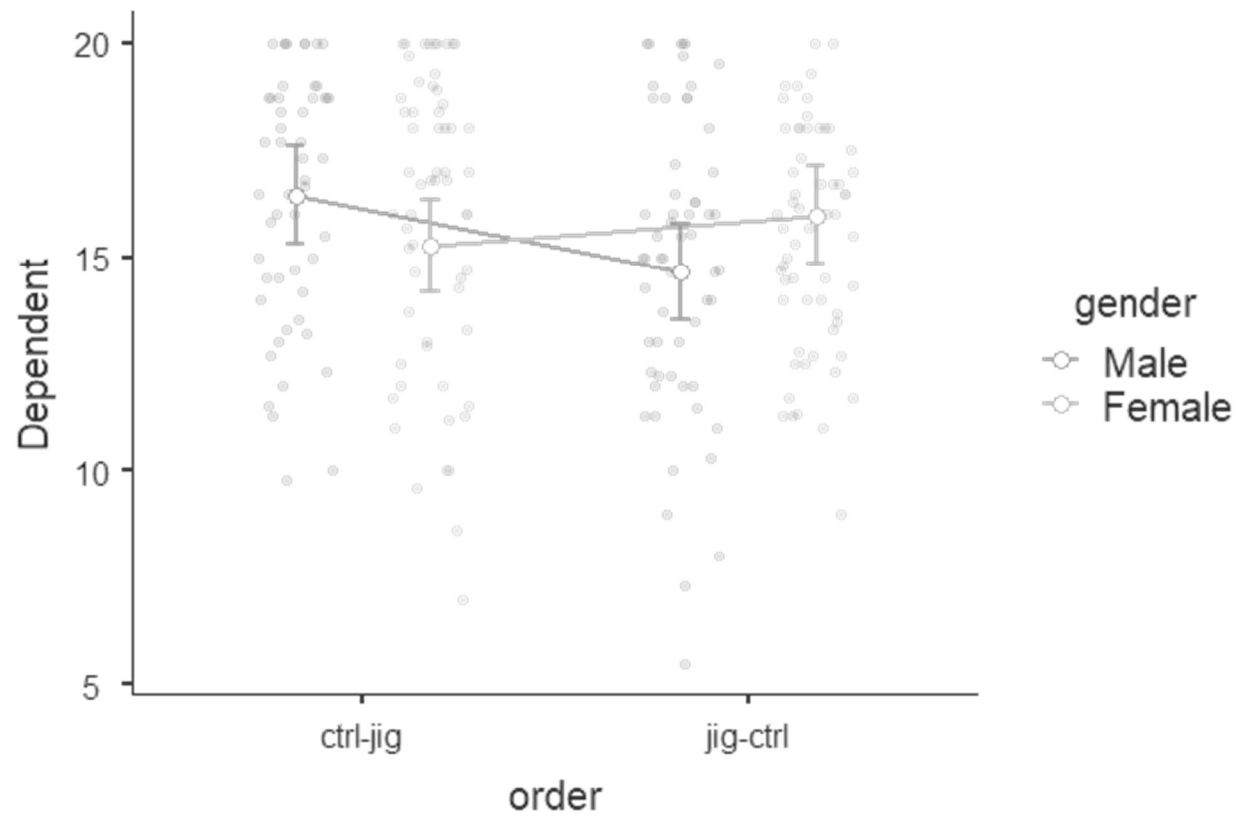
	Sum of Squares	df	Mean Square	F	p	η^2p
cc	84.392	3	28.1307	4.7786	0.003	0.0889
cc * ses	11.260	3	3.7535	0.6376	0.582	0.0128
cc * gender	11.550	3	3.8500	0.6540	0.559	0.0132
cc * order	12.216	3	4.0719	0.6917	0.935	0.0139
cc * niveau_initial(cov)	2.500	3	0.8333	0.1416	0.394	0.0029
cc * ses01 * gender	17.688	3	5.8961	1.0016	0.754	0.0200
cc * ses01 * order	7.045	3	2.3484	0.3989	0.511	0.0081
cc * gender * order	13.661	3	4.5535	0.7735	0.735	0.0155
cc * ses01 * gender * order	7.508	3	2.5028	0.4252	0.592	0.0086
Residual	865.366	147	5.8868			

Note : **order** (0 = controle-jigsaw ; 1 = jigsaw-contrôle) ; **Gender** : genre (0 = fille, 1 = garçon) ; **ses** (0 = bas ; 1 = haut) ; **cc** = **continuous evaluation** (time of measurement)

Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2p
ses	3.8437	1	3.8437	0.29293	0.591	0.0059
gender	0.2197	1	0.2197	0.01674	0.898	0.0003
order	11.6675	1	11.6675	0.88919	0.350	0.0178
ses * gender	1.8070	1	1.8070	0.13772	0.712	0.0028
ses * order	7.3466	1	7.3466	0.55989	0.458	0.0113
gender * order	62.2635	1	62.2635	4.74515	0.034	0.0883
ses * gender * order	25.7006	1	25.7006	1.95867	0.168	0.0384
niveau_initial(cov)	358.9669	1	358.9669	27.35717	<.001	0.3583
Residual	642.9533	49	13.1215			

Annexe L: Résultats de l'ANOVA à mesures répétées pour l'étude 3B



Annexe M: Décomposition de l'interaction entre genre et ordre (E3B)

Within Subjects Effects

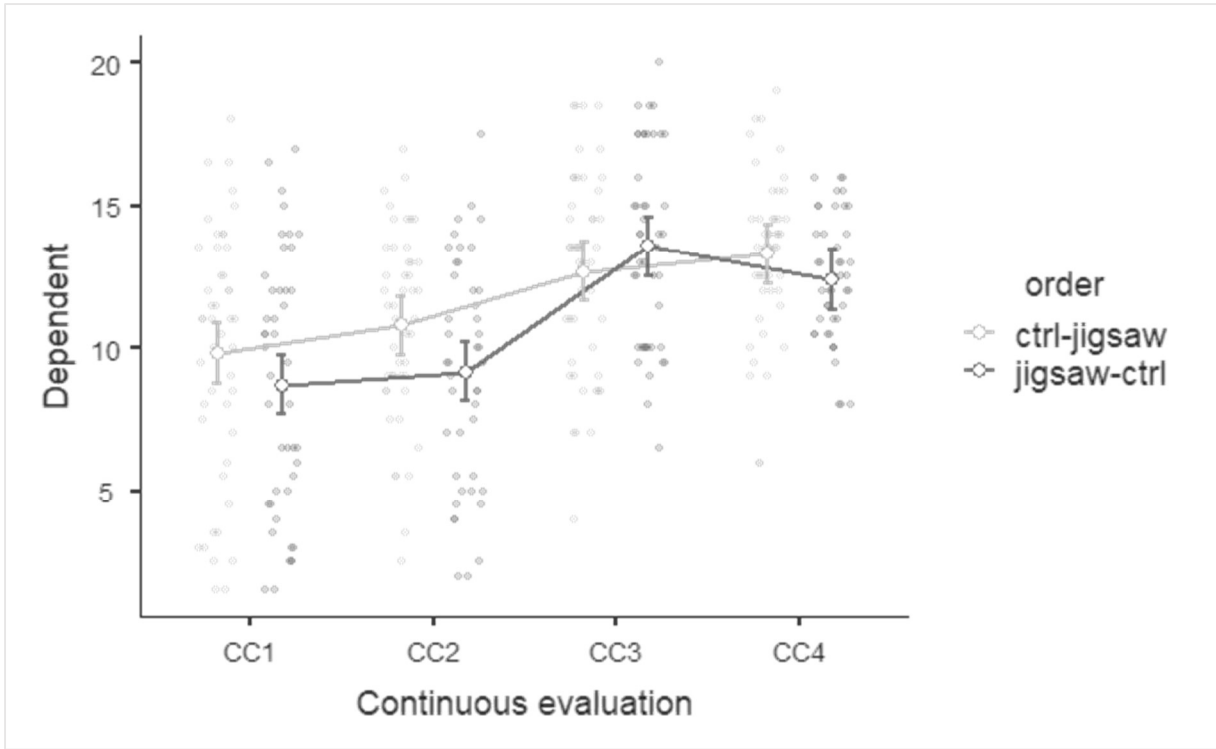
	Sum of Squares	df	Mean Square	F	p	η^2p
cc	888.288	4	222.072	32.1717	< .001	0.0705
cc * Gender	18.397	4	4.599	0.6663	0.616	0.0015
cc * order	7.862	4	1.966	0.2848	0.888	0.0006
cc * Gender * order	34.757	4	8.689	1.2588	0.286	0.0028
Residual	2429.757	352	6.903			

Note : **order** (0 = controle-jigsaw ; 1 = jigsaw-contrôle) ; **Gender** : genre (0 = fille, 1 = garçon) ; **ses** (0 = bas ; 1 = haut) ; **cc** = **continuous evaluation** (time of measurement)

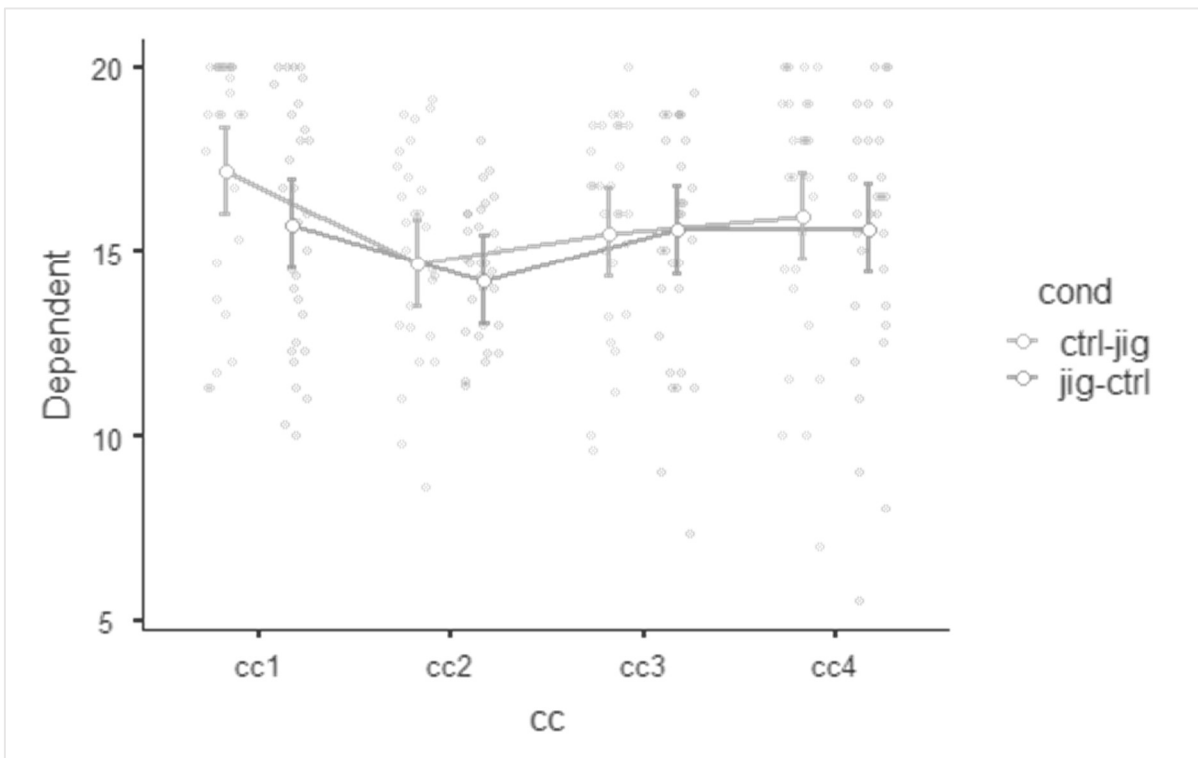
Between Subjects Effects

	Sum of Squares	df	Mean Square	F	p	η^2p
Gender	13.0842	1	13.0842	0.37492	0.542	0.0010
order	0.4283	1	0.4283	0.01227	0.912	0.0000
Gender * order	1.1807	1	1.1807	0.03383	0.854	0.0001
Residual	3071.0684	88	34.8985			

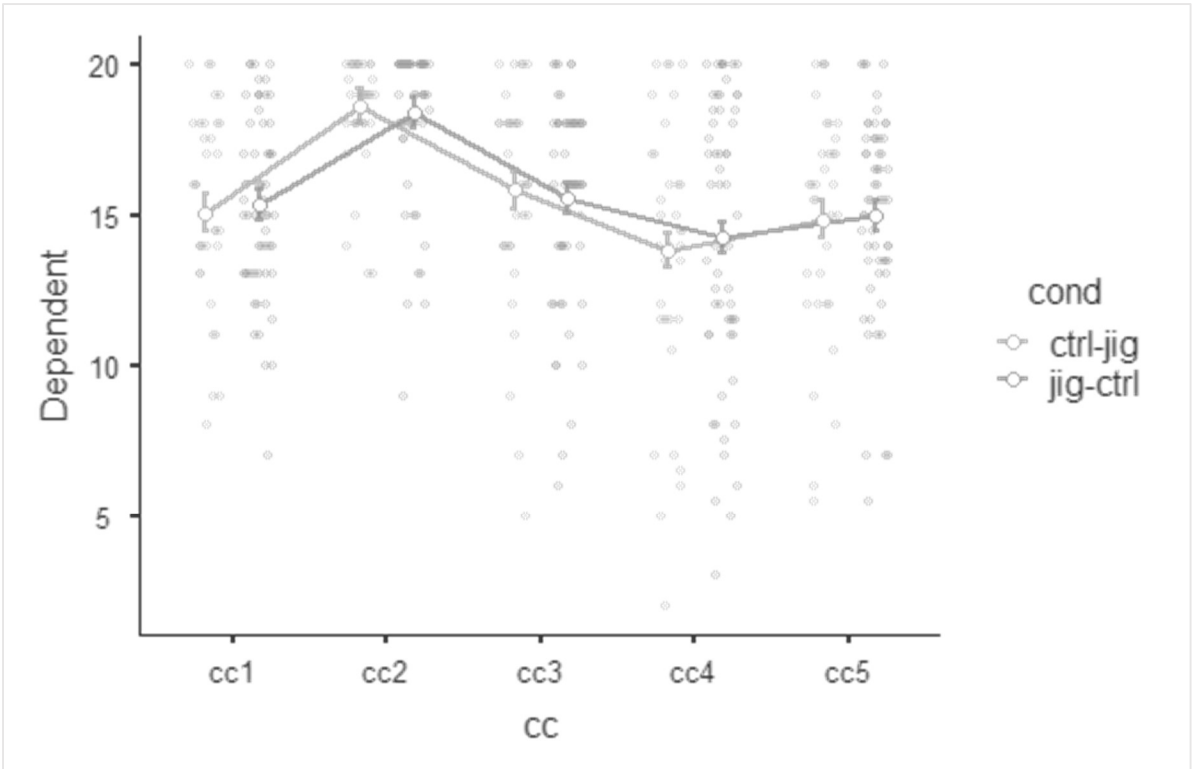
Annexe N: Résultats de l'ANOVA à mesures répétées pour l'étude 3C



Annexe O: Scores aux évaluations de connaissances en fonction de la condition (E3A)



Annexe P: Scores aux évaluations de connaissances en fonction de la condition (E3B)



Annexe Q: Scores aux évaluations de connaissances en fonction de la condition (E3C)