

Table des matières

Introduction.....	5
Partie 1. Les méta-analyses du WWC.....	7
Le site internet <i>Find What Works</i>	8
Les <i>Rapports d'intervention</i>	9
Conformité des études (étape 4)	11
Calcul des résultats pour chaque étude (étape 5.1)	15
Évaluation de l'efficacité de l'intervention (étape 5.2)	22
Examen d'une intervention : le bilan	30
Publier les conclusions (étape 5.3)	31
Un exemple : l'intervention <i>Odyssey Math</i>	33
Les <i>Guides des pratiques</i>	34
Partie 2. L'enseignement des mathématiques	41
Les <i>Rapports d'interventions</i> en mathématiques	41
Les <i>Guides des pratiques</i> en mathématiques	52
Partie 3. Discussion sur les méta-analyses	65
Défauts et qualités des <i>Rapports d'intervention</i> et des <i>Guides des pratiques</i>	65
Discussion sur les méta-analyses	69
Les méta-analyses dans la recherche sur l'enseignement	76
Glossaire.....	79
Références	82

Introduction

Tous les enseignants sont confrontés au quotidien à des élèves dont les apprentissages ne sont pas à la hauteur de leurs attentes. L'institution scolaire, les jeunes, leurs familles et les professionnels qui les entourent sont trop souvent déçus par des résultats jugés insatisfaisants. Toutes les disciplines sont concernées par ce constat, et les mathématiques n'y échappent pas. Elles ont la particularité d'être universelles et leur langage est compris de tous dans le monde : sur les 5 continents, une équation du second degré se résout de la même façon. Et même si les contextes nationaux donnent aux systèmes scolaires, aux méthodes d'enseignement et aux compétences et connaissances des élèves des couleurs spécifiques et uniques, on peut raisonnablement penser que cette discipline et son enseignement partagent un certain nombre de caractéristiques communes à tous ces environnements particuliers.

Trouver « ce qui marche » pour aider des élèves en difficulté est pour chaque enseignant une longue et difficile quête. Au moment où les élèves français semblent perdre pied en mathématiques comme l'atteste les résultats d'enquêtes internationales récentes^a, nous devons chercher des solutions pour les aider à surmonter ses difficultés. Nos voisins américains se posent les mêmes questions que nous, et il peut être intéressant de jeter un œil sur leurs réponses, même si celles-ci sont incomplètes et doivent être sans nul doute adaptées à notre contexte. C'est ce que je vous propose ici avec une présentation des

^a L'enquête TIMSS 2019 place les élèves français du CM1 et de 4^{ème} en dernière position en Europe.

méta-analyses menées depuis 2002 aux Etats-Unis par le What Works Clearinghouse (WWC). Dans une première partie il s'agira de comprendre comment procède le WWC pour examiner les études scientifiques quantitatives portant sur des questions relatives à l'enseignement ; l'objectif ici sera de donner du sens aux résultats publiés dans leurs synthèses. Dans un second temps, nous ferons le point sur leurs conclusions les plus fiables concernant plus particulièrement l'enseignement des mathématiques dans le secondaire, avant de conclure dans une troisième partie en pointant les qualités et les défauts de cette approche scientifique fondée sur les règles de l'*evidence based education* si souvent décriée en France.

Des explications sont données pour les mots écrits en **gras** dans la suite de ce texte, soit dans le glossaire soit dans un schéma annexe.

Les résultats portant sur l'enseignement des mathématiques mis en avant dans la partie 2 ont été sélectionnés par l'auteur et ne reflète en aucun cas un choix du What Works Clearinghouse. Ils peuvent être lus indépendamment des deux autres parties.

Les lecteurs qui souhaiteraient comprendre de plus près les calculs statistiques dont seules les principales conclusions sont présentées dans ce livre, peuvent accéder librement au document explicatif *Mesurer l'effet d'un traitement. Méta-analyses et enseignement* sur le site www.mathadoc.fr

Partie 1. Les méta-analyses du WWC

Depuis les années 80, de nombreuses études scientifiques quantitatives ont été menées dans les pays anglo-saxons, notamment aux Etats-Unis, dans le domaine de l'enseignement. Un grand nombre d'entre elles ont examiné une **intervention** mise en œuvre par des enseignants auprès de leurs élèves au sein d'un établissement scolaire. Les conclusions de ces études se présentent alors sous la forme de chiffres, parfois de natures différentes, qui sont autant de données dont il faut se saisir pour enfin conclure sur l'efficacité de telle ou telle intervention. Comment, à partir de dizaines d'études disparates, répondre aux attentes pragmatiques des enseignants, des parents et des administrations concernées par l'enseignement ? C'est l'objectif que se sont fixés les auteurs de **méta-analyses** qui synthétisent les conclusions d'études identifiées comme dignes d'intérêt dans un domaine particulier, pour ensuite publier leurs résultats sous une forme adaptée à un public dépassant le cercle étroit des chercheurs en sciences de l'éducation.

C'est de tout cela dont il sera question ici en nous intéressant plus particulièrement aux travaux du What Works Clearinghouse (WWC). Cela fait 18 ans maintenant que cette émanation de l'Institut des sciences de l'éducation du département de l'Éducation des États-Unis réalise des méta-analyses aussi bien sur les enseignements disciplinaires que sur l'attitude des élèves face à l'école en passant par les élèves à profil particuliers et la formation des enseignants. La mission confiée par les autorités au WWC, clairement énoncée par les intéressés eux-mêmes en introduction d'un de leur document de

référence^a, est « *d'être une source centrale et fiable de données scientifiques probantes sur ce qui marche dans le domaine de l'enseignement* ». Bien entendu, le cœur de leur cible (enseignants, chefs d'établissements, cadres administratifs) est essentiellement américaine, ce qui aura des conséquences sur l'utilisation concrète que nous pourrons faire de leurs conclusions comme nous le verrons plus loin.

Le site internet *Find What Works*

L'un des éléments marquant de la stratégie développée par le WWC concerne la publication et la diffusion de leurs travaux. Le site internet *Find What Works*^b est la porte d'entrée donnant accès à un énorme gisement d'informations. Depuis la page d'accueil du site, en cliquant sur le logo *Mathematics*, on obtient la liste des 151 interventions examinées par le WWC à ce jour^c, avec en tête de liste celles qui sont considérées comme les plus efficaces. Il est possible sur cette page de filtrer les résultats en fonction des niveaux d'étude^d par exemple.

En cliquant sur le nom d'une intervention, on ouvre une page qui fournit les éléments clés rendant compte des conclusions du WWC, et notamment, pour chaque **domaine** concerné, le classement de l'efficacité de l'intervention (voir plus loin). On peut alors télécharger le *Rapport d'intervention* complet mais également un résumé de 4 pages et un aperçu d'une page. Des données numériques peuvent également être importées pour être librement utilisées à l'aide d'un tableur. Vous trouverez aussi sur ce site internet des *Guides des*

^a p.1 de *What Works Clearinghouse : Standards Handbook* (WWC, 2020)

^b <https://ies.ed.gov/ncee/wwc/>

^c En mai 2020.

^d Ces niveaux d'études sont ceux définis par le système éducatif américain ; ils ont été traduits dans ce texte.

pratiques qui émettent des préconisations pragmatiques à l'attention des enseignants ; ces *Guides* s'appuient sur une analyse rigoureuse de la littérature mais aussi sur l'avis de leurs auteurs. Leur conception diffère en plusieurs points de celle des *Rapports d'intervention* dont nous allons dans un premier temps analyser la genèse en présentant les grandes lignes du processus mis en place par le WWC pour procéder à ses examens explicites et systématiques de la recherche scientifique.

Les Rapports d'intervention

Pour examiner puis évaluer une intervention, le What Works Clearinghouse réalise une méta-analyse : les **données** de certaines études scientifiques sont collectées, des **résultats** sont calculés par le WWC et des conclusions sont tirées puis publiées. Ces méta-analyses suivent toutes une même procédure constituée de cinq étapes chronologiques : un protocole d'examen est défini (étape 1), une recherche bibliographique est menée (étape 2), des études sont identifiées comme étant éligibles (étape 3) puis sélectionnées et classées par rapports à un ensemble de normes (étape 4). La cinquième et dernière étape a ici été divisée en trois sous-étapes : pour chacune des études sélectionnées des résultats sont calculés (étape 5.1), puis l'ensemble de ces résultats est synthétisé (étape 5.2) pour être enfin publié (étape 5.3) (**tableau 1**).

Toutes les procédures dont il sera question ici sont décrites dans le *Manuel des procédures du WWC* et le *Manuel des normes du WWC*. Le premier détaille les étapes 1, 2, 3 et 5 (et donne notamment les formules mathématiques utilisées pour le calcul des résultats), et le second propose des informations sur l'étape 4 en explicitant les normes WWC qui encadrent la sélection des études. Ces textes ont été mis à jour en janvier 2020. Tous les examens d'interventions dont il

sera question dans ce document se sont appuyées sur des textes antérieurs. Ces évolutions^a traduisent la vitalité de ce pan de la recherche scientifique qui souhaite constamment améliorer ses procédures ; elles compliquent malheureusement la présentation des résultats qu'il conviendra toujours de replacer dans le temps et le contexte du moment. Les points sujets à des changements profonds en janvier 2020 seront soulignés par la suite dans ce texte. Tous ces manuels sont en téléchargement libre sur le site internet du WWC^b.

Tableau 1 : les 5 étapes suivies par les méta-analyses du WWC

<p>Étape 1. Identification du protocole de l'examen</p> <p>L'examen d'une intervention suit un protocole établi en fonction des caractéristiques de l'intervention. Il concerne une catégorie d'interventions (par exemple le protocole <i>Mathématiques de la 3^{ème} à bac+ 1</i> est utilisé quand les interventions ont comme objectif d'améliorer les compétences mathématiques des élèves de ces niveaux scolaires). Ce document définit des paramètres comme les caractéristiques de la population cible et les domaines analysés, ainsi qu'une liste de mots clés et de bases de données à interroger. Si besoin, un protocole est créé.</p>
<p>Étape 2. Identification des études</p> <p>Les études scientifiques sont alors rassemblées grâce à une recherche exhaustive de la littérature publiée et non publiée (« littérature grise ») en utilisant les mots clés définis à l'étape précédente.</p>

^a Six versions des *Manuels* ont été publiées à ce jour : 1.0 en mai 2008 ; 2.0 en décembre 2008 ; 2.1 en septembre 2011 ; 3.0 en mars 2014 ; 4.0 en octobre 2017 et 4.1 en janvier 2020.

^bDes traductions partielles sont disponibles sur le site www.mathadoc.fr

Étape 3. Premier tri des études

Les études sont alors examinées une première fois et, si elles répondent bien à la question de recherche, sont déclarées éligibles.

Étape 4. Sélection des études

Chaque étude éligible est alors examinée une seconde fois et classée par rapport aux normes élaborées par le WWC (normes WWC) qui concernent plus particulièrement la relation de cause à effet au sein de l'échantillon de l'étude (on parle de **validité interne**). Les études non conformes sont écartées de la méta-analyse.

Étape 5. Calcul et publication des résultats

5.1 Pour chaque étude conforme aux normes WWC, des résultats sont calculés à partir des données.

5.2 Les résultats de toutes ces études sont combinés pour évaluer l'efficacité de l'intervention.

5.3 Les détails de l'examen et ses conclusions sont publiés sur le site internet du WWC et dans un *Rapport d'intervention*.

Nous allons nous intéresser ici plus particulièrement aux deux dernières étapes du processus suivi pour examiner une intervention : l'étape 4 et l'étape 5. On supposera donc qu'un protocole a été identifié (ou élaboré) qui a permis, après une recherche exhaustive de la littérature, de sélectionner des études dites éligibles. Une fois ces étapes franchies, la conformité de ces études doit donc être évaluée par rapport aux normes développées par le WWC pour *in fine* les classer : c'est l'étape 4.

Conformité des études (étape 4)

Cette étape vise à examiner la qualité de la méthode suivie par une étude éligible. L'objectif est ici de s'assurer que c'est bien

l'intervention qui est la cause de l'effet mesuré sur les élèves dans l'étude en question. L'intervention et uniquement l'intervention.

La plupart des études analysées par le WWC comparent deux groupes d'élèves au cours d'une expérience : un premier groupe d'élèves qui a été soumis à l'intervention (c'est le groupe d'intervention), un second groupe d'élèves qui n'a pas été soumis à l'intervention (c'est le groupe de comparaison). En fin d'expérience, tous les élèves passent un même test, et c'est sur la différence observée entre les moyennes des scores obtenues par chacun des deux groupes que les chercheurs vont se baser pour conclure. Mais en dehors de l'exposition (ou non) à l'intervention, d'autres facteurs pourraient expliquer (au moins en partie) la différence mesurée. On sait bien par exemple que le niveau scolaire d'un élève est un paramètre important. Et si l'un des groupes d'élèves montre un meilleur niveau en mathématiques que l'autre groupe avant l'expérience, cette différence de niveau initiale pourrait expliquer (au moins en partie) la différence observée après l'intervention. Pour éviter ce type de biais, la meilleure solution consiste à répartir les élèves entre ces deux groupes de façon aléatoire : l'étude est alors un **essai contrôlé randomisé** (voir **schéma 1**). Quand la répartition des élèves entre les deux groupes n'est pas aléatoire, on dit que l'étude est **quasi-expérimentale**.

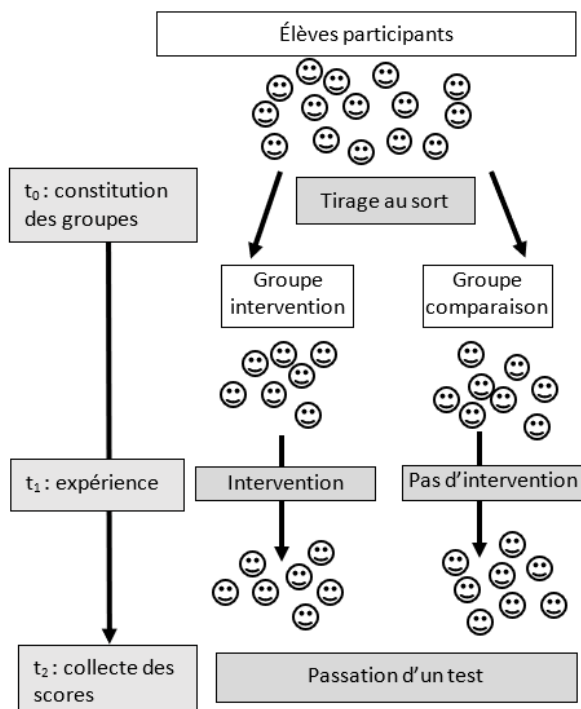
La qualité du design d'un essai contrôlé randomisé ou d'une étude quasi-expérimentale va dépendre notamment du taux d'**attrition** des échantillon et/ou de l'équivalence des groupes d'intervention et de comparaison avant l'intervention.

L'étude va être classée dans l'une des trois catégories suivantes :

- Étude conforme sans réserve aux normes WWC
- Étude conforme avec réserve aux normes WWC
- Étude non conforme aux normes WWC.

Pour être conforme sans réserve aux normes WWC, l'étude doit être un essai contrôlé randomisé avec une attrition faible. Les études conforme avec réserve aux normes WWC, sont soit des essais contrôlés randomisés avec attrition importante ou soit des études quasi-expérimentales ; elles doivent alors montrer que les groupes comparaison et intervention sont équivalents avant l'intervention.

Schéma 1 : essai contrôlé randomisé



Les autres études sont classées comme non conformes aux normes WWC^a, et leurs données ne sont pas exploitées par les chercheurs. Le niveau de conformité aux normes WWC (une étude peut être conforme avec ou sans réserve) jouera un rôle important dans le classement de l'effet de l'intervention comme on le verra plus loin.

Le **schéma 2** ci-dessous présente les principales étapes permettant de déterminer le classement de ces deux types d'études quand ce sont les élèves qui sont distribués entre les groupes intervention et comparaison. Un schéma un peu plus complexe s'applique quand ce sont des classes d'élèves qui sont affectées à la condition d'intervention ou de comparaison.

Les études conformes (avec ou sans réserve) aux normes WWC vont ensuite être analysées pour déterminer l'efficacité de l'intervention et les résultats seront publiés. Cette 5^{ème} et dernière étape va être effectuée en trois temps :

Étape 5.1. Pour chaque étude conforme aux normes WWC, des résultats sont calculés par le What Works Clearinghouse à partir des données publiées et l'efficacité de l'intervention est évaluée.

Étape 5.2. Pour l'ensemble des études sélectionnées, les résultats sont agrégés et l'efficacité globale de l'intervention est évaluée.

Étape 5.3. L'ensemble de ces résultats et les conclusions du WWC sont publiés.

Dans tous les cas, un résultat est toujours associé à un domaine unique (comme par exemple *mathématiques générales*).

^a Le WWC a également produit des normes pour des études qui suivent d'autres design que les comparaisons de groupes (non traitées ici).

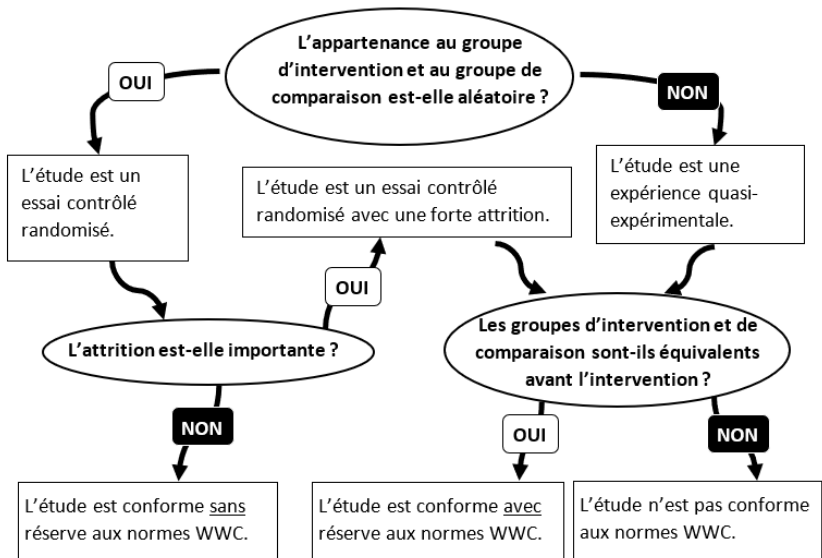


Schéma 2 : études conformes ou non aux normes WWC

Calcul des résultats pour chaque étude (étape 5.1)

Dans un premier temps, le WWC va calculer des résultats à partir des données publiées par l'étude examinée pour, dans un second temps et à partir de ces résultats, procéder au classement de l'efficacité de l'intervention telle qu'elle est mesurée par cette étude.

→ Résultats calculés par le WWC

Les moyennes des scores des élèves pour les deux groupes (intervention et comparaison) seront notées ici m_i et m_c . C'est l'amplitude de la différence de ces moyennes et la significativité (au

sens statistique) qu'il convient de lui attribuer que le WWC va évaluer en calculant trois résultats :

- la **taille d'effet** : différence standardisée des moyennes,
- l'**indice d'amélioration**,
- l'**estimation de la signification statistique** de la taille d'effet (valeur- p) ; ce résultat va permettre de déterminer si la taille d'effet est statistiquement et significativement positive. Ou bien encore d'écarter l'hypothèse selon laquelle la différence entre les deux moyennes serait due au hasard qui prévaut dans tout échantillonnage.

Taille d'effet

Comme on vient de le voir, la taille d'effet calculée par le WWC est la différence standardisée des moyennes, c'est-à-dire la différence des moyennes des scores de chacun des groupes, divisée par l'écart-type des scores de la population^a que l'on va noter s (voir égalité 1)^b. C'est donc aussi le nombre d'écart-types qui sépare le score d'un élève moyen du groupe comparaison du score d'un élève moyen du groupe intervention (voir égalité 2). Dans le domaine des sciences de l'éducation, il est courant d'obtenir des tailles d'effets, notée *ES* (pour *Effect Size*), inférieures à 1^c.

^a Cet écart-type sera estimé à partir des écarts-types des scores des groupes intervention et comparaison, s_i et s_c (le WWC calcule un écart-type groupé).

^b Les statisticiens du WWC utilisent le g de Hedges pour les variables continues, avec un terme correctif pour les petites tailles d'échantillon ; pour ne pas alourdir le texte, ce terme correctif a été négligé ici.

^c En mai 2020, 74% des tailles d'effets calculées par le WWC sont comprises entre -0,1 et 0,5. C'est le g d e Hedges qui est calculé par le WWC.

On utilise les données suivantes :

m_i : moyenne des scores du groupe intervention

m_c : moyenne des scores du groupe comparaison

s : écart-type des scores. Dans la même unité que les scores, ce nombre évalue la dispersion des scores autour de leur moyenne.

On a alors :

$$ES = \frac{m_i - m_c}{s} \quad (1)$$

$$m_i - m_c = ES \times s \quad (2)$$

La taille d'effet n'a pas de dimension et ne dépend donc pas de l'unité des scores. Pour s'assurer que seul l'intervention soit la cause d'une différence des moyennes, un grand nombre d'études quantitatives prend en compte le niveau initial des élèves. Et dans ce cas, chaque élève va en fait subir deux fois le même test : une fois avant l'expérience^a, et une fois après. Les moyennes des scores mesurés après l'expérience (on parle de scores posttests) de chacun des deux groupes sont alors ajustées aux moyennes des scores calculée avant l'intervention (les scores prétests). Ce sont ces moyennes ajustées (par les auteurs des études) qui seront utilisées par le WWC pour calculer la taille de l'effet.

Indice d'amélioration du WWC

L'indice d'amélioration (noté *IA*) est calculé à partir de la taille d'effet. Il peut être interprété comme le changement attendu du rang

^a Si la différence entre les deux moyennes avant l'intervention est trop importante, le WWC ne prendra pas en compte les données de l'étude.

centile^a d'un élève du groupe de comparaison se situant dans la moyenne, s'il avait bénéficié de l'intervention. Ainsi, si la taille d'effet est par exemple égale à $+0,5$, un élève du groupe de comparaison qui se situe au 50^{ème} centile de l'échantillon de l'étude (un élève « moyen » donc) aurait obtenu un score supérieur de $0,5$ écart-type s'il avait reçu l'intervention, et se serait situé au 69^{ème} centile si on suppose que la distribution des scores est normale^b. L'indice d'amélioration obtenu est donc de $+19$ ($69 - 50$). Le **schéma 3** permet de visualiser ce résultat, la courbe noire représentant la distribution des scores du groupe de comparaison, et la courbe en pointillés représentant la distribution des scores du groupe d'intervention (les courbes ont été centrées et réduites).

Signification statistique

Les études quantitatives comparent des groupes qui sont des échantillons d'une population. La différence observée entre le groupe comparaison et le groupe intervention peut être le fruit de l'expérience menée (et donc de l'exposition ou non des élèves à l'intervention) mais aussi de l'échantillonnage^c. On doit donc se demander si la taille d'effet calculée est statistiquement significative, c'est-à-dire si on peut, en acceptant de prendre un risque, rejeter l'hypothèse selon

^a Mesure de position d'une donnée dans une série statistique ordonnée et divisée en 100 intervalles de même effectif.

^b Cette transformation de *ES* en *IA* s'appuie sur les propriétés des courbes normales.

^c Deux échantillons extraits d'une même population sont presque toujours différents ; on sait par exemple que, dans certaines conditions (quand les tailles d'échantillons sont supérieures à 30 notamment), les moyennes des scores des échantillons seront distribuées normalement autour de la moyenne des scores de la population entière.

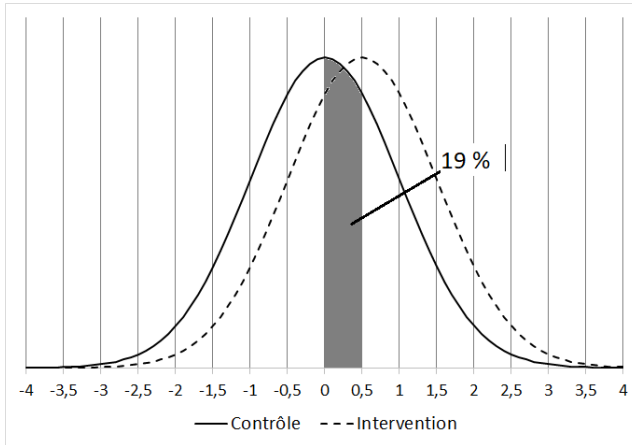


Schéma 3 : l'indice d'amélioration du WWC

laquelle c' est l'échantillonnage qui expliquerait que la différence des moyennes (et donc la taille d'effet) ne soit pas nulle. Le risque sera évalué par le calcul d'une probabilité, la valeur $-p$, ou encore p . Ainsi, pour le WWC, une taille d'effet sera considérée comme statistiquement et significativement différente de 0 (et, dans les faits, supérieure à 0, tant il est rare de trouver des tailles d'effet négative et encore plus statistiquement et significativement négative) quand la probabilité (dénommée valeur $-p$ ou encore p) d'observer une taille d'effet aussi grande dans le cas où l'intervention n'aurait eu aucun impact, est inférieure à 1 sur 20. Cela revient, pour les statisticiens^a, à calculer la valeur $-p$ et à montrer que $p < \frac{1}{20} = 0,05$.

^a On procède ici à des tests d'hypothèse ; le WWC utilise un test de Student bilatéral avec $\alpha = 0,05$

Ce calcul est sensible aux conditions expérimentales, et doit tenir compte de celles-ci en apportant des corrections quand cela est nécessaire. C'est le cas par exemple quand on affecte aléatoirement des classes d'élèves (et non des élèves) dans les groupes intervention ou comparaison et que les scores analysés sont ceux des élèves ; ou dans le cas de mesures multiples, c'est-à-dire quand plusieurs données sont disponibles.

La complexité du terrain

Il est pratiquement impossible d'affecter des élèves issus d'une population cible de façon aléatoire dans un groupe intervention (ou contrôle) ; ce sont le plus souvent des classes (voire des établissements) qui sont tirés au sort pour participer à l'expérience. Les calculs statistiques devront en tenir compte car ne pas le faire surestime l'influence de l'intervention. De la même manière, la plupart du temps, les études présentent plusieurs données. Elles peuvent par exemple fournir des données pour des sous-ensembles de l'échantillon ; ou bien utiliser plusieurs tests ; ou bien mesurer une même caractéristique à plusieurs moments pendant l'expérience ou la période de suivi. Quand la qualité des données le permet, le WWC calculera alors la moyenne des tailles d'effet de chacune des données ainsi que l'indice d'amélioration moyen correspondant. L'estimation de la signification statistique de la taille moyenne de l'effet devra ici tenir compte de la multiplicité des données qui est susceptible de la surévaluer.

La leçon à tirer de ces remarques qui ne font qu'effleurer un grand nombre de questions techniques, est que les examens des études scientifiques quantitatives doivent être menés par des experts en la matière. Ils sont de fait des membres à part entière des équipes de chercheurs du What Works Clearinghouse. Ces calculs sont expliqués

pour des cas simples, dans *Mesurer l'effet d'un traitement. Méta-analyses et enseignement* (Nathalie ROQUES, 2021).

Les chiffres à retenir

- Une taille d'effet (ES) supérieure ou égale à $+0,25$ est considérée comme importante par le WWC.
- Un indice d'amélioration (IA) supérieur ou égal $+10$ (correspondant à une taille d'effet de $+0,25$) est donc important.
- Si $p < 0,05$ la taille d'effet est statistiquement et significativement différente de zéro.

Pour chaque étude on dispose donc d'une taille d'effet, d'un indice d'amélioration et d'une valeur- p (**schéma 4**). Ces résultats vont permettre dans un second temps d'évaluer l'efficacité de l'intervention telle qu'elle est évaluée par cette étude.

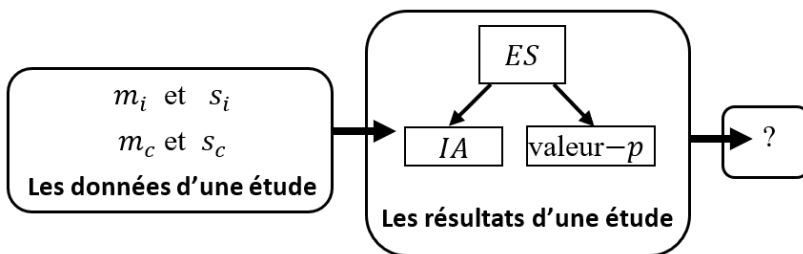


Schéma 4 : les résultats de l'examen d'une étude

→ Classement de l'efficacité de l'intervention

Jusqu'en janvier 2020, l'effet de l'intervention était classé dans l'une des cinq catégories suivantes :

1. Effet positif statistiquement significatif
2. *Effet positif important*
3. Effet indéterminé
4. *Effet négatif important*
5. Effet négatif statistiquement significatif

Les mentions d'effets importants (lignes 2 et 4) disparaîtront lors des prochains examens : depuis janvier 2020 il n'y a plus que trois échelons. L'effet de l'intervention est considéré comme positif et statistiquement significatif (ligne 1) quand la taille de l'effet est positive et statistiquement significative. On doit donc avoir $ES > 0$ et $p < 0,05$. L'effet positif important (ligne 2, supprimé depuis janvier 2020) reconnaissait jusqu'à cette date un certain crédit aux études montrant une taille d'effet supérieure ou égale à +0,25 mais sans significativité statistique (souvent des études menées sur de petits échantillons). L'effet est considéré comme indéterminé si la taille d'effet n'est ni statistiquement significativement positive ni importante. L'effet négatif statistiquement significatif (ligne 5) n'est pratiquement jamais rencontré^a.

Évaluation de l'efficacité de l'intervention (étape 5.2)

C'est la synthèse des résultats de toutes les **études primaires** examinant l'effet d'une intervention et conformes (avec ou sans

^a En mai 2020, 4% des tailles d'effets calculées par le WWC étaient dans ce cas.

réserve) aux normes WWC qui permet d'attribuer un classement à l'effet global de cette intervention dans chaque domaine considéré.

À cela s'ajoute l'évaluation du niveau de preuve associé à ce classement : ce niveau mesure en quelque sorte le degré de confiance avec lequel on peut considérer le classement proposé.

Le **schéma 5** représente une méta-analyse (on parle aussi d'**étude secondaire**) qui aurait accompli toutes les étapes précédentes et qui doit maintenant exploiter les résultats calculés pour chacune des études.

→ Calcul des résultats

On va utiliser les tailles d'effet de chacune des études primaires calculées à l'étape précédente. Ces tailles d'effet, rappelons-le, sont des indicateurs standardisés et indépendants des échelles de scores parfois différentes utilisées dans les études. Il va donc être possible de les agréger et de calculer la taille d'effet globale de l'intervention. Et de la même façon que pour l'examen individuel d'une étude, on cherchera une réponse à la fameuse question : cet indicateur est-il statistiquement et significativement positif ?

Dans les procédures antérieures à 2020, l'indicateur calculé par le WWC qui permet de calculer la taille d'effet globale de l'intervention est la moyenne arithmétique des tailles d'effet de chacune des études : la somme des tailles d'effets est divisée par le nombre des études.

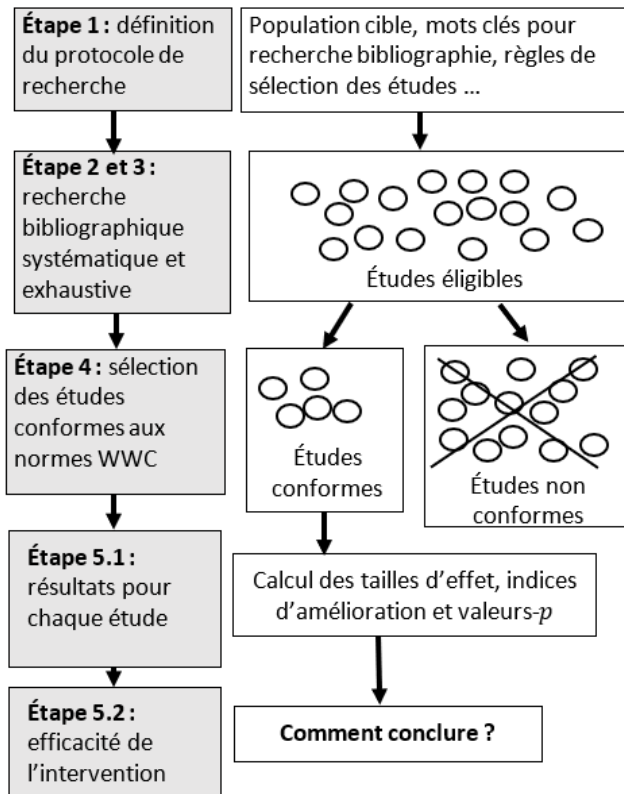


Schéma 5 : méta-analyse menée par le WWC

Depuis janvier 2020 c'est la moyenne des tailles d'effet pondérée par l'inverse des variances des tailles d'effet de chacune des études qui est calculée. La taille d'effet ainsi obtenue est en fait la meilleure

estimation^a de l'effet global réel de l'intervention, mais son calcul est plus compliqué à comprendre^b.

De la même façon que pour une étude primaire, un indice d'amélioration est alors déduit de la taille d'effet globale, et c'est ce résultat qui est le plus largement communiqué auprès du public, car il permet de visualiser l'efficacité globale d'une intervention.

La signification statistique de la moyenne pondérée des tailles d'effet va être estimée, et c'est cette estimation (calcul de la valeur- p) qui va servir de base à l'évaluation de l'efficacité des interventions pour les examens menés après janvier 2020. La signification statistique de la moyenne arithmétique des tailles d'effet n'était pas estimée.

→ **Classement de l'efficacité de l'intervention**

Des changements importants ont été apportés en janvier 2020 sur la façon d'évaluer l'efficacité globale de l'intervention, en lien direct avec le nouveau mode de calcul de la taille d'effet globale de l'intervention que nous venons de décrire. On tient également compte à ce niveau de la qualité du design des études (c'est-à-dire de leur niveau de conformité aux normes WWC). Ces deux éléments vont être combinés afin de déterminer le classement de l'efficacité de l'intervention.

^a La variance de cet estimateur est minimale.

^b De la même façon que pour une moyenne d'échantillonnage que l'on peut considérer comme une variable aléatoire et à laquelle on associe donc une variance (le carré de l'écart-type), une taille d'effet possède une variance qui intègre dans son calcul la taille des échantillons. Plus les tailles des groupes comparaison et intervention sont grandes et semblables, plus la variance est petite et plus le poids de l'étude est important (Michael BORENSTEIN, 2009).

Jusqu'en 2020, il y avait 6 échelons possibles (**tableau 2**). En absence de l'estimation de la signification statistique de la moyenne arithmétique des tailles d'effet, le WWC procédait par comptage des études montrant des tailles d'effet statistiquement et significativement positives.

Ainsi, on peut retenir qu'avant janvier 2020, pour que l'effet d'une intervention soit qualifié de **positif**, il fallait que pour deux études au moins les tailles d'effet soient positives et statistiquement significatives et qu'au moins une de ces études soit conforme sans réserve aux normes WWC (une de ces études devait donc être un essai contrôlé randomisé avec une faible attrition), et que les tailles d'effets calculées pour les autres études ne soient pas négatives et statistiquement significatives ou importantes.

L'effet était considéré comme **potentiellement positif** quand au moins pour une étude la taille d'effet calculée était statistiquement et significativement positive ou importante, et que le nombre d'études montrant des effets indéterminés était inférieur ou égal au nombre d'études montrant un effet positif et que pour aucune étude les tailles d'effets calculées n'étaient négatives et statistiquement significatives ou importantes.

Tableau 2. Classement de l'efficacité d'une intervention en vigueur avant janvier 2020

Effet positif	Preuves fortes d'un effet positif sans preuve contraire et conformité <u>sans</u> réserve aux normes WWC pour au moins une étude.
Effet potentiellement positif	Preuve d'un effet positif sans preuve contraire majeure.
Aucun effet discernable	Aucune preuve affirmative d'effet.

Effet mixte	Preuve d'effets incohérents.
Effet potentiellement négatif	Preuve d'un effet négatif sans preuve contraire majeure.
Effet négatif	Preuves fortes d'un effet négatif sans preuve contraire et conformité <u>sans</u> réserve aux normes WWC pour au moins une étude.

A partir de 2020, il n'existe plus que 5 échelons (**tableau 3**). On dispose maintenant de la moyenne des tailles d'effet pondérée par l'inverse des variances des tailles d'effet. Et cette fois, il sera possible d'estimer sa significativité statistique en lui associant une valeur- p qui va intervenir dans le classement de l'efficacité de l'intervention.

On peut retenir que depuis janvier 2020, pour qu'un effet soit qualifié de **positif**, il faut que la taille d'effet globale calculée à partir d'au moins deux études soit statistiquement et significativement positive et que plus de 50% de cette taille moyenne pondérée provienne d'études conformes sans réserve aux normes WWC : au moins l'une de ces études est un essai contrôlé randomisé avec une faible attrition et la part de ce type d'études est prépondérante.

Tableau 3. Classement de l'efficacité d'une intervention en vigueur depuis janvier 2020

Effet positif	Taille d'effet globale statistiquement et significativement positive et 50% de cette taille d'effet provient d'études conformes <u>sans</u> réserve aux normes WWC
Effet potentiellement positif	Taille d'effet globale statistiquement positive
Effet incertain	Pas d'effet statistiquement significatif

Effet potentiellement négatif	Taille d'effet globale statistiquement négative
Effets négatif	Taille d'effet globale statistiquement et significativement négative et une majorité d'études sont conformes <u>sans</u> réserve aux normes WWC

Encore une fois, pour les interventions analysées avant janvier 2020 et dont il sera question dans la suite de ce texte, les procédures s'appuient sur des versions antérieures à celles qui ont été mises en place depuis.

→ Niveau de preuve

La dernière étape consiste à évaluer le niveau de preuve associé au classement de l'intervention, toujours dans un domaine particulier, en s'appuyant sur les critères présentés dans le **tableau 4** (ce tableau n'a pas subi de modifications importantes en janvier 2020). Il existe deux niveaux de preuves : un niveau moyen à important et un niveau faible. Trois critères vont intervenir ici : le nombre d'études conformes aux normes WWC, le nombre de terrains analysés par ces études et la taille des échantillons.

Tableau 4. Critères utilisés pour déterminer le niveau de preuve^a

Niveau de preuve moyen à important	L'examen comprend plus d'une étude conforme aux normes WWC ET concerne plusieurs terrains ET les résultats sont basés sur un échantillon total d'au moins 350 élèves.
------------------------------------	---

^a Traduction de la figure IV.4 p.26 de *What Works Clearinghouse : Procedures Handbook (Version 4.1)*

Niveau de preuve faible	Le domaine comprend une seule étude conforme aux normes WWC OU concerne un seul terrain OU les résultats sont basés sur un échantillon total de moins de 350 élèves.
-------------------------	--

Si une seule étude est analysée, on ne peut écarter l'hypothèse selon laquelle certaines caractéristiques de cette étude, telles que la façon de mesurer les données ou le moment choisi pour mener l'expérience, aient pu affecter les résultats. L'examen de plusieurs études réduit également le biais potentiel dû à une erreur d'échantillonnage.

De même, avec l'étude d'un seul terrain (par exemple un établissement), il est possible que certaines caractéristiques de ce terrain (par exemple les données démographiques des élèves de l'établissement) aient pu affecter les résultats.

Enfin, la taille d'échantillon totale (la somme des tailles des échantillons de chacune des études) doit être supérieure à 350 de manière à ce que la probabilité de détecter une taille d'effet égale à 0,30 et statistiquement significative soit au moins égale à 80%^a.

Pour toutes ces raisons, le WWC considère que l'ampleur des preuves est faible lorsque les conclusions ne reposent que sur une seule étude ou que sur l'étude d'un seul terrain ou que la taille d'échantillon totale est inférieure à 350. Et inversement, le niveau de preuve sera moyen à important si plusieurs études ont été menées sur des terrains différents avec des échantillons de taille suffisante.

^a Ces chiffres concernent un essai contrôlé randomisé simple avec des groupes intervention et comparaison de taille égale en l'absence de covariables dans l'analyse des données avec $p < 0,05$.

Examen d'une intervention : le bilan

L'examen d'une intervention a donc deux objectifs : évaluer l'efficacité de l'intervention (en agrégeant les résultats calculés pour chacune des études et en tenant compte de leur niveau de conformité vis-à-vis des normes WWC) et évaluer le niveau de preuve. A titre d'illustration, le **schéma 6** donne une vue d'ensemble de l'examen d'une intervention qui se serait basé sur trois études conformes aux normes WWC (c'est un nombre raisonnable d'études, car peu franchissent cette sélection rigoureuse). Deux questions différentes sont posées mais qui sont tout de même liées : si trois études sont conformes sans réserve aux normes WWC avec des tailles d'effets statistiquement et significativement positives, l'efficacité de l'intervention sera démontrée^a. Cela va également dans le sens d'un niveau de preuve important, car taille d'effet, significativité statistique et taille d'échantillon sont associées, et il est donc vraisemblable que la taille d'échantillon totale soit supérieure à 350.

Le nombre de terrains restera dans cet exemple un élément à considérer. C'est ce qu'illustre le **tableau 5** ci-contre qui prend à nouveau comme exemple le cas d'une méta-analyse ayant sélectionné 3 études conformes aux normes WWC et qui toutes montrent des tailles d'effet statistiquement et significativement positive^b (c'est bien sûr un cas extrêmement favorable, il s'agit ici avant tout d'illustrer les concepts vus auparavant).

^a La taille d'effet globale calculée par le WWC depuis janvier 2020 sera alors forcément statistiquement significativement positive.

^b On peut montrer que dans ce cas la taille d'effet globale est statistiquement significativement différente de zéro.

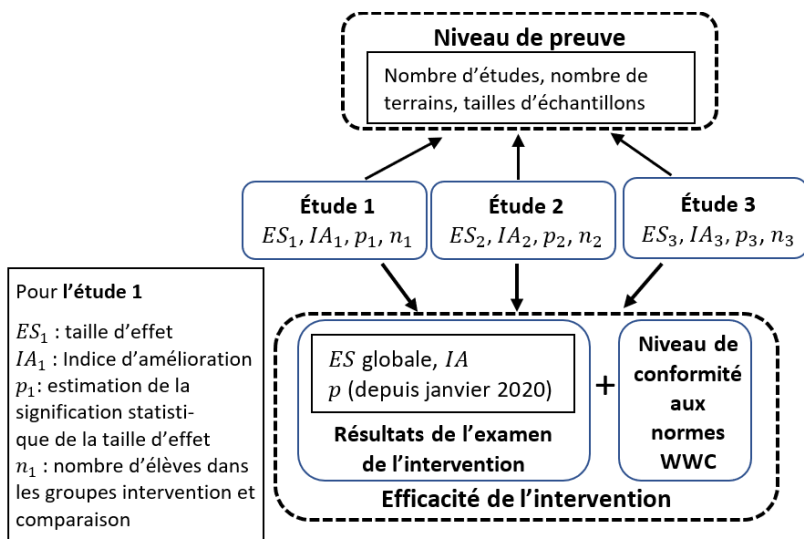


Schéma 6 : examen d'une intervention basé sur 3 études

Publier les conclusions (étape 5.3)

Les informations concernant l'examen d'une intervention sont publiées sur les pages du site internet *Find What Works* et de façon plus détaillée dans le *Rapport d'intervention* en accès libre sur le site. Ces *Rapports* présentent en introduction les éléments clés sur l'efficacité de l'intervention pour chacun des domaines considérés (certaines interventions concernent plusieurs domaines, mais les résultats sont toujours attachés à un domaine spécifique).

Tableau 5. Conclusions de l'examen d'une intervention basé sur 3 études (pour les notations voir le schéma 6).

	Étude 1	Étude 2	Étude 3
Niveau de conformité	<u>sans</u> réserve	<u>avec</u> réserve	<u>avec</u> réserve
Résultats des études	$ES_1 > 0$ $p_1 \leq 0,05$	$ES_2 > 0$ $p_2 \leq 0,05$	$ES_3 > 0$ $p_3 \leq 0,05$

Efficacité de l'intervention	Avant janvier 2020	Après janvier 2020
Résultats	$ES = \frac{ES_1 + ES_2 + ES_3}{3}, IA$	$ES =$ moyenne pondérée $IA ; p$
Évaluation de l'effet : effet positif	<ul style="list-style-type: none"> • au moins deux études (ici 3) montrent des effets positifs statistiquement significatifs • une étude est conforme <u>sans</u> réserve aux normes WWC • aucun effet négatif 	$ES > 0$ et $p \leq 0,05$ <ul style="list-style-type: none"> • Si plus de 50% de la taille d'effet provient d'études conformes <u>sans</u> réserve aux normes WWC^a

Niveau de preuve	
Moyen à fort	<ul style="list-style-type: none"> • si plusieurs terrains sont étudiés • si $n_1 + n_2 + n_3 > 350$

On y trouve aussi les résultats calculés pour chacune des études examinées ainsi que la liste de toutes les publications analysées : les références concernent bien sûr les études sélectionnées dont les

^a En simplifiant un peu et pour donner un ordre de grandeur, les tailles d'échantillons des études conformes sans réserve aux normes WWC doivent être supérieures aux tailles d'échantillons des études conformes avec réserve.

données ont été exploitées, mais aussi les études qui n'ont pas été sélectionnées à l'étape 3 (les études inéligibles), ni à l'étape 4 (les études éligibles mais non conformes aux normes WWC) en donnant une courte description des motifs de leur exclusion. Tous les résultats (et certaines données) des études exploitées dans le cadre d'une intervention sont également téléchargeables au format csv afin d'être librement exploités à l'aide d'un tableur. Il est alors notamment possible de reproduire les calculs menés par les auteurs de la méta-analyse. Les *Rapports d'intervention* reprennent également certains des éléments publiés dans les deux *Manuels* (*Manuel des procédures du WWC* et le *Manuel des normes du WWC*) dans le but d'explicitier les procédures et les critères appliqués pour, *in fine*, justifier les conclusions du WWC. Ces éléments sont issus bien évidemment des *Manuels* en vigueur au moment de l'examen de l'intervention par le WWC.

D'une façon générale, les *Rapports d'intervention* du What Works Clearinghouse se limitent à établir un constat. Aucune préconisation n'est énoncée, aucun commentaire n'est formulé. Il en va autrement dans les *Guides des pratiques* publiés par le WWC dont il va être question par la suite.

Un exemple : l'intervention *Odyssey Math*

C'est maintenant une vue de l'ensemble de toutes les étapes de l'examen de l'intervention *Odyssey Math* qui est proposée ci-contre (**tableau 6**). Vous trouverez plus loin des informations sur cette intervention spécifique à l'enseignement des mathématiques ; le seul domaine analysé ici est *mathématiques générales*. Cette intervention a été examinée par le WWC en 2017 en utilisant donc les critères en vigueur avant janvier 2020. Une estimation de l'efficacité globale de

l'intervention utilisant les critères appliqués depuis janvier 2020 est proposée^a.

Tableau 6 : l'intervention *Odyssey Math*

Étape 1	Protocole <i>Mathématiques de la maternelle à la 4ème</i>		
Étape 2	22 études identifiées		
Étape 3	9 études éligibles (13 études non éligibles)		
Étape 4	3 études conformes aux normes WWC (6 études éligibles non conformes) ; elles sont repérées par le nom de l'auteur principal		
	DILEO <i>sans réserve</i>	WIKEJUMAR <i>sans réserve</i>	CORNELIUS <i>avec réserve</i>
Étape 5.1	n= 207 ; ES=0,44 IA = +17 ; p = 0,151 (non significatif)	n= 2 456 ; ES=0,02 IA = +1 ; p = 0,61 (non significatif)	n= 105 ; ES=0,446 IA = +18 ; p = 0,021 (significatif)
Étape 5.2	Critères antérieurs à janvier 2020 Effets potentiellement positifs ; IA moyen = 12 Critères postérieurs à janvier 2020 Effets incertains (la taille d'effet globale est positive mais n'est pas statistiquement significativement positive). Étendue des preuves : moyenne à importante		

Les Guides des pratiques

En plus des *Rapports d'intervention* dont nous venons de parler, le What Works Clearinghouse publie des *Guides des pratiques* portant sur un thème en lien direct avec l'enseignement et les apprentissages des élèves, thème clairement annoncé dans leur titre (par exemple

^a Indication donnée par l'auteur sans validation par le WWC (Nathalie ROQUES, 2021)

Aider les élèves en difficulté en mathématiques : le soutien scolaire à l'école et au collège).

Dans ces documents, des **recommandations** concrètes sont adressées aux enseignants, basées sur un examen rigoureux de la littérature scientifique qui s'appuie sur un protocole spécialement conçu pour chacun des *Guides*. Des études sont sélectionnées et examinées avec des critères proches de ceux détaillés auparavant pour les examens d'interventions, avec à la clé le calcul de tailles d'effets accompagnées de l'estimation de leur significativité statistique. Mais ces recommandations se fondent également sur l'avis des experts qui s'appuient sur les résultats précédemment évoqués mais aussi sur leur expérience professionnelle. Pragmatiques, ces *Guides des pratiques* illustrent chacune de leurs recommandations par des exemples d'exercices ou de problèmes mathématiques à proposer aux élèves et proposent également des solutions aux enseignants pour surmonter les obstacles qu'ils pourraient rencontrer dans la mise en œuvre de ces conseils.

→ **Trois niveaux de preuve**

A chacune des recommandations est associé l'un des 3 niveaux de preuve suivant :

- preuves solides
- preuves modérées
- preuves minimales.

Pour classer une recommandation dans l'une de ces 3 catégories, sept éléments (comme le niveau de preuve évalué pour l'ensemble des études, la qualité des tests utilisés lors de l'évaluation des élèves, l'intensité du lien entre la recherche analysée et la recommandation) vont être examinés à la lumière de critères définis pour chacun des

trois niveaux de preuve. Et là aussi des modifications ont été apportées en janvier 2020 qui intégreront de manière plus explicite et cohérente les résultats de l'examen des études tels qu'ils ont été décrits dans les chapitres précédents.

Le **tableau 7** propose une vue d'ensemble des 7 éléments et des critères de validation associés au niveau « preuves solides » (c'est la version mise à jour en janvier 2020 qui est proposée ici). Par exemple, si la recommandation est spécifiquement évaluée par les études sélectionnées et examinées par le WWC, alors l'intensité du lien entre recherche et recommandation (4^{ème} élément du tableau 7) sera considérée comme forte et donc le niveau « preuves solides » sera attribué à cet élément ; par contre si l'importance de la recommandation en tant que composante des interventions évaluées dans les études est variable, alors seul le niveau « preuves modérées » sera atteint (et ce sera le niveau « preuves minimales » si cette composante est considérée comme faible).

Tableau 7 : critères de validation pour le niveau de preuves solides

<i>Les 7 éléments</i>	<i>Pour obtenir le niveau « preuves solides »</i>
1. Niveau de preuve (qualité des études)	La recherche inclue des études conformes aux normes WWC avec un niveau de preuve de moyen à important (tableau 4)
2. Efficacité des actions évaluées	Les effets mesurés par les études examinées sont majoritairement positifs et aucun effet négatif n'est observé (tableau 2 ou 3).
3. Pertinence par rapport à l'objectif du <i>Guide</i>	La recherche a un rapport direct avec l'objectif – contextes, échantillons et résultats évalués adéquats.

4. Lien entre la recherche et la recommandation	La recommandation est directement évaluée par les études.
5. Confiance du comité	Le comité a un haut niveau de confiance sur l'efficacité de la recommandation.
6. Opinion des experts	Non applicable (concerne le niveau « preuves minimales » uniquement).
7. Qualité des tests	Les tests suivent des normes strictes.

Le niveau de preuve de la recommandation est le plus bas niveau obtenu pour l'un de ces 7 éléments. Ainsi, quand la recommandation est annoncée comme ayant une base de preuves solides cela signifie que tous les éléments ont le niveau « preuves solides ».

On retiendra qu'une recommandation de niveau « preuves solides » est soutenue par les résultats positifs d'études au design irréprochable et correctement menées, qui ne laissent pratiquement aucun doute quant au lien de cause à effet existant entre la recommandation et l'objectif fixé par les auteurs du *Guide des pratiques* pour une large population d'élèves.

Si le niveau « preuves modérées » est annoncé, cela signifie que les études au design irréprochable montrent des impacts positifs mais que des questions demeurent quant à la possibilité d'étendre les résultats à une large population d'élèves (par manque de répliques par exemple) ou quant à la démonstration du lien de cause à effet. Il est possible également que les études n'aient pas directement analysé la recommandation en question, mais que cette dernière puisse tout de même être reliée à l'analyse publiée.

Une recommandation est créditée du niveau « preuves minimales » quand les données suggèrent une relation entre la pratique recommandée et des résultats positifs mais que la recherche n'a pas

démontré que la recommandation est la seule cause de ces résultats positifs. Il est possible aussi que la recommandation soit tout simplement difficile à étudier d'une façon expérimentale rigoureuse, ou que les chercheurs n'ont pas encore étudié directement cette recommandation ou que l'efficacité de cette recommandation repose sur un ensemble de preuves faibles ou peu cohérentes.

Ainsi, les auteurs du *Guide des pratiques* intitulé *Aider les élèves en difficulté en mathématiques : le soutien scolaire en primaire et au collège* (avril 2009) assurent que « l'enseignement durant ces interventions devrait être explicite et systématique » (il s'agit de la recommandation n°3, voir plus loin sa description complète). Cette recommandation est associée à un niveau de preuve solide pour trois raisons. Tout d'abord, elle est basée sur un nombre important d'essais contrôlés randomisés et d'études quasi expérimentales de haute qualité (6 études au total sont conformes aux normes WWC) et qui ont examiné les effets d'un enseignement mathématique explicite et systématique chez des élèves dans des contextes variés. Également, ces études ont calculé 8 tailles d'effet toutes positives, dont 6 sont statistiquement et significativement supérieures à zéro. Enfin, ces études ont été menées par des équipes de chercheurs différentes et indépendantes.

→ **Publier les conclusions**

Le site *Find What Works* propose une synthèse de chaque *Guide* en présentant les recommandations et leur niveau de preuve. Le *Guide des pratiques* est gratuitement téléchargeable et fournit des informations sur les études conformes aux normes WWC exploitées par les auteurs, avec notamment leur taille d'effet calculée par le WWC et la durée d'implémentation de l'intervention étudiée.

Des vidéos sont également proposées pour certains *Guides* ainsi que des résumés qui ont comme objectifs d'être clairs et opérationnels (ils concernent soit les recommandations pratiques, soit leur assise scientifique).

Partie 2. L'enseignement des mathématiques

Maintenant que nous savons comment le What Works Clearinghouse examine les interventions et va parfois jusqu'à émettre des recommandations, il est temps d'analyser les résultats qui concernent plus particulièrement l'enseignement des mathématiques dans les établissements du secondaire (collège ou lycée) et qui ont été publiées dans les *Rapports d'intervention* et les *Guides des pratiques*. Comme cela a été souligné en introduction, c'est certainement dans ce champ disciplinaire que les informations issues d'études étrangères (anglo-saxonnes pour la plupart) seront les plus pertinentes car plus facilement transposables à notre contexte français si on les compare par exemple aux conclusions d'analyses portant sur la lecture.

Les *Rapports d'interventions* en mathématiques

Depuis la page du site *Find What Works* qui recense toutes les interventions examinées par le WWC, vous pourrez consulter 582 *Rapports d'interventions* dont 220 ont comme sujet les mathématiques. Certains examens n'ont pas identifié d'études conformes (avec ou sans réserve) aux normes WWC ; c'est le cas pour l'intervention *Singapore Math* par exemple. Dans ce cas, le *Rapport* se contente de mentionner qu'il n'est pas possible de conclure sur l'efficacité ou non de l'intervention par le biais d'une méta-analyse et donne la liste des articles dont l'examen a montré qu'aucun n'était exploitable. Pour d'autres, le WWC a identifié des études conformes aux normes WWC et exploité leurs données, pour calculer puis publier des résultats ; et c'est ce dont il va être question ci-dessous.

Un *Rapport d'intervention* concerne toujours l'examen d'une intervention associée à un *Protocole d'examen*^a qui définit notamment les niveaux scolaires et les compétences des élèves ciblés par le WWC. Certaines interventions (comme *I CAN Learn*) ont été examinées en utilisant plusieurs protocoles d'examens (en l'occurrence *Mathématiques de la maternelle à la 4^{ème}* et *Mathématiques de la 3^{ème} à bac + 1*), donnant lieu à autant de *Rapports d'intervention*. Dans chaque *Rapport*, les résultats sont présentés par domaine. Les trois domaines mathématiques sont *mathématiques générales, algèbre* et *géométrie*^b. Notons également que chaque *Rapport* se base sur le *Manuel des normes WWC* et le *Manuel des procédures WWC* en vigueur au moment de sa rédaction : ces deux ouvrages ont connu des versions différentes et les règles qui conduisent aux classements annoncés ci-dessous sont susceptibles d'avoir variées dans le temps. Toutes les interventions examinées jusqu'à présent par le WWC l'ont été avant la dernière mise à jour de janvier 2020.

Un fichier au format csv recensant les informations de toutes les interventions examinées par le WWC permet de les filtrer sur de très nombreux critères. C'est à partir de ce fichier que les interventions ont été filtrées ici sur les trois champs « domaine », « protocole d'examen » et « niveau d'étude », avec (par ordre chronologique de tri) :

^a L'architecture des protocoles d'examen a été modifiée en 2015 et leur nombre a été diminué (les 3 niveaux, école élémentaire, collège et lycée sont aujourd'hui regroupés en deux niveaux partagés au niveau de la classe de 4^{ème}).

^b Des domaines non mathématiques sont également parfois pris en compte lors des examens d'interventions, nous les avons laissés de côté ici.

1. Les domaines *algèbre, géométrie, mathématiques générales*
2. Tous les protocoles d'examen, sauf les protocoles concernant les jeunes enfants, les élèves présentent des handicaps et la formation des enseignants.
3. Tous les niveaux sauf les niveaux maternelle, CP, CE1, CE2 et CM1 (les interventions^a doivent donc s'intéresser aux élèves du CM2 jusqu'au niveau bac+1).

On obtient alors une liste de 15 examens d'interventions (c'est-à-dire de 15 *Rapports d'intervention*) proposant 19 résultats. Rappelons que chaque résultat est la combinaison d'un domaine, d'une taille d'effet et d'un indice d'amélioration, d'un classement de l'efficacité de l'intervention et d'un niveau de preuve (voir les **tableaux 2, 3 et 4**). Le nombre de résultats est supérieur à celui des examens d'interventions car pour certains d'entre eux, plusieurs domaines mathématiques ont été analysés.

→ **Les interventions dont les effets sont basés sur des preuves solides**

Sur les 19 résultats obtenus, seuls 6 montrent un niveau de preuve moyen à important (**tableau 8**)^b. Ce qui veut dire que 13 résultats ont un niveau de preuve faible (par exemple avec une seule étude, ou un seul terrain, ou une taille d'échantillon trop faible). Chaque résultat provient de l'examen d'une intervention. Cinq résultats, sont associés au domaine *mathématiques générales* ; le résultat présenté pour

^a On notera que (par exemple) le protocole *Mathématiques de la maternelle à la 4^{ème}* peut intégrer des études qui ne s'intéressent qu'au niveau maternelle.

^b La dernière colonne de ce tableau donne l'indice d'amélioration moyen et l'étendue de ces indices calculés pour chaque étude.

l'intervention *Cognitive Tutor*® concerne quant à lui le domaine *algèbre*^a. Tous ces examens ont été conduits en suivant les règles du *What Works Clearinghouse Procedures and Standards Handbook 3.0* (2014).

Tableau 8. Interventions et résultats avec une niveau de preuve moyen à important

Intervention	Nbre études	Taille échantillon	Classement de l'efficacité	Indice amélioration
<i>Accelerated Math</i> ®	6	5 206	Effets mixtes	5 (de -7 à 12)
<i>Cognitive Tutor</i> ®	5	12 182	Effets mixtes	4 (de -7 à 19)
<i>Connected Mathematics Project (CMP)</i>	2	3 062	Pas d'effet discernable	2 (de 0 à 4)
<i>Knowledge is Power Program (KIPP)</i>	4	19 542	Effets positifs	12 (de 7 à 20)
<i>Odyssey</i> ® <i>Math</i>	3	2 768	Effets potentiellement positifs	12 (de 1 à 18)
<i>Saxon Math</i>	5	8 855	Effets mixtes	8 (de -1 à 16)

Ces examens sont repris en détail ci-dessous avec une courte description de l'intervention (issues des *Rapports d'intervention*

^a Ce *Rapport d'intervention* publie également des résultats dans les deux autres domaines mathématiques mais l'étendue de leurs preuves est faible.

concernés, voir références en annexe) suivie d'un tableau synthétique présentant les principaux résultats pour chacune des études examinées par le WWC. Ces dernières sont repérées par le nom de leur auteur principal, vous en trouverez les références complètes dans les *Rapports d'intervention* concernés (les études conformes aux normes WWC sont signalées en **gras**).

→Les deux interventions qui « marchent »

Les enseignants sont bien sûr à la recherche d'interventions « qui marchent ». Les deux interventions qui montrent des effets positifs basés sur des preuves plutôt solides sont *Knowledge is Power Program (KIPP)* et *Odyssey® Math*.

Knowledge Is Power Program (KIPP)

L'examen a été mené en suivant les règles fixées par le protocole d'examen des établissements à charte^a.

L'objectif du WWC était ici d'évaluer l'effet sur les élèves d'être scolarisés dans un établissement *KIPP*. Les établissements *KIPP* suivent un ensemble de règles fixées par leur charte *Engagement pour l'excellence*. Cette charte est signée par les élèves, leurs parents et les enseignants : les élèves s'engagent à être assidus, les parents à soutenir leurs enfants dans ce sens et les enseignants à être disponibles. Financés par des fonds publics, les établissements *KIPP* fonctionnent de manière autonome, en dehors du contrôle direct du district scolaire local et sont exemptés de certaines règles et réglementations nationales ou locales. Ils jouissent notamment d'une

^a Ce protocole a comme vocation d'encadrer l'examen d'interventions s'intéressant à des établissements dont l'organisation s'apparente à nos établissements privés.

grande autonomie dans le recrutement, la formation et la révocation des chefs d'établissement ainsi que des enseignants. Ces établissements ont également des amplitudes horaires journalières plus importantes et une année scolaire plus longue que les établissements publics traditionnels.

Quatre études conformes aux normes WWC ont été examinées (voir **tableau 9** ci-dessous). Si les critères en vigueur depuis janvier 2020 avaient été utilisés, la conclusion aurait été que l'effet est potentiellement positif (et non positif comme annoncé dans le *Rapport d'intervention* publié en 2018), car si la taille d'effet globale est statistiquement positive, moins de 50 % de cette taille d'effet provient d'études conformes sans réserve aux normes WWC.

Tableau 9. Les études examinées pour l'intervention *Knowledge Is Power Program*

Références	Taille	Niveau	ES	Significatif ?
Tuttle (2015)	455	6 ^{ème} - 4 ^{ème}	0,18	Oui
Tuttle (2015)	13 624	6 ^{ème} - 4 ^{ème}	0,27	Oui
Tuttle (2015)	1 928	6 ^{ème} - 4 ^{ème}	0,20	Oui
Woodworth (2008)	3 535	CM2 - 6 ^{ème}	0,52	Oui
Bilan	19 542	CM2 - 4^{ème}	0,29	

Odyssey® Math

L'examen a été mené en suivant les règles fixées par le protocole d'examen *Mathématiques de la maternelle à la 4^{ème}*.

Odyssey® Math est une application informatique déployée sur ordinateurs et dédiée à l'enseignement des mathématiques de la maternelle à la 4^{ème}. Les élèves débutent par une évaluation diagnostique dont les résultats sont utilisés par les enseignants pour

créer des parcours pédagogiques personnalisés en fonction des compétences à acquérir. Les enseignants peuvent modifier les parcours d'apprentissage générés par le logiciel en ajoutant ou supprimant des tâches pour renforcer la pratique dans un domaine ou aligner le parcours d'apprentissage sur le plan des cours. Ce programme comprend des outils d'évaluation qui génèrent des résultats que les enseignants peuvent utiliser pour adapter leur enseignement. Trois études conformes aux normes WWC ont été examinées (voir **tableau 10** ci-dessous). En utilisant les critères en vigueur depuis janvier 2020, la conclusion aurait été que l'effet est incertain (et non potentiellement positif), car la taille d'effet globale n'est pas statistiquement significativement supérieure à zéro (comme cela a déjà été signalé au chapitre précédent).

Tableau 10. Les études examinées pour l'intervention *Odyssey® Math*

Références	Taille	Niveau	ES	Significatif ?
DiLeo (2017)	207	CM2	0,45	Non
Wikéjumar (2009)	2 456	CM1	0,02	Non
Cornelius (2013)	105	5 ^{ème} - 4 ^{ème}	0,46	Oui
Bilan	2 768	CM1 - 4^{ème}	0,31	

→ **Les quatre interventions qui ne « marchent pas »**

Les quatre autres interventions ont des effets mixtes ou non discernables, toujours avec un niveau de preuve moyen à important.

Accelerated Math

L'examen a été mené en suivant les règles fixées par le protocole d'examen *Mathématiques de la maternelle à la 4^{ème}*.

Accelerated Math est un logiciel qui propose des exercices et problèmes aux élèves de la maternelle à la terminale. Il fournit également aux enseignants des comptes-rendus leur permettant de suivre les progrès de leurs élèves. Ce logiciel crée des exercices personnalisés, donne immédiatement aux élèves des informations sur leurs réussites et échecs et génère pour les enseignants des rapports sur les progrès et le niveau de compétence de leurs élèves. Utilisé en lien avec le programme de mathématiques appliqué en classe, il complète ce dernier en proposant des exercices aux élèves en fonction de leur niveau. Cet outil permet également aux enseignants de différencier leur enseignement grâce au suivi individuel des progrès de chaque élève. Une fois que les élèves montrent qu'ils maîtrisent une compétence mathématique, le logiciel leur attribue automatiquement de nouvelles activités en fonction de la série d'activités à laquelle ils étaient affectés.

Six études conformes aux normes WWC ont été examinées (voir **tableau 11** ci-dessous).

Tableau 11. Les études examinées pour l'intervention *Accelerated Math*

Références	Taille	Niveau	ES	Significatif ?
Caputo (2007)	70	6 ^{ème}	-0,01	Non
Lambert (2014)	666	CE1-CM2	0,22	Non
Lambert (2009)	633	5 ^{ème} - 3 ^{ème}	0,07	Non
Lehmann (2005)	1 131	CM1 - 6 ^{ème}	0,06	Non
Ysseldyke (2007)	700	CE1-CM2	0,15	Non
Ysseldyke (2007)	2 006	CE2 - 6 ^{ème}	0,29	Oui
Bilan	5 206	CE1 - 3^{ème}	0,13	

Cognitive Tutor

L'examen a été mené en suivant les règles fixées par le protocole d'examen *Mathématiques de la 3^{ème} à bac + 1*. Le domaine associé aux résultats est *algèbre*.

Ce programme de mathématiques se concentre sur la façon dont les élèves du secondaire conçoivent et apprennent les mathématiques. Les élèves utilisent des manuels qui proposent un travail de type collaboratif. Ils doivent résoudre des problèmes en s'engageant dans des investigations puis expliquer et comparer leurs solutions. Les discussions entre élèves sur des sujets mathématiques sont favorisées car ils sont encouragés à expliciter leurs raisonnements. Deux séances par semaine se déroulent en salle informatique où les élèves utilisent une application qui propose des activités diverses adaptées à leur propre rythme. Les problèmes mathématiques sont conçus pour souligner les liens entre les représentations verbales, numériques, graphiques et algébriques.

Six études conformes aux normes WWC ont été examinées (voir **tableau 12** ci-dessous).

Tableau 12. Les études examinées pour l'intervention *Cognitive Tutor*

Références	Taille	Niveau	ES	Significatif ?
Cabalo (2007)	344	4 ^{ème} -terminale	-0,18	Non
Campuzano (2009)	270	4 ^{ème} - 3 ^{ème}	-0,23	Non
Pane (2014)	11 066	4 ^{ème} -terminale	0,08	Non
Ritter (2007)	255	3 ^{ème}	0,38	Non
Wolfson (2008)	247	3 ^{ème} -terminale	0,49	Oui
Bilan	19 542	CM2 - 4^{ème}	0,29	

Connected Mathematic Project (CMP)

L'examen a été mené en suivant les règles fixées par le protocole d'examen *Mathématiques de la maternelle à la 4^{ème}*.

Ce programme de mathématiques destiné aux élèves de la 6^{ème} à la 4^{ème} est basé sur l'investigation. Il confronte les élèves à des problèmes interactifs et des situations de la vie quotidienne pour qu'ils puissent explorer des concepts mathématiques. Ce programme cherche à faciliter l'acquisition de connaissances et de compétences mathématiques tout en permettant aux élèves de découvrir les liens qui existent entre les différents domaines mathématiques mais aussi entre les mathématiques et les autres disciplines scientifiques.

Deux études conformes aux normes WWC ont été examinées (voir **tableau 13** ci-dessous).

Tableau 13. Les études examinées pour l'intervention *Connected Mathematic Project*

Références	Taille	Niveau	ES	Significatif ?
Cai (2011)	606	4 ^{ème}	0,10	Non
Ridgway (2002)	2 456	6 ^{ème} - 4 ^{ème}	0,00	Non
Bilan	3 062	6^{ème} - 4^{ème}	0,05	

Saxon Math

L'examen a été mené en suivant les règles fixées par le protocole d'examen *Mathématiques de la maternelle à la 4^{ème}*.

Ce programme destiné aux élèves de la maternelle à la terminale propose des séances qui suivent le déroulé suivant :

1. L'enseignant présente un ou plusieurs nouveaux concepts mathématiques, en utilisant des exemples, en animant des conversations et en facilitant l'intégration de ces nouveaux concepts à ceux déjà connus.

2. L'enseignant guide les élèves sur des problèmes pratiques liés aux nouveaux concepts.

3. Les élèves s'engagent individuellement dans une pratique écrite visant à maîtriser les nouvelles compétences et à conserver la maîtrise des concepts enseignés précédemment.

Toutes les cinq leçons, les élèves complètent des évaluations écrites et sommatives. Les résultats de ces évaluations permettent aux enseignants d'orienter leur enseignement et d'informer les élèves et les parents sur le niveau de compétences acquises.

Cinq études conformes aux normes WWC ont été examinées (voir **tableau 14** ci-dessous).

Tableau 14. Les études examinées pour l'intervention *Saxon Math*

Références	Taille	Niveau	ES	Significatif ?
Agodini (2013)	2 045	CE1	0,07	Non
Crawford (1986)	78	4 ^{ème}	0,41	Non
Good (2006)	745	CE1 et CE2	0,11	Non
Resendez (2005)	3 054	5 ^{ème}	0,17	Non
Resendez (2005)	2 933	6 ^{ème}	0,25	Non
Bilan	8 855	CE1 - 4^{ème}	0,20	

Les Guides des pratiques en mathématiques

Nous savons déjà que le What Works Clearinghouse s'adresse également d'une façon pragmatique aux enseignants en exposant dans ces *Guides des pratiques* des recommandations concrètes sur des sujets précis. Vous trouverez quatre *Guides des pratiques* concernant l'enseignement des mathématiques et s'intéressant plus particulièrement à la réussite des élèves au secondaire^a sur le site *Find What Works*.

- *Améliorer la résolution des problèmes du CMI à la 4^{ème}* (mai 2012, révisé en mai 2018)
- *Enseigner des stratégies pour renforcer les connaissances en algèbre au collège et au lycée* (avril 2015)
- *Développer un enseignement efficace des fractions de la maternelle à la 4^{ème}* (septembre 2010)
- *Aider les élèves en difficulté en mathématiques : le soutien scolaire à l'école primaire et au collège* (avril 2009)

A eux quatre, ils proposent en tout 21 recommandations (liste complète dans le **tableau 15**, à la fin de ce chapitre). Parmi ces recommandations, quatre seulement sont basées sur des preuves solides ; elles sont issues des deux *Guides* intitulés *Améliorer la résolution des problèmes du CMI à la 4^{ème}* et *Aider les élèves en difficulté en mathématiques : le soutien scolaire à l'école primaire et au collège*. Je vous en propose une description détaillée ci-dessous,

^a Les deux guides *Enseigner les maths aux jeunes enfants* (2013) et *Encourager les filles en maths et en sciences* (2007) ont été laissés de côté ici.

qui sont des traductions libres d'extraits de textes originaux publiés en américains.

→ *Améliorer la résolution des problèmes du CMI à la 4^{ème}*

Ce *Guide* propose 5 recommandations, dont deux (la recommandation n°2 et la recommandation n°3) sont basées sur des preuves solides.

Recommandation n°2. Aider les élèves dans leur réflexion et leur analyse du processus de résolution d'un problème

Les élèves apprennent plus facilement les mathématiques et résolvent mieux les problèmes lorsqu'ils comprennent les différentes étapes permettant de résoudre ces derniers. Analyser les procédures mises en œuvre lors de la résolution de problèmes permet aux élèves de comprendre ce qu'ils font et pourquoi ils le font, d'élaborer les étapes qu'ils envisagent de suivre et de relier de nouveaux concepts à ceux qu'ils connaissent déjà. Plus les élèves réfléchissent au processus de résolution de problèmes qu'ils mettent en œuvre, meilleur sera leur raisonnement mathématique - et leur capacité à appliquer ce raisonnement à de nouvelles situations. Les enseignants devraient apprendre aux élèves à analyser les stratégies qu'ils emploient lors de la résolution des problèmes. Bien que l'objectif ultime soit que les élèves analysent et réfléchissent par eux-mêmes lorsqu'ils résolvent un problème, les enseignants peuvent avoir besoin de les aider quand ils introduisent une nouvelle activité ou un nouveau concept. Par exemple, un enseignant peut fournir des instructions et les mettre en œuvre pour modéliser sa réflexion lorsqu'il résout un problème à voix haute. En outre, un enseignant peut utiliser les idées énoncées par les élèves pour leur permettre d'améliorer leurs analyses et leurs réflexions autour de la résolution de problèmes.

Cette recommandation s'appuie sur l'examen de 9 études conformes aux normes WWC qui a permis de calculer 17 tailles d'effet dont 9 sont statistiquement significativement supérieures à zéro, 8 sont positives et non statistiquement significativement supérieures à zéro, et une est négative et non statistiquement significativement inférieure à zéro ; elle s'articule autour des trois axes décrits ci-dessous.

1. Fournir aux élèves une liste d'instructions pour les aider à analyser leur travail et à réfléchir au processus de résolution des problèmes

Cela peut être une liste de questions que les élèves doivent se poser et auxquelles ils doivent répondre pendant la résolution de problèmes ou des listes d'instructions qui les aident à franchir les étapes permettant de résoudre un problème. Les questions posées par les enseignants doivent inciter les élèves à réfléchir au processus de résolution de problèmes ; de la même manière, des listes de tâches peuvent soutenir la réflexion des élèves tout au long de ce processus. Il convient d'inviter les élèves à expliquer et à justifier leur réponse pour chaque instruction, oralement ou par écrit. Les élèves peuvent utiliser ces instructions pour travailler de manière indépendante, en petits groupes ou même pour résoudre des problèmes sur un ordinateur.

2. Modéliser le processus de résolution des problèmes

L'enseignant devrait énoncer à voix haute non seulement la réponse apportée pour chacune des instructions, mais également les raisons pour lesquelles ces instructions ont été mises en œuvre. Il peut également donner la solution de l'étape qui a été franchie et demander aux élèves d'expliquer pourquoi cela fonctionne. Veiller à utiliser une instruction à chaque étape du processus de résolution du problème, par exemple, lors de la première lecture du problème, lors de la recherche de stratégies pour résoudre le problème, et après la résolution du problème.

3. Utiliser les idées des élèves lors de la résolution d'un problème afin qu'ils puissent développer leur aptitude à analyser et à réfléchir

En s'appuyant sur les idées des élèves, les enseignants peuvent aider ces derniers à clarifier et affiner leur façon d'analyser et de réfléchir lors de la résolution d'un problème. Le dialogue enseignant-élève peut inclure des questions guidées pour permettre aux élèves d'établir une méthode d'analyse et de réflexion qui leur semble logique et peut se révéler particulièrement pertinent quand les élèves n'aiment pas utiliser les instructions fournies par l'enseignant ou quand ils ont des difficultés à comprendre et à utiliser ces instructions.

Recommandation n°3. Enseigner aux élèves comment utiliser des représentations visuelles

Une tâche majeure pour tout élève engagé dans la résolution de problèmes consiste à traduire les informations d'un problème en une équation utilisant des symboles (une expression arithmétique ou algébrique). L'utilisation de représentations visuelles aide les élèves à résoudre des problèmes en permettant de faire le lien entre les relations qui unissent les grandeurs présentes dans le problème d'une part, et les opérations mathématiques qui vont permettre de résoudre le problème, d'autre part. Les enseignants doivent montrer aux élèves comment s'engager dans la résolution d'un problème à l'aide d'une représentation visuelle en sélectionnant les représentations visuelles appropriées.

Cette recommandation s'appuie sur l'examen de 6 études conformes aux normes WWC qui a permis de calculer 15 tailles d'effet dont 8 sont statistiquement significativement supérieures à zéro, 6 sont positives et non statistiquement significativement supérieures à zéro, et une est négative et non statistiquement significativement

inférieure à zéro ; elle s'articule autour des trois axes décrits ci-dessous.

1. Sélectionner des représentations visuelles adaptées aux élèves et aux problèmes à résoudre

Les enseignants doivent choisir la représentation visuelle qui conviendra le mieux à leurs élèves et l'utiliser systématiquement pour des problèmes similaires. Par exemple, si un enseignant représente un problème de ratio ou de proportion à l'aide d'un diagramme que les élèves ont trouvé utile pour déterminer l'équation nécessaire à la résolution de ce problème, il doit continuer à utiliser ce même diagramme lorsque les élèves travaillent sur des problèmes similaires. Les élèves peuvent avoir besoin de temps pour s'exercer à utiliser des représentations visuelles et peuvent avoir du mal à les utiliser avec succès.

2. Mettre en place des groupes de discussions et de « réflexions à voix haute » pour apprendre aux élèves à représenter les problèmes visuellement

Lorsqu'une nouvelle représentation visuelle ou un nouveau type de problème est enseigné, l'enseignant devrait montrer comment représenter le problème en réfléchissant à voix haute aux décisions qu'il prend lorsqu'il relie le problème à la représentation, en explicitant ses idées au moment d'aborder le problème et en expliquant ses choix. Il doit expliquer comment il a identifié le type de problème - tel qu'un problème de proportion, de ratio ou de pourcentage - en se basant sur les éléments mathématiques du problème. Il doit également apprendre aux élèves à identifier les informations pertinentes ou essentielles à la résolution du problème.

3. Montrer aux élèves comment convertir les informations représentées visuellement en notation mathématique

Après avoir représenté visuellement les informations pertinentes dans un problème, montrer comment chaque partie de la représentation visuelle peut être traduite en notation mathématique. Quand la correspondance entre la représentation visuelle et l'équation n'est pas immédiate, les enseignants doivent expliciter les liens entre cette représentation et la résolution du problème.

→ *Aider les élèves en difficulté en mathématiques : le soutien scolaire en primaire et au collège*

Ce *Guide* propose 8 recommandations, dont deux (la recommandation n°3 et la recommandation n°4) sont basées sur des preuves solides et sont décrites ci-dessous. Elles concernent les séances de soutien dédiées à des groupes d'élèves en difficulté et le soutien mis en place individuellement.

Recommandation n°3. L'enseignement durant ces interventions devrait être explicite et systématique : proposer des modèles et des guides pratiques pour résoudre les problèmes, expliquer les processus intellectuels mis en œuvre, revenir sur les erreurs commises et proposer des révisions fréquentes.

Un enseignement explicite s'appuie sur des modèles clairs qui ont comme but la résolution de problèmes types en utilisant une large palette d'exemples ; les élèves doivent s'entraîner fréquemment pour utiliser de nouvelles compétences et stratégies ; ils doivent avoir l'opportunité de réfléchir à voix haute (par exemple pour expliquer leurs décisions et les étapes qu'ils ont suivies) ; les élèves ont également besoin de retours explicites sur leur travail.

Cette recommandation s'appuie sur l'examen de 6 études conformes aux normes WWC qui a permis de calculer 8 tailles d'effet dont 6 sont statistiquement significativement supérieures à zéro et 2 sont

positives et non statistiquement significativement supérieures à zéro ; elle s'articule autour des trois axes décrits ci-dessous.

1. S'assurer que l'enseignement est systématique et explicite, avec des exemples clairs, nombreux, de difficultés variées et expliqués par l'enseignant

Pour être considéré comme systématique, l'enseignement mathématique (ou les leçons) devrait permettre une construction progressive des compétences en suivant un ordre logique et en fournissant aux élèves de nombreuses applications pour chaque concept. Il doit débiter par un exposé clair et sans ambiguïté des concepts et des méthodes (de calculs, de résolution de problèmes) décrits pas à pas pour permettre aux élèves de comprendre les éléments mathématiques qui se cachent derrière toutes ces procédures. Les enseignants ne doivent pas se limiter à n'expliquer qu'une ou deux façons d'aborder un problème, ni à des problèmes trop simples.

2. Donner aux élèves l'opportunité de résoudre des problèmes en groupe et de communiquer leurs stratégies de résolution

L'enseignant qui s'engage dans un enseignement explicite doit s'appuyer sur un enseignement structuré où il joue un rôle actif en transférant progressivement le travail aux élèves. Cette phase débute par la résolution de problèmes par l'enseignant avec les élèves. De façon graduelle, les élèves complètent un nombre plus important d'étapes durant la résolution avec une guidance de la part de l'enseignant en diminution progressive. Les élèves devraient finalement accéder à une pratique indépendante et résoudre les problèmes avec peu ou pas de soutien de la part de l'enseignant. Les enseignants doivent donner des retours clairs sur ce que l'élève a fait correctement et sur les progrès qu'il lui reste à faire. Cette étape de feedback peut aussi être l'occasion d'apporter une explication ou une clarification supplémentaire sur les instructions.

3. S'assurer que l'enseignement comprend des révisions fréquentes à chaque session

Des révisions fréquentes donnent l'opportunité aux élèves de s'exercer sur des sujets précédemment vus de façon approfondie. Ces révisions assurent que les connaissances sont maintenues dans le temps et permettent aux élèves de faire des connexions entre des concepts mathématiques variés.

Recommandation n° 4. Les interventions devraient inclure un enseignement sur la résolution des problèmes verbaux^a basé sur l'identification de structures communes

Les intervenants doivent mettre en œuvre un enseignement explicite et systématique portant sur la résolution de problèmes verbaux en s'appuyant sur la structure sous-jacente de ces problèmes. Les problèmes verbaux simples donnent un sens aux opérations mathématiques telles que la soustraction ou la multiplication. Lorsque les élèves réussissent à définir la structure sous-jacente d'un problème verbal, ils ont non seulement plus de réussite dans la résolution des problèmes, mais ils peuvent également mieux comprendre les idées mathématiques plus profondes sur lesquelles sont basés ces problèmes. L'enseignement systématique devrait s'intéresser également aux liens structurels entre les problèmes connus et familiers et les problèmes nouveaux et non familiers. En explicitant ces liens, les élèves sauront quand appliquer les méthodes de résolution qu'ils ont apprises.

Cette recommandation s'appuie sur l'examen de 9 études conformes aux normes WWC qui a permis de calculer 14 tailles d'effet dont 11

^a Les énoncés de ces problèmes contiennent beaucoup de texte et peu de notations mathématiques.

sont statistiquement significativement supérieures à zéro et 3 sont positives et non statistiquement significativement supérieures à zéro ; elle s'articule autour des deux axes décrits ci-dessous.

1. Apprendre aux élèves à identifier les structures associées à des types de problèmes particuliers, puis à utiliser les méthodes adaptées à chaque type de problème

Les élèves doivent être explicitement informés des principales caractéristiques structurelles sous-jacentes à chaque type de problème. Les types de problèmes sont définis par des structures mathématiques communes à un ensemble de problèmes. Par exemple, les problèmes d'évolution dans le temps décrivent des situations dans lesquelles une quantité (comme un nombre d'enfants ou de crayons) est soit augmentée soit diminuée dans le temps. En revanche, les problèmes de comparaison ne mettent pas en jeu le facteur temps et sont construits sur des comparaisons entre des éléments appartenant à des ensembles différents (poires et pommes, garçons et filles, articles chauds et froids).

Pour permettre aux élèves de comprendre et de résoudre chaque type de problème, les enseignants devraient d'abord expliquer les règles qui structurent la solution type (ou les questions-guides qui mènent à l'équation permettant de résoudre le problème) à travers des exemples entièrement ou partiellement corrigés, puis laisser les élèves travailler en binôme.

2. Apprendre aux élèves à reconnaître la structure sous-jacente commune à des problèmes familiers et non familiers et à utiliser les méthodes de résolution des problèmes familiers pour résoudre des problèmes non familiers

Un problème familier et connu paraît parfois à l'élève comme étant un problème non familier et nouveau en raison de changements aussi superficiels que le format utilisé, des changements dans le vocabulaire

clé (moitié, un demi, $\frac{1}{2}$), ou l'inclusion d'informations non pertinentes. Pour faciliter le transfert de la solution d'un problème familier connue à un problème non familier, les élèves doivent d'abord identifier les informations du problème qui ne sont pas pertinentes pour ensuite reconnaître la structure sous-jacente du problème. Les élèves qui savent reconnaître et résoudre un type de problème avec des nombres entiers devraient savoir qu'ils peuvent appliquer la même stratégie à un problème similaire sur le plan structurel mais qui semble différent en raison de changements de formulations et de la présence d'éléments de récit supplémentaires.

Tableau 15 : liste des 21 recommandations des *Guides des pratiques* en mathématiques et leur niveau de preuve

<i>Améliorer la résolution des problèmes du CMI à la 4^{ème}</i>	
1. Préparer des problèmes et les utiliser durant les cours en classe entière	Preuves minimales
2. Aider les élèves dans leur réflexion et leur analyse du processus de résolution d'un problème	Preuves solides
3. Enseigner aux élèves comment utiliser des représentations visuelles	Preuves solides
4. Exposer les élèves à des stratégies multiples de résolution de problèmes	Preuves modérées
5. Aider les élèves à reconnaître et articuler concepts mathématiques et notations	Preuves modérées

<i>Aider les élèves en difficulté en mathématiques : le soutien scolaire à l'école primaire et au collège</i>	
En classe entière	
1. Évaluer tous les élèves pour identifier ceux qui risquent d'être en difficulté et leur proposer une aide	Preuves modérées
Séances de soutien aux élèves en difficulté (en groupe ou individuellement avec parfois des enseignants spécialisés)	

2. Les séances de soutien devraient se concentrer intensément sur le traitement approfondi des nombres entiers de la maternelle au CM2 et sur les nombres rationnels du CM1 à la 4 ^{ème} en s'appuyant sur des notions clés validées par des experts.	Preuves minimales
3. L'enseignement durant ces interventions devrait être explicite et systématique : proposer des modèles et des guides pratiques pour résoudre les problèmes, expliquer les processus intellectuels mis en œuvre. revenir sur les erreurs commises et proposer des révisions fréquentes	Preuves solides
4. Les interventions devraient inclure un enseignement sur la résolution des problèmes verbaux basées sur l'identification de structures communes.	Preuves solides
5. L'enseignement devrait permettre aux élèves de travailler sur les concepts mathématiques à l'aide de représentations visuelles et les intervenants devraient être formés à leur utilisation.	Preuves modérées
6. A tous les niveaux d'étude, les interventions devraient à chaque session revenir pendant 10 minutes sur des notions d'arithmétique de base.	Preuves modérées
7. Évaluer les progrès des élèves recevant cet enseignement supplémentaire et tous les élèves à risque.	Preuves minimales
8. Inclure des stratégies sur la motivation durant ces interventions	Preuves minimales

<i>Développer un enseignement des fractions efficace, de la maternelle à la 4^{ème}</i>	
1. Favoriser chez les élèves une compréhension informelle des notions de partage et de proportionnalité pour développer les premiers concepts concernant les fractions.	Preuves minimales
2. Aider les élèves à concevoir que les fractions sont des nombres et qu'elles élargissent le système numérique au-delà des nombres entiers. Utiliser les droites numériques	Preuves modérées

comme outils de représentation central en enseignant ce fait ou d'autres concepts liés aux fractions depuis les premiers niveaux d'étude et au-delà.	
3. Aider les élèves à comprendre les procédures de calcul avec les fractions.	Preuves modérées
4. Permettre aux élèves de comprendre les stratégies à mettre en œuvre pour résoudre des problèmes en utilisant des fractions, des ratios et la proportionnalité avant de leur proposer la technique du « produit en croix » comme procédure pour résoudre de tels problèmes.	Preuves minimales
5. La formation professionnelle devrait donner la priorité à l'amélioration de la compréhension par les enseignants des fractions et à la manière de les enseigner.	Preuves minimales

<i>Enseigner des stratégies pour renforcer les connaissances en algèbre au collège et au lycée</i>	
1. Utiliser des problèmes résolus pour engager les élèves à analyser des raisonnements et des stratégies algébriques.	Preuves minimales
2. Apprendre aux élèves à reconnaître précisément la structure d'une expression algébrique.	Preuves minimales
3. Enseigner aux élèves comment choisir intentionnellement une stratégie algébrique parmi d'autres alternatives pour résoudre des problèmes.	Preuves modérées

Partie 3. Discussion sur les méta-analyses

Comme annoncé au début de ce livre, l'utilité même des méta-analyses portant sur l'enseignement et plus largement le bien fondé des règles de décision posées par l'*evidence based education* (voir plus loin) font débat dans la communauté des chercheurs en sciences de l'éducation, et ce particulièrement en France. Mais même chez les scientifiques qui considèrent ces outils avec un peu plus d'intérêt et de confiance, certains points de méthode font l'objet de discussions. Quelques éléments de réflexion sont proposés ci-dessous pour y voir plus clair.

Défauts et qualités des *Rapports d'intervention* et des *Guides des pratiques*

Comme tout élément participant au débat scientifique, les deux outils phares publiés par le What Works Clearinghouse peuvent être (et ont été) discutés voire critiqués à plusieurs niveaux^a. Mettons tout de même au crédit du WWC de mettre à la disposition d'un large public une grande quantité de données et de résultats accompagnés de documents qui clarifient les processus structurant la réalisation de leurs méta-analyses. Également on peut souligner l'évolution constante de leurs procédures qui atteste d'une véritable volonté d'améliorer les méthodes utilisées (depuis 2008, six versions des *Manuels de procédures et de normes WWC* se sont succédées ; pour les dernières, des versions provisoires pouvaient être commentées par les internautes).

^a Straight talk of evidence, *How What Works Clearinghouse reviews sometimes make ineffective education programs appear effective (part two in a series)*, décembre 2017

Revenons une dernière fois sur ces deux outils de communication que sont les *Rapports d'intervention* et les *Guides des pratiques*.

→ **Les Rapports d'intervention**

Ils sont au cœur même du travail mené par le What Works Clearinghouse et symbolisent à eux seuls la volonté de mettre en œuvre les concepts de l'*evidence based education*. En effet, les *Rapports d'intervention* constatent méthodiquement et scientifiquement les effets d'une intervention sur les élèves dans un domaine spécifique à partir de données d'études quantitatives sélectionnées pour la qualité de leur design. Aucune interprétation des résultats n'est proposée.

Mais quelle que soit la rigueur qui anime ici les chercheurs, la subjectivité s'invite sans même qu'on s'en aperçoive. C'est le cas notamment quand plusieurs éléments d'appréciation sont fournis pour un même examen, et qu'il reste encore à décider vers lequel tourner son attention. Deux résultats permettent d'apprécier les résultats d'une intervention : son efficacité (**tableaux 2** ou **3**) et le niveau de confiance qu'on peut lui attribuer (**tableau 4**). C'est l'efficacité de l'intervention qui est mise en avant par le site *Find What Works*. Dans ce document, j'ai choisi de m'appuyer plutôt sur le niveau de preuve (curieusement absent des résumés proposés sur le site internet). Elle me semble en effet avoir au moins autant d'importance dans un processus mettant clairement en avant la force de la preuve. Quand le WWC détermine par exemple qu'une intervention a des effets positifs (nous avons vu que cela se traduit par une taille d'effet statistiquement et significativement positive pour au moins deux études conformes aux normes WWC dont au moins un essai contrôlé randomisé), mais que le niveau de preuve est faible (et donc qu'un seul terrain est étudié

ou bien que la taille totale de l'échantillon est inférieure à 350), on se trouve face à un résultat dont la **validité externe** n'est pas démontrée. Et inversement, quand l'efficacité est indiscernable ou mixte mais que le niveau de preuve est moyen ou important, ne conviendrait-il pas de considérer avec attention cette probable absence d'effets ?

L'intérêt des *Rapports d'intervention* se trouve également limité par leur manque d'utilité pratique. Les interventions examinées sont en effet la plupart du temps des programmes clé en main complets. Leur examen ne répond donc pas à l'attente des enseignants qui ne souhaitent pas bouleverser entièrement leur pratique (c'est de toute façon pratiquement impossible en France) mais plutôt les modifier en y intégrant par exemple un élément pédagogique qu'ils identifieraient comme pertinent dans le contexte qui est le leur. Et ce sont finalement plutôt les éléments qui constituent les interventions qui intéressent les enseignants. Ils ne sont pas différenciés dans l'analyse menée par le WWC et il n'est donc pas possible de savoir comment tel ou tel élément de l'intervention^a participe à l'effet évalué.

Mais la perplexité des enseignants va encore plus loin, quand ils se trouvent face à des résultats contradictoires qui mettent la pertinence même des examens menés par le WWC en question. Ainsi la lecture des deux *Rapports* concernant les interventions *Odyssey Math* et *Accelerated math*, toutes les deux basées sur l'utilisation d'un logiciel permettant une différenciation pédagogique en fonction du niveau de l'élève, et affichant des efficacités diamétralement opposées avec un même niveau de preuve (plutôt important), ne permet pas de comprendre pourquoi ces deux interventions ont des effets différents.

^a Les éléments constitutifs et caractéristiques de l'intervention *Odyssey Math* pourraient être le fait de travailler sur ordinateur ou la qualité des exercices proposés ou celle des évaluations (ces éléments devant être définis en rapport avec l'enseignement déployé dans les groupes comparaison).

Toujours dans le même registre, le nombre limité des domaines étudiés (seuls trois domaines sont identifiés pour les mathématiques : *algèbre, géométrie, mathématiques générales*) empêche une analyse plus fine des compétences acquises par les élèves grâce à telle ou telle intervention^a.

Pour toutes ces raisons, qui finalement se résument à un manque de finesse dans l'analyse des interventions, les conclusions des *Rapports d'intervention* peuvent donc ne pas présenter d'utilité concrète immédiate pour les enseignants.

→ **Les Guides des pratiques**

Bien évidemment, les *Guides des pratiques* répondent mieux à cette demande, car ils proposent des recommandations ciblées sur des segments précis du travail des enseignants pouvant être rapidement mises en œuvre dans leurs classes. Certaines d'entre elles ne sont basées que sur des preuves considérées comme minimales, et s'appuient essentiellement sur l'avis des experts consultés et donc au moins partiellement sur leur subjectivité, ce qui pourrait être considéré comme un défaut. Ainsi les auteurs du *Guide* intitulé *Améliorer la résolution des problèmes du CMI à la 4^{ème}*, tout en reconnaissant que son niveau de preuve est minimal, placent volontairement la recommandation n°1 en tête de liste des 5 recommandations énoncées dans ce *Guide* et soulignent même lui apporter tout leur crédit. Ils annoncent également que, d'une façon générale, « *une base de preuve minimale ne veut pas dire qu'elle (la recommandation, NDLR) est moins importante qu'une recommandation dont le niveau de preuve est modéré ou solide* ».

^a Voir à ce sujet la discussion proposée par James HIEBERT et Douglas GROUWS (2007)

Finalement, on pourrait présenter la coexistence des *Rapports d'intervention* et des *Guides des pratiques* comme un mariage plus ou moins heureux où les défauts de l'un seraient compensés par les qualités de l'autre, et vice versa : la rigueur objective et le faible intérêt pratique des *Rapports d'intervention* sont effectivement contrebalancés par les éléments concrets mais parfois subjectifs des *Guides des pratiques*.

Discussion sur les méta-analyses

L'élaboration, la conduite et l'exploitation de méta-analyses se sont déployées dans le cadre plus vaste d'un mouvement baptisé *evidence based education*^a. On pourrait traduire cette expression anglaise par « pratiques éducatives basées sur des preuves ». L'*evidence based education* s'inspire de l'*evidence based medicine* qui organise la prise de décision dans le domaine médical en se fondant notamment sur les résultats et les conclusions de méta-analyses. Les règles qui ont été mises en place dans les années 90^b ont profondément bouleversé la médecine et il ne viendrait à l'idée de personne actuellement de les remettre en question.

Le fondement même de ces méthodes de recherche est d'exclure toutes les conclusions qui ne seraient pas construites sur la base d'un raisonnement scientifique et d'écarter par là-même toute interprétation subjective. Cet objectif extrêmement séduisant attire de nombreux chercheurs qui affirment alors avoir démontré de façon

^a Voir à ce sujet Olivier REY (2006)

^b Cette date correspond au début de l'enseignement de l'*Evidence Based Medicine*. C'est pourtant dans un article écrit par Gene GLASS en 1976 que le terme « méta-analyse » apparaît pour la première fois et que les bases de cette nouvelle science sont posées en s'appuyant sur les résultats d'études issues du domaine des sciences de l'éducation.

indiscutable leurs conclusions. Mais la démarche est exigeante et doit contrecarrer la puissance de nos convictions qui peuvent infiltrer les discours les plus austères et les mieux habillés. Se prévaloir de l'*evidence based education* sans accepter d'en suivre rigoureusement les préceptes pourrait alors à juste titre paraître suspect : déguiser (parfois involontairement) une opinion en une preuve scientifique est une tentation à laquelle beaucoup succombe. Revenons sur certains points qui sont souvent évoqués par les promoteurs des méta-analyses.

→ **La sélection des études**

Cette étape est bien évidemment déterminante et conditionne toute la suite du processus. Des critères de sélection ont été définis par le WWC pour plusieurs types d'études (nous nous sommes uniquement intéressés ici aux études procédant par comparaison de groupes, d'autres critères peuvent être utilisés, par exemple pour les études de cas). Depuis 2002, les normes WWC qui fixent ces règles se sont étoffées et précisées, et elles ont été publiées pour la première fois en 2017 dans un document séparé de celui définissant les procédures : le *Manuel des procédures et des normes WWC (3.0)* a donné naissance au *Manuel des procédures WWC (4.0)* et au *Manuel des normes WWC (4.0)*, démontrant par là même l'importance toute particulière de cette étape.

Sélectionner des études c'est bien entendu en écarter quelques unes, et si vous jetez un œil sur la liste des publications rejetées lors de l'examen d'une intervention (voir l'examen de l'intervention *Odyssey Math*), vous verrez qu'elles sont bien plus nombreuses que celles qui ont été retenues. La satisfaction des auteurs de ces dernières (qui se mesure à la mention en première page d'un label « conforme aux normes WWC ») est aussi intense que la déception de ceux dont le travail n'est pas retenu. Exclure, c'est évidemment poser un jugement

négligé sur certains travaux de recherche et l'intransigeance de cette sélection associée à l'importance donnée de nos jours aux conclusions des méta-analyses, donne aux auteurs de méta-analyses des airs de supériorité qui en agace plus d'un.

→ **Les calculs statistiques**

Les méta-analyses conduites par le WWC et les études primaires qui les alimentent mettent en œuvre des concepts statistiques qui peuvent paraître incompréhensibles pour une majorité de lecteurs car elles requièrent une expertise qui dépassent largement les connaissances des enseignants à la recherche d'informations concrètes. Il est probable que ce traitement mathématique de données quantitatives détourne certains membres de la communauté éducative de ces publications parfois considérées comme austères et inabordables^a. C'est pour cette raison et pour faciliter l'exploitation des résultats par le plus grand nombre que le WWC propose une simplification des résultats sous la forme d'un indice d'amélioration (*IA*) ; cet indice est calculé pour chaque étude primaire sélectionnée, mais également pour l'ensemble des études collectées lors de l'examen d'une intervention, toujours pour un domaine spécifique. Mais simplifier ne peut se faire sans une perte d'informations. Et le choix des informations mises de côté ne peut se faire sans une part de subjectivité (et donc un affaiblissement de l'objectivité pourtant si importante aux yeux des auteurs de méta-analyses).

Des critiques ont également été émises sur la façon de calculer cet indice d'amélioration, car il ne prenait pas en compte la taille des

^a C'est d'ailleurs la raison qui m'a poussée à rédiger un document expliquant les méthodes de calculs utilisées par les statisticiens dans des cas simples (Nathalie ROQUES, 2021).

échantillons des études. Une étude menée sur un échantillon de 50 élèves avait donc, jusqu'en janvier 2020, le même poids qu'une étude menée sur 2000 élèves. Le WWC a modifié ce calcul dans la version 4.1 du *Manuel des Procédures WWC* qui cette fois tient compte de la taille des échantillons, comme nous l'avons vu dans la première partie. A ce jour, aucun examen utilisant ces nouvelles procédures n'a été publié.

J'ai volontairement sélectionné les étapes finales des méta-analyses dans ce livre, en privilégiant donc la façon dont le WWC sélectionne les études et calcule ses résultats. Ce faisant, d'autres questions ont été laissées de côté qui ont pourtant tout autant d'importance et qui concernent les méta-analyses conduites par le WWC ou d'autres. Elles font elles aussi l'objet de discussions parmi les scientifiques dont quelques éléments sont repris ci-dessous.

→ **Mise en œuvre de l'intervention**

L'absence de l'évaluation de la qualité de la mise en œuvre de l'intervention a fait l'objet de plusieurs critiques. Des auteurs, comme Timothy W. WOOD (2017) ont ainsi fait remarquer que des résultats qui paraissent décevants pouvaient s'expliquer par une mauvaise application de l'intervention (les personnes qui font ce genre de remarque sont en général les promoteurs de l'intervention « mal appliquée »). On peut rajouter ici que le WWC ne fixe aucune durée minimale concernant la mise en œuvre de l'intervention contrairement à d'autres auteurs de méta-analyses, comme Robert SLAVIN (2019) qui a lui exclu les études où l'intervention était appliquée moins de 12 semaines.

→ **Le groupe comparaison**

Dans les études par comparaison de groupes, la plupart du temps peu d'informations sont données sur le groupe comparaison. Et comme le fait remarquer James HIEBERT (2007, p.396), « *si la méthode d'enseignement identifiée comme la moins bonne^a est de très mauvaise qualité, ou insuffisamment décrite, alors les résultats qui montrent que l'autre méthode est meilleure sont de faible importance* ». De plus, si on peut penser que les études utilisées dans une même méta-analyse appliquent toute à peu près de la même façon l'intervention dans les groupes traités, il n'en est pas de même pour l'enseignement dispensé dans les groupes de comparaison qui peut varier énormément d'une étude à l'autre. Et finalement, on risque de se retrouver avec un ensemble d'analyses disparates qu'il devient difficile d'intégrer dans une même méta-analyse.

→ **La validité externe des études**

Pour être conformes aux normes WWC, une étude doit satisfaire un certain nombre de critères qui permettent de vérifier la qualité du lien de cause à effet mis en avant par l'étude. L'accent est donc clairement mis par le WWC sur ce que l'on nomme la validité interne des études examinées. Ce faisant, la question de la généralisation des résultats, en d'autres termes la validité externe de l'étude, est laissée de côté. Elle est pourtant au cœur des attentes des professionnels de l'enseignement qui ont besoin de savoir si les conclusions peuvent être étendues à leurs élèves, c'est-à-dire à une population bien souvent différente de l'échantillon analysé par les chercheurs. Un enseignant qui travaille avec des élèves dans un contexte particulier (c'est-à-dire dans un pays déterminé, en milieu urbain ou rural, dans un environnement socioéconomique et culturel spécifique par exemple) peut-il utiliser telle ou telle intervention qui a montré des résultats

^a Et c'est presque toujours celle mise en œuvre dans le groupe comparaison.

solides dans un autre contexte ? Répondre à cette question exige une réplication des études dans des contextes variés (ce qui augmente le niveau de preuve telle qu'évalué par le WWC)^a. Cela peut être problématique pour des enseignants français qui ne travaillent ni dans le même cadre structurel ni avec les mêmes outils que les enseignants américains à qui le WWC s'adresse en priorité puisque c'est le cœur même de sa mission.

→ **Les biais de publication et de mesure**

Les méta-analyses ne considèrent que des résultats disponibles, et les manuscrits (publiés ou non) concernent la plupart du temps des résultats positifs. En effet, une absence de résultats ou des résultats négatifs sont souvent perçus comme un échec ou inintéressants par les chercheurs, et sont de ce fait moins souvent publiés. Ainsi, sur 4 346 tailles d'effet calculées par le WWC, 24,1 % sont négatives et 4% sont statistiquement et significativement négatives ; elles sont 74,4 % à être positive et 31,2 % à être statistiquement et significativement positives^b. C'est pour éviter ce biais (ou erreur systématique) de publication que les méta-analyses étendent leur recherche à la « littérature grise », moins susceptibles de mettre de côté des résultats

^a Nous ne sommes pas rentrés dans les détails du calcul de la taille globale de l'effet par le WWC qui a choisit d'utiliser le modèle des effets fixes en janvier 2020. Un autre modèle, le modèle des effets aléatoires, aurait permis d'élargir les conclusions des méta-analyses mais n'est pas actuellement mis en œuvre. Plus de détails dans Nathalie ROQUES (2021)

^b 42% des tailles d'effet positives sont statistiquement et significativement positives mais seulement 16,8% des tailles d'effet négatives sont statistiquement et significativement négatives (calculs effectués sur les données publiées en mai 2020 à partir du fichier csv).

négatifs^a. Mais dans ce cas, la qualité des publications (qui ne sont donc plus soumises à un comité de lecture) peut être interrogée.

Dans un registre différent, la qualité des tests utilisés pour mesurer les compétences des élèves doit être questionnée. Ainsi, Robert SLAVIN (2008) fait-il remarquer que les résultats issus de tests élaborés par les auteurs même d'une étude sont plus susceptibles de donner des résultats positifs que les résultats mesurés par des tests extérieurs à l'intervention. En effet, les tests « maison » courent le risque d'évaluer la qualité de la mise en œuvre de l'intervention plutôt que l'évolution des acquisitions des élèves soumis à l'intervention. C'est pour cette raison que ce chercheur a tenu compte de cet élément et écarté certaines études de sa méta-analyse.

D'autres éléments, comme l'effet Hawthorne^b ou l'impact des tests effectués sur les élèves avant l'intervention (ce sont les prétests, indispensables pour tenir compte du niveau des élèves avant l'intervention), sont susceptibles d'avoir un effet sur les résultats calculés, le plus souvent en « gonflant » ces derniers et donc en surestimant les effets des interventions.

Finalement, de nombreux pièges méthodologiques guettent les chercheurs qui parfois, en toute bonne foi et à leur insu, laissent leur opinion s'inviter dans les procédures mises en œuvre dans leurs méta-analyses. On remarquera tout de même qu'ils ne sont pas les seuls à tomber dedans. Et quand bien même le chercheur se montrerait d'une neutralité exemplaire, c'est alors son manque d'implication et

^a Depuis janvier 2020 le WWC prend en compte les thèses de master des sciences de l'éducation qui étaient jusqu'alors laissées de côté.

^b Qui est aux sciences de l'éducation ce qu'est l'effet placebo est à la médecine.

probablement de connaissances pour l'intervention examinée qui risquerait de diminuer son expertise.

Les méta-analyses dans la recherche sur l'enseignement

La grande question de la réussite scolaire des élèves est le sujet principal d'un nombre colossal de recherches qui ne s'appuient pas toutes, loin s'en faut, sur des essais contrôlés randomisés.

Si certains laissent la part belle à ce type d'étude, et c'est bien sûr le cas du What Works Clearinghouse, d'autres préconisent de leurs associer des études dites qualitatives qui ont comme objectif d'expliquer pourquoi telle intervention a tel effet sur tel élève. C'est le point de vue de James HIEBERT (2007) qui expose ses arguments dans un article dont je vous recommande la lecture. Dans la nouvelle version 4.1 des *Manuels des procédures WWC*, un calcul de taille d'effet pour les études de cas est proposé, et la place de ces études en est ainsi renforcée.

Toujours dans le domaine des études scientifiques quantitatives, l'expérience accumulée durant ces 30 dernières années de recherche conduit certains chercheurs à accorder plus de crédit à une étude menée sur un échantillon important qu'à une méta-analyses conduite sur plusieurs études de tailles modestes et, par nature, hétérogènes. C'est véritablement au plus haut niveau de décision que toutes les questions relatives à l'élaboration des recherches scientifiques quantitatives devraient être posées pour attribuer de manière raisonnée et efficace des moyens techniques, financiers et humains^a en fonction des objectifs fixés.

^a Les besoins pour mener ce type d'études sont importants et constituent un frein non négligeable à leur mise en œuvre.

D'autres encore laissent tout simplement de côté les études quantitatives et par là-même les méta-analyses qui en sont le fruit. Les questions de recherche et les méthodes mises en œuvre sont alors totalement différentes. C'est le cas en France où certains didacticiens considèrent que les études qui comparent des groupes entre eux ne peuvent pas rendre compte des mécanismes complexes dont la connaissance est indispensable pour comprendre comment un concept exposé par un enseignant devient une compétence chez un élève. C'est ainsi que Ghilsaine GEUDET et Aline ROBERT (2018, p.379) assurent que « *la validité des recherches et résultats en didactique des mathématiques qui partent de question de terrain ne peut relever d'une évaluation directe des effets sur les apprentissages des élèves, encore moins d'une analyse statistique à grande échelle* ». On pourrait aussi associer les neuro sciences à ce pan de la recherche qui n'a pas comme premier objectif de recommander une pratique ou une méthode d'enseignement, mais plutôt de comprendre finement les processus mis en œuvre quand un élève comprend, ou apprend. On regrettera ici que ni la presse, ni l'institution scolaire ne fassent écho des résultats d'une recherche française confidentielle qui n'a pas su pour l'instant médiatiser ses conclusions.

L'objectif de ce texte était de permettre aux enseignants, aux parents, et plus largement à toute personne jouant un rôle dans la construction des apprentissages chez nos jeunes, de comprendre les résultats publiés par le What Works Clearinghouse, résultats peu connus en France^a. Il me semble que ces informations pourraient être prises en considération lors des débats qui agitent la communauté scientifique

^a Je n'ai pour ma part trouvé aucune publication ni article en langue française à ce sujet.

appelée à la rescousse de nos élèves en difficulté. Et c'est tout particulièrement vrai quand ces débats portent sur des méthodes d'enseignement, ce qui est souvent le cas. Le dernier événement en date de ce type a concerné la méthode *Singapour* mise en avant dans le rapport *21 mesures pour l'enseignement des mathématiques* (Cédric VILLANI et Charles TOROSSIAN, 2018). Cette méthode n'a pour l'instant pas fait l'objet d'une étude scientifique conforme aux normes WWC qui, interrogé sur son efficacité, n'a pas pu se prononcer.

Glossaire

Attrition	Perte de sujets au cours de l'expérience qui entraîne une perte de données.
Domaine	Ensemble de compétences disciplinaires. En mathématiques le WWC en a défini trois : <i>mathématiques générales, algèbre et géométrie.</i>
Donnée	Nombre publié par une étude primaire qui sera utilisée dans une méta-analyse (par exemple la moyenne des scores des élèves obtenus lors d'un test).
Écart-type	Nombre qui évalue la dispersion des scores autour de leur moyenne, dans la même unité que les scores.
Étude primaire	Étude expérimentale prospective dont proviennent les données utilisées par les méta-analyses.
Étude secondaire	Étude rétrospective utilisant les données provenant d'études primaires (voir méta-analyse)
Intervention	Cela peut être un programme scolaire ou un ensemble de pratiques qualifié de méthode pédagogique ou un produit commercial (comme une application informatique ou un manuel), dont l'objectif est d'améliorer les compétences des élèves. L'intervention <i>Odyssey Math</i> a été utilisée dans ce texte comme exemple.
Méta-analyse	Synthèse systématique d'études quantitatives indépendantes répondant à une même question de recherche qui se traduit notamment par le calcul d'indicateurs mesurant l'effet d'une intervention. Une méta-analyse est une étude secondaire

	rétrospective ; les études incluses dans la méta-analyse sont des études primaires.
Quasi-expérimentale (étude)	Étude dans laquelle la constitution des groupes intervention et comparaison ne repose pas sur la randomisation mais sur d'autres méthodes de manière à ce que les profils des deux groupes soient similaires (comme l'appariement ou la régression par discontinuité).
Recommandation	Ensemble d'éléments concrets et pratiques préconisés concernant des aspects pédagogiques, des concepts mathématiques ou des compétences et des comportements à favoriser chez les élèves.
Résultat	Nombre calculé par les auteurs des méta-analyses à partir des données de l'étude (par exemple une taille d'effet).
Taille d'effet	Différence standardisée des moyennes : écart entre les moyennes divisé par l'écart-type des scores.
Validité interne	Qualité d'une étude associée à la fiabilité des conclusions internes (lien de cause à effet)
Validité externe	Qualité d'une étude associée à la possibilité de généraliser les résultats observés au-delà de la situation expérimentale elle-même.

Références

Toutes les publications du What Works Clearinghouse sont téléchargeables librement sur leur site internet <https://ies.ed.gov/ncee/wwc/>. Des traductions partielles (*Manuel des procédures WWC* et *Manuel des normes WWC*) sont disponibles sur mathadoc.fr

Les Manuels de procédures et les Manuels des normes WWC

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2017, October). *What Works Clearinghouse : Procedures Handbook (Version 4.0)*.

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2017, October). *What Works Clearinghouse : Standards Handbook (Version 4.0)*.

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2019, October). *What Works Clearinghouse : Procedures Handbook (Version 4.1)*.

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2019, October). *What Works Clearinghouse : Standards Handbook (Version 4.1)*.

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2014, March). *What Works Clearinghouse : Procedures and Standards Handbook (Version 3.0)*.

Les Guides des pratiques (Practice Guides)

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools* (NCEE 2009-4060). Washington, DC : National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Traduction : Aider les élèves en difficulté en mathématiques, le soutien scolaire à l'école primaire et au collège

Star, J. R., Caronongan, P., Foegen, A., Furgeson, J., Keating, B., Larson, M. R., Lyskawa, J., McCallum, W. G., Porath, J., & Zbiek, R. M. (2015). *Teaching strategies for improving algebra knowledge in middle and high school students* (NCEE 2014-4333). Washington, DC : National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.

Traduction : Enseigner des stratégies pour favoriser les connaissances en algèbre au collège et au lycée

Siegler, R., Carpenter, T., Fennell, F., Geary, D., Lewis, J., Okamoto, Y., Thompson, L., & Wray, J. (2010). *Developing effective fractions instruction for kindergarten through 8th grade : A practice guide* (NCEE #2010-4039). Washington, DC : National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education

Traduction : Développer un enseignement efficace des fractions de la maternelle à la 4^{ème}

Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., Koedinger, K. R., & Ogbuehi, P. (2012). *Improving mathematical problem solving in grades 4 through 8 : A practice guide* (NCEE 2012-4055). Washington, DC : National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Traduction : Améliorer la résolution des problèmes du CM1 à la 4^{ème}
Halpern, D., Aronson, J., Reimer, N., Simpkins, S., Star, J., and Wentzel, K. (2007). *Encouraging Girls in Math and Science* (NCER 2007-2003). Washington, DC : National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Traduction : Encourager les filles en maths et en sciences

Frye, D., Barody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide* (NCEE 2014-4005). Washington, DC : National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.

Traduction : Enseigner les maths aux jeunes enfants

Les *Rapports d'intervention (Intervention Reports)*

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2017, December). Primary Mathematics intervention report : Accelerated Math®

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2016, June). Secondary Mathematics intervention report : Cognitive Tutor®

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2017, February). Primary Mathematics intervention report : Connected Mathematics Project.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2016, May). Secondary Mathematics intervention report : Saxon Math.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2017, May). Primary Mathematics intervention report : Saxon Math.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2018, January). Charter Schools intervention report : Knowledge is Power Program (KIPP)

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2017, January). Primary Mathematics intervention report : Odyssey Math.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse (2015, December). Primary Mathematics intervention report : Singapore Math.

Les études conformes aux normes WWC pour l'examen de l'intervention *Odyssey Math*

DiLeo, J. (2007). A study of a specific language arts and mathematics software program : Is there a correlation between usage levels and achievement? (Unpublished doctoral dissertation). Indiana University of Pennsylvania, Indiana.

Wijekumar, K., Hitchcock, J., Turner, H., Lei, P., & Peck, K. (2009). A multisite cluster randomized trial of the effects of Compass Learning Odyssey Math on the math achievement of selected grade 4 students in the mid-Atlantic region. Final report (NCEE 2009-4068). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://eric.ed.gov/?id=ED507314>

Cornelius, A. S. (2013). The effects of response to intervention on the mathematics achievement of seventh and eighth grade students (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI 1422044724)

Les autres publications

BORENSTEIN Michael, HEDGES Larry V ;, HIGGINS Julian P.T. et ROTHSTEIN Hannah R, *Introduction to meta-analysis*, Wiley, 2009.

GEUDET, Ghislaine et ROBERT, Aline *La didactique des mathématiques en France : panorama du champ de recherche* dans Enseigner les mathématiques. Didactique et enjeux de l'apprentissage, Belin, 2018.

HIEBERT James et GROUWS Douglas, *The effect of classroom mathematics teaching on student's learning*, Second Handbook of Research on Mathematics Teaching and Learning, Chapter 9, p.371 – 404, 2007, traduction partielle sur mathadoc.fr en vente sur bookelis.fr

REY, Olivier (2006), *Qu'est-ce qu'une « bonne » recherche en éducation*, Dossiers de veille de l'Ifé, n°18, mai 2006

ROQUES, Nathalie (2021) *Mesurer l'effet d'un traitement. Méta-analyses et enseignement*. En cours de publication

SLAVIN Robert E., *Effective Programs in Middle and High School Mathematics : A Best-Evidence Synthesis*, juin 2019, DOI : 10.3102/0034654308330968

VILLANI, Cédric et TOROSSIAN, Charles, *21 mesures pour l'enseignement des mathématiques*, février 2018

WOOD, Timothy W. , NIFDI, *Does the What Works Clearinghouse Really Work ? Investigations into Issues of Policy, Practice, and Transparency*, janvier 2017

Du même auteur

Inégalités sociales et mathématiques dans l'OCDE. Volume 1 et 2 (2019)

Une année avec les 3èmeB (2019)

Comment la manière d'enseigner les mathématiques influence les apprentissages de James HIEBERT et Douglas GROUWS (2007, traduit en 2019)