Annexe technique

Instruments (questionnaires et tests)

Le Secrétariat de l'OCDE a mis au point cinq instruments différents, à savoir : trois questionnaires pré- et post-expérience (pour les élèves, enseignants et chefs d'établissement) et deux tests de performance à l'intention des élèves en mathématiques et sciences ainsi qu'en arts visuels et musique. Un sixième outil a également été utilisé pour l'évaluation de la créativité des élèves, il s'agit du test EPoC (évaluation du potentiel créatif) spécifique par domaine et élaboré par Lubart, Besançon et Barbot (Université Paris-Descartes, France).

Questionnaires

Les questionnaires « Élève » ont été soumis à l'ensemble des élèves, qu'ils appartiennent au groupe expérimental ou au groupe de contrôle. La longueur et la complexité de ces questionnaires variaient en fonction du niveau d'enseignement. Ainsi, dans l'enseignement primaire, le questionnaire soumis en amont de l'expérience comportait 75 items (informations contextuelles, sentiments, attitudes, activités, capsules d'ancrage) tandis que celui soumis après l'expérience en comportait 53 (plusieurs questions contextuelles n'étant pas répétées). Dans l'enseignement secondaire, le questionnaire pré-expérience comptait 85 items (sur le même modèle que celui proposé en primaire) tandis que le questionnaire post-expérience n'en comportait que 58.

Le questionnaire « Enseignant » pré-expérience était composé de 105 items portant sur des informations relatives à la classe cible des enseignants, leurs pratiques pédagogiques, leurs points de vue quant au rôle et à l'évaluation de la créativité et de l'esprit critique, les capsules et des informations contextuelles. Comme pour les élèves, le questionnaire post-expérience était plus bref et ne comptait que 87 items. Toutefois, celui soumis aux enseignants du groupe expérimental comportait également une batterie de 37 items supplémentaires, afin de rendre compte des sentiments de ces enseignants à l'égard de l'expérience et de leur perception de ses impacts. Les questionnaires « Enseignant » ont été soumis à l'ensemble des enseignants, qu'ils appartiennent au groupe expérimental ou au groupe de contrôle.

Les questionnaires « Établissement » ont été administrés avant et après l'expérience, la majorité d'entre eux ayant été remplis par les chefs d'établissement mais également parfois par les coordinateurs locaux du projet. Le questionnaire pré-expérience était composé de 39 items sur les caractéristiques et les ressources des établissements ainsi que sur leur participation à de précédents projets d'innovation. Le questionnaire post-expérience comptait quant à lui 19 items présentant le retour d'expérience relatif à la participation de l'établissement au projet.

Tests de performance en sciences, technologie, ingénierie et mathématiques (STIM)

Les tests de performance en sciences et mathématiques ont été conçus par le Secrétariat de l'OCDE grâce à des items tirés des deux enquêtes à grande échelle suivantes : l'étude TIMSS (*Trends in International Mathematics and Science Study*) pour les élèves du primaire et l'évaluation PISA (Programme international pour le suivi des acquis des élèves) pour les élèves du secondaire. Ces

tests comprenaient des items de sciences et de mathématiques à réponse ouverte et fermée, des questions intégrées sur l'intérêt des élèves à l'égard de ces deux matières ainsi que des questions sur les pratiques utilisées par les enseignants dans leurs cours de sciences et de mathématiques.

Pour chaque niveau d'enseignement, deux livrets de test équivalents – livret de test A et livret de test B – ont été conçus et soumis tour à tour aux élèves, permettant ainsi de comparer les résultats pré- et post-expérience. Dans l'enseignement primaire, le livret de test A comportait 45 items et le livret de test B en contenait 46. Dans l'enseignement secondaire, 48 items étaient présentés dans le livret de test A contre 50 dans le livret de test B. Tous les livrets de test contenaient un ensemble d'items équilibré abordant tant le contenu (formes géométriques, affichage de données, sciences de la Terre, etc.) que les domaines cognitifs en mathématiques et en sciences (connaître, appliquer et raisonner, par exemple). Ces livrets incluaient également quelques items examinant les attitudes et pratiques des élèves. Des directives spécifiques ont été fournies aux coordinateurs locaux du projet concernant l'administration en classe des tests de STIM.

Tests de performance en arts visuels et musique (AVM)

Les tests de performance en arts visuels et musique ont été entièrement élaborés par le Secrétariat de l'OCDE. Ils contenaient des items d'arts visuels et de musique à réponse fermée et reposaient sur l'écoute d'airs de musique et sur l'observation d'images de peintures et de sculptures. En musique, les élèves étaient évalués sur leur vocabulaire, leurs observations (auditives et visuelles), leur compréhension des effets expressifs recherchés ainsi que sur leurs connaissances théoriques du domaine. En raison du contexte multiculturel, le test était davantage axé sur la capacité à écouter et entendre ainsi qu'à voir et observer que sur la culture relative aux disciplines artistiques. À l'instar du test de STIM, certains items intégrés évaluaient l'intérêt des élèves pour ces deux matières et recueillaient des informations sur les pratiques pédagogiques utilisées dans leurs cours d'arts visuels et de musique.

Dans l'enseignement primaire, les livrets de test comportaient 78 items, tandis qu'ils en contenaient 119 dans l'enseignement secondaire. Deux livrets de test différents mais équivalents (A et B) ont été conçus tant pour les élèves de l'enseignement primaire que secondaire, afin de permettre une comparaison des résultats pré- et post-expérience.

Des directives spécifiques ont été communiquées aux coordinateurs locaux du projet concernant l'administration en classe des tests d'AVM. La majorité des questions de musique invitait les élèves à répondre au test tout en écoutant de courts airs de musique (provenant de diverses aires culturelles). Avant l'administration du test, les enseignants devaient brièvement expliquer la tâche à effectuer. De la même manière, une partie des questions d'arts visuels invitait les élèves à observer certaines œuvres d'art (peintures, dessins, sculptures).

Dans le cadre des tests d'AVM, aucune tâche n'exigeait que les élèves disposent de bonnes connaissances ou même de quelques notions sur la musique ou les d'œuvres d'art utilisées lors des tests. Les élèves étaient interrogés sur leur degré de familiarité avec la musique écoutée (qui, conformément à ce qui avait été souhaité, n'a pas constitué un facteur déterminant pour l'exactitude des réponses).

Sélection des items dans les tests d'AVM

La phase pilote du projet s'est déroulée sur deux années scolaires complètes tant pour l'hémisphère sud que l'hémisphère nord. Entre la première et la seconde année, l'équipe du projet OCDE-CERI a réalisé plusieurs tests de validation pour s'assurer du bon fonctionnement des instruments.

Dans un premier temps, les items qui présentaient un taux de réussite trop faible (généralement inférieur à 40 % pour les questions dichotomiques) ont été exclus lors du second cycle. Puis, les items dont le profil de réponse laissait entendre qu'ils pouvaient faire l'objet d'une mauvaise interprétation ou de problèmes de traduction ont également été supprimés des instruments. Afin d'identifier ces cas, les écarts de variance entre les pays ont été étudiés. Enfin, des diagnostics fondés sur la théorie de réponse à l'item (TRI), comme le fonctionnement différentiel des items, ont permis d'en signaler d'autres qui ont également été exclus de l'analyse finale des tests (ces derniers n'étant donc composés que d'items ayant donné les résultats escomptés).

Tests EPoC d'évaluation de la créativité

Les tests EPoC ont été conçus par Lubart, Besançon et Barbot afin de mesurer le potentiel créatif des enfants et adolescents dans différents domaines de la pensée et de la production créatives : expression graphique-artistique, expression verbale et littéraire, résolution de problèmes de société, composition musicale et productions scientifiques et mathématiques. Ces tests invitaient les individus à produire une création (un dessin, un récit, une solution à un problème mathématique, une composition, etc.) qui était ensuite évaluée de manière standardisée par une personne en charge de l'évaluation ou au moyen d'un outil informatique. Il existait dans chaque domaine d'évaluation deux types de tâches mettant en œuvre soit la pensée divergente-exploratoire soit la pensée convergente-intégrative (synthèse créative). La mesure finale du potentiel créatif englobait ces deux aspects de la créativité, même s'ils pouvaient être évalués séparément. Deux livrets de test équivalents ont été conçus pour chaque domaine – qualifiés de livret de test A et livret de test B – permettant ainsi de mieux comparer les résultats pré- et post-expérience.

Élaboration des scores aux tests de STIM et d'AVM

Il existe plusieurs méthodes pour élaborer les différents scores et indices, allant de techniques simples comme les sommes et les moyennes (pondérées), à des techniques plus complexes telles que les modèles inspirés de la théorie de réponse à l'item (TRI). La présente section fournit des précisions sur le calcul des scores aux tests de performance ainsi que des explications concernant le fait d'avoir privilégié les moyennes pondérées par pays aux dépens d'autres méthodes possibles.

Sommes et scores pondérés

L'examen a d'abord porté sur les « scores bruts » qui correspondent au ratio simple entre la somme des bonnes réponses apportées par chaque élève et le nombre total de questions. Toutefois, ces scores ne font aucune distinction entre les différents items et leur attribuent à tous la même importance.

Les scores pondérés ont été le second type de scores à être calculé. Ces scores ont attribué un coefficient de pondération à chaque réponse en fonction du taux de bonnes réponses observé dans l'enquête internationale dont l'item provenait (ou dans les données du projet OCDE-CERI s'agissant des tests d'AVM spécialement conçus pour l'occasion).

Par exemple, l'étude TIMMS 2011 a fait état d'un taux de bonnes réponses à l'échelle internationale estimé à 90 % pour l'item suivant : « Quelle force est à l'origine du mouvement d'un bateau ? ». Dans la mesure où cette question a été considérée comme relativement simple dans l'ensemble des pays, un coefficient de pondération de 0.10 lui a été attribué. Répondre correctement à cette question n'aurait donc pas de forte incidence sur le score final au test de STIM.

Des différences importantes ont néanmoins été observées entre les pays au niveau des scores obtenus à des items spécifiques. L'analyse des résultats de l'étude pilote ayant été effectuée par pays, les coefficients de pondération propres à chaque pays ont été utilisés afin de rendre compte de ces différences. Cela signifie que pour une question donnée les coefficients de pondération ont été calculés pour chaque pays (et pouvaient donc varier d'un pays à l'autre).

Par exemple, la proportion d'élèves à avoir bien répondu à l'item précédemment cité s'élevait à 79 % en Thaïlande, contre 94 % aux Pays-Bas. On peut ainsi partir du principe que cet item était plus difficile pour les élèves thaïlandais que pour les élèves néerlandais, et c'est pourquoi lors du calcul des scores pondérés par pays un coefficient de pondération plus élevé a été attribué à cet item dans l'équipe thaïlandaise (0.21) par rapport à l'équipe néerlandaise (0.06).

Étant donné que le Brésil, la France et l'Inde n'ont pas pris part à l'étude TIMSS 2011 (enseignement primaire) et que l'Inde n'a pas participé à l'évaluation PISA 2012 (enseignement secondaire), les coefficients de pondération moyens à l'échelle internationale ont été utilisés pour calculer les scores de ces pays. La même procédure a été adoptée lorsqu'un pays ne disposait pas de données pour une question spécifique dans l'une ou l'autre desdites études.

D'autres modifications ont été apportées dans le calcul des scores au test de STIM afin de corriger les problèmes mineurs survenus pour deux équipes nationales dans le cadre de l'administration des tests : les équipes indienne et thaïlandaise.

Équivalence des livrets de test

L'équivalence des livrets de test a été obtenue grâce à la conception des instruments, à leur mise à l'essai initiale et aux ajustements mineurs apportés aux instruments post-expérience après leur administration, de façon à corriger les légers déséquilibres apparus entre les domaines et sous-domaines.

Pour tenir compte de ces différences, le ratio entre les scores pondérés et le nombre de questions a fait l'objet d'une comparaison entre les livrets de test A et B. Afin de garantir la comparabilité des données, les scores des élèves ont été ajustés à chaque fois que le ratio d'un sous-thème (des questions ouvertes en sciences par exemple) était supérieur dans le livret de test A que dans le livret de test B (en d'autres termes, la question moyenne était plus difficile dans le livret de test A), ou inversement.

Prenons à titre d'exemple deux batteries d'items à réponse ouverte en mathématiques dans un pays donné : l'une était intégrée au livret de test A et l'autre au livret de test B. Le livret de test A contenait cinq items qui présentaient un coefficient de pondération total estimé à 2.20 (en moyenne, un coefficient de pondération de 0.44 ayant été attribué à chaque item). Le livret de test B contenait quant à lui sept items, pour un coefficient de pondération total estimé à 2.52 (en moyenne, un coefficient de pondération de 0.36 ayant été attribué à chaque item). Dans cet exemple, la batterie de questions dans le livret de test A était, en moyenne, plus difficile que celle dans le livret de test B. Pour chaque élève, le score obtenu au livret de test A a donc été ajusté à l'aide d'un coefficient de difficulté relative. Si le score résultant était supérieur au score total du livret de test B, il était alors établi à ce même niveau.

Théoriquement, dans cet exemple, lorsque le livret de test A est plus difficile que le livret de test B, l'ajustement apporté au livret de test A se présente comme suit :

$$Score\ final\ A = \mathrm{Min}\left(Score\ initial\ A \times \frac{\sum coef.de\ pond\ A \times \sum items\ B}{\sum coef.de\ pond.B \times \sum items\ A}, \sum coef.de\ pond.A\right)$$

$$\Leftrightarrow$$

$$Score\ final\ A = \mathrm{Min}\left(Score\ initial\ A \times \frac{2.20 \times 7}{2.52 \times 5}, 2.20\right)$$

$$\Leftrightarrow$$

$$Score\ final\ A = \mathrm{Min}\left(Score\ initial\ A \times 1.22, 2.20\right)$$

Cette approche a été privilégiée à celle de la standardisation, dans la mesure où cette dernière aurait affecté les valeurs aux deux extrémités de l'échelle des scores, compliquant ainsi l'interprétation. Toutefois, il convient d'admettre qu'avec l'approche adoptée certaines informations ont été perdues pour les élèves qui avaient obtenu des scores plus élevés au livret de test le plus difficile.

Scores obtenus à l'aide de la théorie de réponse à l'item

Une autre méthode pouvant servir à calculer les scores aux tests de performance repose sur les modèles inspirés de la théorie de réponse à l'item, qui modélisent la réponse de chaque participant d'un niveau de compétences donné à chaque item du test. Ces modèles comprennent également des paramètres relatifs à la difficulté des items, à leur pouvoir discriminant et à la pseudo-chance.

Le paramètre de difficulté (*b*) décrit le degré de difficulté de l'item. Plus ce paramètre est élevé, plus grande est la probabilité que l'élève ne réponde pas correctement à l'item. Le paramètre de discrimination (*a*) décrit le pouvoir discriminant de l'item, c'est-à-dire sa capacité à différencier les répondants dont les compétences sont élevées de ceux dont les compétences sont faibles. Le paramètre de pseudo-chance (*c*) décrit la probabilité de répondre correctement à l'item par pur hasard (c'est-à-dire 1/k, où k correspond au nombre de réponses possibles à l'item).

Dans un tel modèle à trois paramètres, la probabilité de répondre correctement à un item dichotomique *i* pour un répondant dont la capacité correspond à θ est la suivante :

$$p_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

Dans le cadre de cette étude pilote, des modèles à un, deux et trois paramètres ont été adaptés aux données séparément par niveau d'enseignement. Toutefois, la taille des échantillons n'a pas permis de convergence lorsque plus d'un paramètre était utilisé. Avec des échantillons bien plus grands que ceux présents dans cette étude pilote, l'évaluation PISA a mis en œuvre des modèles inspirés de la théorie de réponse à l'item à deux paramètres depuis 2015 uniquement.

S'agissant des tests de STIM et d'AVM, des modèles à un paramètre ont servi pour l'élaboration des scores qui réunissaient les informations fournies par l'ensemble des pays (score transnational) ou qui étaient calculés pour chaque pays (score par pays).

Engagement des élèves

Les tests de STIM et d'AVM administrés dans le cadre de ce projet étant des évaluations sans enjeu pour les élèves, des contre-mesures ont été prises afin de minimiser les conséquences d'un éventuel faible engagement de la part de certains élèves. Tous les scores ont notamment été calculés en tenant compte d'un seuil fixé à 30 % s'agissant du nombre de réponses manquantes de la part d'un élève (et en excluant les scores inférieurs à ce seuil). En d'autres termes, dès lors que ce seuil était appliqué, on attribuait aux élèves un score valide uniquement lorsqu'ils avaient répondu à au moins 70 % des items.

La fixation de ce seuil à hauteur de 30 % s'est appuyée, dans les faits, sur plusieurs considérations d'ordre pratique telles que le nombre de résultats qui auraient dû être exclus ou l'éventuelle gamme de seuils disponibles pour l'ensemble des différents groupes et sous-groupes d'items (par exemple, pour des batteries de six items, un seuil supérieur à 33 % aurait supposé d'accepter également les élèves n'affichant que trois réponses valides, soit un seuil jugé inacceptable). Parmi l'ensemble des élèves ayant répondu aux tests pré- et post-expérience, l'application de ce seuil s'est traduite par l'exclusion de 12 % des résultats au test de STIM au niveau de l'enseignement primaire et de 9 % des résultats au niveau de l'enseignement secondaire. Pour le test d'AVM, cela a conduit à l'exclusion de 30 % des résultats au niveau du primaire et de 13 % des résultats au niveau du secondaire (le premier chiffre laissant entendre une charge de travail excessive demandée aux élèves du primaire). Si ces pertes de données étaient escomptées, elles ont toutefois été considérées comme raisonnables pour l'obtention de scores plus fiables, tant à l'échelle individuelle qu'au niveau des équipes nationales.

De plus, seuls les élèves ayant rempli les questionnaires pré- et post-expérience ou ayant passé les tests de performance ou tests de créativité pré- et post-expérience ont été pris en compte dans l'analyse relative à l'instrument en question. Ceux qui ne se trouvaient pas dans cette situation se voyaient tout de même attribuer un score lorsque cela était possible, mais étaient exclus de toutes les analyses portant sur les comparaisons avant et après expérience.

Sélection des scores

Si les scores obtenus à l'aide de la théorie de réponse à l'item sont plus appropriés pour le calcul du score final à un test, ils sont également plus complexes et, surtout, plus difficiles à interpréter (car ils ne conservent pas la plage de valeurs du test initial). Dans ce projet pilote, la corrélation entre les scores obtenus à l'aide de la théorie de réponse à l'item et les scores plus simples était très élevée (toujours supérieure à 0.80).

Par exemple, le Tableau 1 présente les corrélations observées entre les différents types de scores au test de STIM sur la base des réponses apportées lors du test pré-expérience. Les corrélations étant toujours fortes, cela signifie que les modèles inspirés de la théorie de réponse à l'item produisent des résultats similaires à ceux obtenus à l'aide de méthodes plus simples. Par conséquent, le score pondéré par pays, dont le calcul est plus simple, a été choisi comme la méthode définitive pour le calcul des scores aux tests de STIM et d'AVM.

Tableau 1. Tableau des corrélations entre les différents types de scores au test de STIM

Pré-expérience (en amont de l'expérience)

	Score brut	Score international pondéré	Score pondéré par pays	Score pondéré par pays (sans seuil)	Score transnational obtenu avec la TRI	Score transnational obtenu avec la TRI (sans seuil)	Score par pays obtenu avec la TRI
Score brut	1.00						
Score international pondéré	0.98	1.00					
Score pondéré par pays	**0.96**	**0.98**	**1.00**				
Score pondéré par pays (sans seuil)	0.96	0.97	**1.00**	1.00			
Score transnational obtenu avec la TRI	0.87	0.85	**0.84**	0.84	1.00		
Score transnational obtenu avec la TRI (sans seuil)	0.88	0.86	**0.85**	0.84	0.97	1.00	
Score par pays obtenu avec la TRI	0.84	0.82	**0.82**	0.81	0.90	0.87	1.00

Scores aux tests EPoC

La notation des tests EPoC d'évaluation de la créativité a été réalisée ou coordonnée par une équipe de chercheurs externes spécialisés qui a évalué, sur une échelle à 7 points, les performances des élèves dans chaque sous-domaine (les épreuves de pensée convergente-intégrative en expression verbale et littéraire, par exemple).

Tous les élèves n'ont pas passé le même test EPoC. En effet, les pays se concentraient sur différents domaines, en fonction de celui dans lequel se déroulait l'expérience pédagogique. Par ailleurs, ils se pouvaient parfois que les élèves n'effectuent qu'une partie du test (composé de quatre épreuves pour chaque domaine). Étant donné que les scores variaient d'un domaine à l'autre, ils ont tous été standardisés afin d'attribuer à chaque élève un unique score au test EPoC dans une fourchette commune pour tous les domaines. Dans le cadre de l'attribution des scores aux élèves, les règles suivantes ont été observées :

- Si la discipline de l'enseignant était les mathématiques, seul le score obtenu au test EPoC en mathématiques était pris en compte pour le calcul du score des élèves au test EPoC (même si les élèves avaient passé des tests EPoC supplémentaires dans d'autres domaines).
- Si la discipline de l'enseignant était les sciences ou la biologie, seul le score obtenu au test EPoC en sciences contribuait au score des élèves au test EPoC.
- Si la discipline de l'enseignant était les arts visuels, la musique ou d'autres disciplines artistiques, seul le score obtenu à l'épreuve d'expression graphique-artistique du test EPoC contribuait au score des élèves au test EPoC.
- Si la discipline de l'enseignant était la physique ou la technologie, alors les scores obtenus au test EPoC en sciences et en mathématiques contribuaient au score des élèves au test EPoC (lorsque les deux scores étaient disponibles, dans le cas inverse seul celui disponible était pris en compte).
- Enfin, dans tous les autres cas, les scores au test EPoC ont été élaborés à partir de l'ensemble des scores disponibles (mathématiques, sciences, expression graphique-artistique, résolution de problème de société et expression verbale et littéraire).

Élaboration des indices contextuels

Plusieurs indices socio-économiques ont été calculés sur un large éventail de variables. Ils décrivaient le milieu social, économique et culturel des élèves en s'appuyant principalement – mais pas systématiquement – sur les questionnaires contextuels remplis par les élèves, les enseignants et les chefs d'établissement. Dans certains pays, des données administratives sont venues compléter les informations tirées des déclarations des répondants.

Âge

Dans de rares cas, les réponses apportées au questionnaire ont été comparées et enrichies avec d'autres données disponibles provenant des dossiers scolaires, des registres administratifs et d'autres sources d'information. Malgré tous ces efforts, quelque 5 000 élèves ne disposaient toujours pas de valeur valide concernant leur âge, notamment dans les équipes américaine, espagnole, indienne et russe. En pareil cas, leur âge a fait l'objet d'une estimation qui correspondait à l'âge habituel d'un enfant scolarisé dans la même année d'études dans le pays en question.

Indice de l'environnement culturel familial

Dans le droit fil de l'item PISA relatif au patrimoine culturel disponible au domicile, un indice propre à chaque pays a été créé pour décrire l'environnement culturel familial de l'élève à partir du nombre de livres présents dans sa bibliothèque familiale. Cet indice a été élaboré par pays en raison de l'hétérogénéité des bibliothèques familiales observée entre les différents pays : disposer, par exemple, de 50 livres à la maison n'ayant pas la même valeur que l'on se trouve dans une région du monde

ou dans une autre. En vue d'établir des seuils cohérents, les données tirées de l'évaluation PISA sur la répartition du patrimoine culturel ont servi de valeur de référence, dans la mesure où elles étaient représentatives à l'échelle nationale. Tous les pays ont adopté le même seuil et le même plafond s'agissant de la variable relative à la bibliothèque familiale, à savoir de 0 à 10 livres et plus de 200 livres.

Indice du niveau de formation des parents

Cet indice repose sur le niveau de formation le plus élevé des parents, déterminé non pas en fonction du nombre d'années d'études mais du niveau de la CITE. En suivant la même logique que celle appliquée pour le calcul de l'indice de l'environnement culturel familial, des données provenant de l'évaluation PISA (et d'Eurostat) ont servi de valeurs de référence représentatives à l'échelle nationale afin de répartir les élèves dans l'une des trois catégories suivantes :

- Niveau faible, correspondant à un niveau égal ou inférieur au premier cycle de l'enseignement secondaire (ou à l'enseignement primaire pour les équipes brésilienne et thaïlandaise).
- Niveau intermédiaire, allant de l'enseignement de faible niveau à l'enseignement post-secondaire non tertiaire.
- Niveau élevé, correspondant à un niveau égal ou supérieur à la licence ou une formation tertiaire de cycle court.

Cet indice ne concernait que les élèves de l'enseignement secondaire, car le questionnaire « Élève » proposé dans l'enseignement primaire ne comportait aucune question relative au niveau de formation des parents.

Indice de statut au regard de l'immigration

Conformément à la définition PISA, on entend par élève issu de l'immigration celui dont les deux parents sont nés à l'étranger (et ce, peu importe le pays de naissance de l'élève). Dans le cadre de cette étude pilote, le pourcentage d'élèves issus de l'immigration variait considérablement d'une équipe à l'autre, et on observait dans certaines équipes une très faible présence de ces élèves, du moins si l'on se référait à la définition PISA. Toutefois, lorsqu'on utilise séparément les informations sur le pays de naissance des élèves et de chacun de leurs parents, les données font ressortir plusieurs profils d'élèves différents, et ces différences sont d'autant plus grandes que sont prises en compte les données fournies par la variable relative à la langue principale parlée en famille par les élèves.

Afin de disposer de résultats pertinents pour l'ensemble des équipes, un nouvel indice décrivant le statut des élèves au regard de l'immigration a été créé grâce à toutes les informations fournies par les variables relatives au pays de naissance des élèves et de leurs parents, et à la langue parlée en famille par les élèves. Les caractéristiques de ce nouvel indice sont illustrées dans le Tableau 2 et le Tableau 3. Ces deux tableaux permettent notamment de constater que les pays ont été répartis en deux groupes en fonction du pourcentage d'élèves qui présentent au moins certains aspects que l'on retrouve chez les personnes issues de l'immigration.

Tableau 2. Indice de statut au regard de l'immigration pour les équipes qui comprennent un grand nombre d'élèves issus de l'immigration

Langue parlée en famille / Né(s) dans le pays	Non pour l'élève et les parents	Oui pour l'élève ; non pour les deux parents	Non pour l'élève ; oui pour un des deux parents	Oui pour l'élève ; non pour un des deux parents	Non pour l'élève ; oui pour les deux parents	Oui pour l'élève et les deux parents
Langue étrangère	Élevé	Élevé	Élevé	Élevé	Modéré	Modéré
Langue secondaire utilisée dans le pays	Élevé	Élevé	Modéré	Modéré	Modéré	Modéré
Langue principale utilisée dans le pays	Modéré	Modéré	Modéré	Modéré	Autochtone	Autochtone

Remarque : seules les équipes américaine (Vista), françaises (CRI et Lamap), galloise et indienne ont été prises en compte dans ce groupe.

Tableau 3. Indice de statut au regard de l'immigration pour les équipes qui comprennent peu d'élèves issus de l'immigration

Langue parlée en famille / Né(s) dans le pays	Non pour l'élève et les parents	Oui pour l'élève ; non pour les deux parents	Non pour l'élève ; oui pour un des deux parents	Oui pour l'élève ; non pour un des deux parents	Non pour l'élève ; oui pour les deux parents	Oui pour l'élève et les deux parents
Langue étrangère	Élevé	Élevé	Élevé	Élevé	Élevé	Élevé
Langue secondaire utilisée dans le pays	Élevé	Élevé	Élevé	Élevé	Modéré	Modéré
Langue principale utilisée dans le pays	Élevé	Élevé	Modéré	Modéré	Autochtone	Autochtone

Remarque : seules les équipes américaine (Montessori), brésilienne, hongroise, néerlandaise, russe, slovaque et thaïlandaise ont été prises en compte dans ce groupe.

Indice de statut socio-économique et culturel (SSC)

L'indice de statut socio-économique et culturel (SSC) regroupe toutes les informations fournies par l'indice de l'environnement culturel familial et l'indice du niveau de formation des parents. Au sein de l'échantillon, un statut dit neutre a été attribué à la majorité des élèves des établissements d'enseignement primaire, tandis qu'ils étaient 20 % à être répartis dans les groupes de statuts dits positif et négatif. Dans l'enseignement secondaire, 85 % des élèves ont été intégrés au groupe du statut neutre, 10 % dans le groupe du statut le plus élevé et seulement 5 % dans le groupe du statut le plus faible.

Deux versions supplémentaires de cet indice ont été calculées, sur la base des réponses apportées aux questionnaires « Établissement » et « Enseignant ». Si elles n'ont fait état d'aucune variation au sein des classes (ou des établissements), elles demeurent néanmoins une précieuse source d'informations pour situer les résultats et conclusions dans leur contexte.

Élaboration des indices liés à la créativité et à l'esprit critique

Tant les questionnaires que les tests de performance comprenaient plusieurs items rendant compte des sentiments, pratiques et attitudes des répondants à l'égard de la créativité et de l'esprit critique.

La majorité des indices regroupant les données contenues dans les différentes batteries d'items a été élaborée grâce à une analyse des facteurs principaux (voir le Tableau 4). Lorsque cette analyse ne pouvait être menée, en raison notamment de résultats incomplets, d'une faible quantité de données disponibles ou du non-respect de certaines hypothèses dans la répartition des questions, d'autres méthodes de sommation plus simples ont été utilisées (voir le Tableau 5).

Des contrôles statistiques (tels que le coefficient alpha de Cronbach et le niveau de variance expliqué par le premier facteur – c'est-à-dire celui tiré de l'analyse) ont été menés pour veiller à la fiabilité des indices. Par ailleurs, le respect de l'invariance de configuration entre l'ensemble des équipes et niveaux d'enseignement avait été garanti. Comme le montre le Tableau 4, certains indices semblent ne bien fonctionner qu'avec les élèves de l'enseignement secondaire, portant ainsi à croire que l'effort cognitif demandé aux élèves était potentiellement trop important pour ceux de l'enseignement primaire.

Tableau 4. Indices conçus à l'aide de l'analyse des facteurs principaux

Indice	Exemples d'items	Échelle	Coefficient alpha de Cronbach combiné (nbre d'items composant l'indice indiqué entre parenthèses)	
			Primaire	Secondaire
Pratiques liées à la C-PC en sciences	En cours de sciences, nous sommes invités à réfléchir à la première action que nous devrions entreprendre pour résoudre un problème scientifique	De 1 (jamais) à 4 (à tous les cours)	0.732 (6)	0.756 (6)
	En cours de sciences, nous sommes invités à expliquer pourquoi certaines choses sont impossibles			
Pratiques liées à la C-PC en mathématiques	En cours de mathématiques, nous travaillons sur des problèmes qui admettent plusieurs solutions	De 1 (jamais) à 4 (à tous les cours)	0.699 (6)	0.696 (6)
	En cours de mathématiques, nous travaillons sur des problèmes pour lesquels il existe plusieurs possibilités d'arriver à la bonne solution			
Pratiques liées à la C-PC en arts visuels	En cours d'arts visuels, je dois utiliser les connaissances que j'ai acquises dans d'autres matières pour effectuer mon travail	De 1 (jamais) à 4 (à tous les cours)	0.632 (6)	0.733 (6)
	En cours d'arts visuels, nous discutons de styles artistiques dont je n'avais jamais entendu parler			

Indice	Exemples d'items	Échelle	Coefficient alpha de Cronbach combiné (nbre d'items composant l'indice indiqué entre parenthèses)	
			Primaire	Secondaire
Pratiques liées à la C-PC en musique	En cours de musique, nous écoutons et jouons des types de musique que je ne connaissais pas auparavant	De 1 (jamais) à 4 (à tous les cours)	0.683 (5)	0.752 (6)
	En cours de musique, nous composons de la musique			
Intérêt pour les sciences	J'aimerais apprendre à concevoir des expériences scientifiques	De 1 (pas du tout) à 4 (très vivement)	0.570 (4)	0.722 (4)
	J'aimerais comprendre pourquoi les scientifiques sont parfois en désaccord			
Intérêt pour les mathématiques	J'aimerais apprendre à penser comme un mathématicien	De 1 (pas du tout) à 4 (très vivement)	0.678 (4)	0.701 (4)
	J'aimerais mieux comprendre pourquoi on dit que les mathématiques sont omniprésentes			
Intérêt pour les arts visuels	J'aimerais acquérir suffisamment de connaissances pour pouvoir créer des œuvres pour mon plaisir après l'école	De 1 (pas du tout d'accord) à 4 (tout à fait d'accord)	0.576 (3)	0.847 (4)
	J'aimerais comprendre comment les arts visuels parviennent à transmettre certaines émotions			
Intérêt pour la musique	J'aimerais avoir suffisamment de connaissances en matière de musique pour pouvoir jouer avec mes amis après l'école	De 1 (pas du tout d'accord) à 4 (tout à fait d'accord)	0.636 (3)	0.796 (4)
	J'aimerais comprendre comment la musique parvient à transmettre certaines émotions			
Sentiments positifs à l'égard de l'apprentissage	J'ai le sentiment de bien travailler dans ce cours	De 1 (jamais) à 4 (toujours)	0.576 (4)	0.757 (6)
	J'ai le sentiment de faire preuve de coopération dans ce cours	De 1 (pas du tout) à 4 (énormément)		
Sentiments négatifs à l'égard de l'apprentissage	J'éprouve du stress dans ce cours	De 1 (pas du tout) à 4 (énormément)	n.a.	0.759 (4)
	Je me sens perdu dans ce cours			
Dispositions à l'égard de la C-PC[1]	J'aime comparer différents angles pour aborder des problèmes lors de mon temps libre en dehors du cadre scolaire	De 1 (pas du tout) à 4 (énormément)	0.750 (6)	0.805 (6)
	En dehors du cadre scolaire, j'essaye de comprendre pourquoi les individus affichent certaines idées			

Indice	Exemples d'items	Échelle	Coefficient alpha de Cronbach combiné (nbre d'items composant l'indice indiqué entre parenthèses)	
			Primaire	Secondaire
Méthode d'apprentissage de la C-PC	Généralement, dans mon travail scolaire, je dois explorer différents points de vue sur un problème ou un sujet	De 1 (pas du tout d'accord) à 4 (tout à fait d'accord)	n.a.	0.722 (6)
	Généralement, dans mon travail scolaire, je dois évaluer les forces et les faiblesses d'idées ou de travaux			
Engagement des parents	Mes parents/tuteurs discutent avec moi de livres, de films ou de musique	De 1 (jamais) à 4 (toujours)	n.a.	0.676 (4)
	Mes parents/tuteurs m'interrogent sur la qualité de mon travail scolaire			
Climat en classe[2]	Quand le cours commence, je dois attendre un long moment avant que les élèves ne se calment[3]	De 1 (pas du tout d'accord) à 4 (tout à fait d'accord)	0.673 (4)	0.718 (4)
	Les élèves de cette classe sont généralement actifs et enclins à participer aux activités et aux discussions en classe			

Remarques : entre trois et six items ont été sélectionnés pour chaque sous-échelle, après réalisation d'une analyse factorielle exploratoire
1. L'indice des dispositions à l'égard de la créativité et de l'esprit critique peut également être décliné en deux sous-échelles : la sous-échelle des dispositions à l'égard de la créativité et celle des dispositions à l'égard de l'esprit critique.
2. L'indice de climat en classe est le seul à reposer sur les items tirés des déclarations des enseignants.
3. Par souci d'uniformisation, la mise à l'échelle de cet item a été inversée.

Tableau 5. Indices de sommation

Indice	Exemples d'items	Échelle	Méthode d'élaboration (nbre d'items composant l'indice indiqué entre parenthèses)	
			Primaire	Secondaire
Efficacité personnelle en arts visuels	Dans mes cours d'arts visuels, je peux même réaliser les tâches les plus difficiles si je m'y efforce Lorsqu'il s'agit d'arts visuels, j'élabore des plans que je respecte par la suite	De 1 (pas du tout) à 4 (tout à fait)	Simple moyenne type (2)	Simple moyenne type (2)
Efficacité personnelle en musique	Dans mes cours de musique, je comprends même les points les plus difficiles Lorsqu'il s'agit de musique, je persévère jusqu'à ce que la tâche soit terminée	De 1 (pas du tout) à 4 (tout à fait)	Simple moyenne type (2)	Simple moyenne type (2)
Sentiments négatifs à l'égard de l'apprentissage	J'éprouve de l'inquiétude dans ce cours Je m'ennuie dans ce cours	De 1 (jamais) à 4 (toujours)	Variable dichotomique[1] (2)	Analyse des facteurs principaux (4)
Activités culturelles	Ces dernières semaines, en dehors de l'école, j'ai dessiné, peint ou créé un objet d'art pour m'amuser Ces dernières semaines, en dehors de l'école, j'ai lu un livre en entier pour mon propre plaisir	De 1 (jamais) à 4 (4 fois ou plus)	Variable dichotomique[2] (4)	Variable dichotomique[2] (4)
Engagement des parents	Ces dernières semaines, mes parents/tuteurs ont discuté avec moi de livres, de films ou de musique Ces dernières semaines, mes parents/tuteurs m'ont interrogé sur la qualité de mon travail scolaire	De 1 (jamais) à 4 (4 fois ou plus)	Variable dichotomique[2] (4)	Analyse des facteurs principaux (4)
Sentiment d'appartenance à l'école	J'éprouve un sentiment d'appartenance à l'école Je m'ennuie souvent à l'école	De 1 (pas du tout d'accord) à 4 (tout à fait d'accord)	n.a.	Simple moyenne type (2)
Climat scolaire en ce qui concerne la créativité et l'esprit critique	L'école est un lieu qui m'aide à faire preuve de créativité L'école est un lieu qui m'aide à envisager différents points de vue	De 1 (pas du tout d'accord) à 4 (tout à fait d'accord)	n.a.	Simple moyenne type (2)

Remarques :
1. Méthodologie décrite dans le document intitulé « *Investigating Optimal Learning Moments in U.S. and Finnish Science Classes* », rédigé par Schneider et al. en 2016.
2. Dans ces cas, la fonction déterminant la valeur de la variable dichotomique indiquait lorsque les élèves répondaient à au moins l'un des items contenus dans l'indice

Appariement des coefficients de propension

En termes de variables socio-démographiques, les équipes locales avaient été invitées à sélectionner des établissements, des enseignants et des élèves provenant, dans la mesure du possible, d'une pluralité de contextes, et à garantir la comparabilité entre le groupe de contrôle et le groupe expérimental sur le plan des caractéristiques des élèves et de la qualité des enseignants. Malgré les efforts déployés par les équipes, des différences substantielles entre ces deux groupes ont été constatées lors de la collecte de données. Afin de réduire au minimum l'impact de ces différences sur les conclusions du projet, un appariement des coefficients de propension a été réalisé. Cette technique consiste à réaligner les situations initiales des deux échantillons en attribuant aux élèves du groupe de contrôle différents coefficients de pondération. Un ensemble spécifique de coefficients de pondération a été calculé pour chacun des instruments d'enquête. Pour le questionnaire, le calcul de ces coefficients visait à corriger des déséquilibres potentiels relatifs au sexe, au milieu socio-économique et à l'âge. Pour les tests de performance, l'objectif consistait également à corriger les déséquilibres potentiels dans les données initiales de leur principale variable à l'étude (à savoir les scores au test EPoC d'évaluation de la créativité, au test de STIM et au test d'AVM).

Pour chaque élève, l'appariement établissait un score de propension qui traduisait sa probabilité d'être intégré au groupe expérimental. Un coefficient de pondération propre à chaque élève a ensuite été calculé et attribué à chaque élève du groupe de contrôle (par construction, celui des élèves du groupe expérimental ayant été fixé à 1), sur la base des similarités observées avec l'élève du groupe expérimental avec lequel il était comparé. Dans la pratique, c'est l'appariement avec fonction noyau qui a été utilisé, un coefficient de pondération supérieur ayant été attribué aux élèves du groupe de contrôle qui se rapprochaient le plus de l'élève considéré dans le groupe expérimental. Une fois les coefficients de pondération inclus dans l'analyse, les différences observées entre les deux groupes ont été considérablement réduites. Dans tous les modèles pris en compte dans le cadre de cette étude pilote – à l'exception d'un seul – les différences résiduelles entre les groupes n'étaient pas significatives. Par voie de conséquence, toutes les évolutions survenues entre les deux groupes ainsi qu'entre le début et la fin de l'expérience peuvent être entièrement imputables à l'expérience et non pas, potentiellement, aux différences préexistantes entre les deux groupes.

Dans le cadre de cette étude pilote, l'appariement des coefficients de propension a été effectué par pays et niveau d'enseignement, et la liste des variables prises en compte pour cet appariement variait en fonction de l'instrument analysé (questionnaires, tests de STIM, d'AVM ou test EPoC). Plus précisément, la liste des variables utilisées pour l'appariement dans ces quatre instruments incluait le sexe, l'âge, l'âge au carré et l'indice de statut socio-économique. De plus, pour chacun des trois tests, la liste comportait également le score obtenu par chaque élève au test pré-expérience.

Une option d'appoint courante a également été utilisée lors de l'appariement, impliquant l'exclusion de certains élèves du groupe expérimental dont le score de propension était supérieur au maximum fixé ou inférieur au score de propension minimum pour les élèves du groupe de contrôle (voir le Tableau 6).

Tableau 6. Appariement des coefficients de propension

Pourcentage d'élèves retenus (et nombre d'élèves exclus) après l'appariement des coefficients de propension

	Questionnaires	Tests EPoC d'évaluation de la créativité	Tests de STIM	Tests d'AVM
Équipe brésilienne	98 % (9)	96 % (25)	75 % (76)	..
Équipe néerlandaise	96 % (22)	87 % (48)	99 % (4)	95 % (13)
Équipe française (CRI)	99 % (3)	100 % (0)	98 % (6)	..
Équipe française (Lamap)	..	95 % (16)	84 % (6)	..
Équipe hongroise	95 % (54)	83 % (131)	94 % (68)	..
Équipe indienne	66 % (103)	..	46 % (177)	..
Équipe russe	96 % (21)	89 % (91)	99 % (4)	..
Équipe slovaque	93 % (26)	79 % (115)	98 % (6)	76 % (77)
Équipe espagnole	..	..	82 % (48)	..
Équipe thaïlandaise	95 % (178)	88 % (383)	97 % (84)	60 % (260)
Équipe britannique (Pays de Galles)	96 % (22)	77 % (167)	91 % (56)	..
Équipe américaine (Montessori)	..	x	x	..
Équipe américaine (Vista)	97 % (24)	86 % (54)	54 % (107)	44 % (142)
Total	**95 % (462)**	**86 % (1 123)**	**90 % (732)**	**67 % (492)**

Évaluation des impacts

Cette section offre un aperçu de la méthodologie utilisée pour évaluer les effets de l'expérience dans les chapitres 6 et 7.

Enseignants

L'analyse de l'évolution des pratiques et attitudes des enseignants a entièrement reposé sur les données tirées des questionnaires « Enseignant » pré- et post-expérience. Ces questionnaires comprenaient plusieurs batteries de questions portant, entre autres, sur les convictions des enseignants à l'égard de la créativité et de l'esprit critique, leurs pratiques pédagogiques et d'évaluation ainsi que le degré de préparation qu'ils estimaient posséder. Par ailleurs, ils contenaient également des informations contextuelles d'ordre général.

Le projet s'est appuyé sur un plan d'échantillonnage dirigé, avec le recrutement d'équipes, d'établissements et d'enseignants essentiellement en fonction d'un intérêt commun pour les priorités du projet. Les échantillons ne prétendaient donc pas être représentatifs de leur système d'éducation respectif.

Sur les 728 enseignants ayant pris part au projet, 90 ont participé à deux cycles de l'expérience au cours de deux années scolaires différentes dans leur équipe respective. Le nombre d'enseignants participants indiqué dans les tableaux statistiques était de 818, dans la mesure où ce chiffre tenait mieux compte du nombre de classes impliquées dans le projet. Les données tirées des questionnaires du panel pour ces 90 enseignants ont été traitées séparément, car l'analyse n'avait fait état d'aucune différence significative dans le profil des réponses entre les deux années scolaires à l'étude.

Les taux de participation et de réponse variaient entre les pays. Dans la majorité des équipes, la taille restreinte de l'échantillon d'enseignants ayant répondu aux questionnaires pré- et post-expérience a empêché toute analyse solide des doubles différences. Une fois divisé pour comparer les résultats entre les enseignants dans le groupe de contrôle et ceux dans le groupe expérimental, le faible nombre de réponses n'a pas permis de disposer d'une puissance statistique suffisante pour distinguer les effets réels de la variation aléatoire. Il a donc été décidé d'axer les analyses des évolutions engendrées par l'expérience uniquement sur les sous-échantillons du panel des équipes hongroise et thaïlandaise. Ces deux sous-échantillons ont été choisis car ils étaient les seuls à respecter les critères suivants :

- inclure au moins 30 enseignants ayant répondu aux questionnaires pré- et post-expérience ;
- s'appuyer sur une juste répartition des répondants entre le groupe expérimental et le groupe de contrôle (entre 45 et 55 % pour chaque groupe) ;
- représenter plus de 50 % du nombre total d'enseignants participants dans leur équipe ;
- représenter plus de 80 % du nombre d'enseignants ayant rempli les questionnaires pré-expérience dans leur équipe.

Dans les autres équipes, les travaux d'analyse sur les professionnels de l'éducation se sont limités aux rapports de référence, aux questionnaires post-expérience ou aux questionnaires « Établissement », c'est pourquoi ils n'ont pas permis de déduire d'inférences causales concernant les impacts de l'expérience.

Pour les équipes hongroise et thaïlandaise, des régressions linéaires avec estimation groupée de la variance (premier niveau : établissement) ont permis d'effectuer des comparaisons avant et après expérience sous forme d'estimations des doubles différences. La variable dépendante était la différence des résultats d'intérêt entre les mesures pré- et post-expérience (pratiques, convictions, efficacité personnelle, etc.), et la seule variable explicative était l'exposition à l'expérience (par apposition au groupe de contrôle). Par construction, les estimations indiquaient l'effet de l'appartenance au groupe expérimental, et l'ampleur de cet effet a fait fonction de test solide des moyens.

Pour les autres équipes, les questionnaires post-expérience ont permis de tirer de précieuses conclusions sur la manière dont les enseignants avaient interagi avec les référentiels de compétences et plans de cours du projet, sur les éléments qu'ils n'avaient pas retenus dans le cadre de l'expérience pédagogique ainsi que sur la mesure dans laquelle ils avaient intégré des pratiques innovantes dans leurs activités d'enseignement au quotidien.

Élèves

À l'instar de la méthode utilisée pour les enseignants, la principale variable explicative étudiée pour les élèves était leur participation à l'expérience (par opposition au fait d'appartenir à une classe du groupe de contrôle). Cette variable muette se retrouvait dans l'ensemble des modèles. Par construction, le coefficient estimé de cette variable montrait si, pour les élèves, le fait d'avoir des enseignants ayant participé aux plans de développement pédagogique avait eu ou non un effet positif sur différents résultats d'intérêt.

La disponibilité des données variait selon les domaines, les équipes et les niveaux d'enseignement. Parallèlement, les informations tirées des questionnaires « Établissement » et « Enseignant » n'étaient pas disponibles pour l'ensemble des élèves. Les variables explicatives intégrées aux modèles divergeaient donc sensiblement d'un contexte à l'autre, même si aucune analyse n'était menée en cas d'absence de données contextuelles sur les élèves.

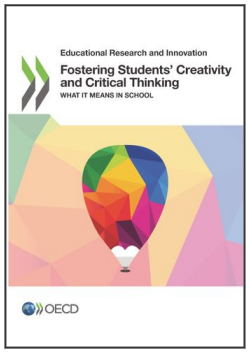
De façon plus spécifique, tous les modèles comportaient des variables relatives à l'âge, au sexe et au milieu socio-économique. Figuraient également parmi ces variables, lorsque les informations étaient disponibles, le temps écoulé entre les mesures pré- et post-expérience ainsi que la durée de l'expérience auprès des élèves (c'est-à-dire le nombre de minutes consacrées aux activités pédagogiques mises en œuvre pour ce projet). Le domaine dans lequel l'expérience avait lieu était aussi inclus dans ces modèles le cas échéant (c'est-à-dire uniquement lorsqu'au moins quelques classes participaient à l'expérience dans plusieurs domaines). Enfin, comme cela avait été fait pour l'appariement des coefficients de propension, les modèles comportaient toujours comme covariable la valeur du résultat d'intérêt lors de la mesure pré-expérience (le score au test d'AVM au début du projet, par exemple).

Chaque modèle consistait en une régression linéaire avec estimation groupée de la variance (premier niveau : établissement) afin de tenir compte de la corrélation intragroupe, et un modèle séparé était exécuté pour chaque résultat, pays et niveau d'enseignement. Dans tous les contextes, les coefficients de pondération obtenus grâce à l'appariement des coefficients de propension ont été intégrés afin de garantir la comparabilité entre le groupe expérimental et le groupe de contrôle.

Comme mentionné au chapitre 7, l'analyse multivariée comportait les deux étapes suivantes :

- une régression des résultats d'intérêt sur toutes les covariables disponibles ;
- une régression des résultats d'intérêt sur toutes les covariables disponibles et sur les paramètres d'interaction (un à la fois), avec l'association de la variable indiquant le groupe dans lequel se trouvent les élèves et les covariables visées.

Cette stratégie d'analyse a permis d'examiner les effets de l'expérience sur les élèves de manière globale, et de se concentrer sur les effets qu'elle avait engendrés pour des sous-groupes d'élèves particuliers afin de déterminer si elle avait eu différents effets pour certains de ces sous-groupes (par exemple, l'expérience a-t-elle mieux fonctionné avec les filles qu'avec les garçons ?). Cette approche a toutefois impliqué l'estimation de centaines de modèles. En vue de les présenter de façon claire et accessible, les tableaux du chapitre 7 indiquent le nombre de résultats statistiquement positifs ou négatifs pour les différents éléments à l'étude. De cette manière, au lieu de se concentrer sur les résultats individuels par pays et niveau d'enseignement – qui ne sont que d'ordre préliminaire dans le cadre de cette phase pilote – les lecteurs sont incités à examiner de manière plus globale les tendances qui ressortent des données pour ce qui a trait aux effets de l'expérience.



Extrait de :

Fostering Students' Creativity and Critical Thinking

What it Means in School

Accéder à cette publication :

https://doi.org/10.1787/62212c37-en

Merci de citer ce chapitre comme suit :

Vincent-Lancrin, Stéphan*, et al.* (2020), « Annexe technique », dans Stéphan Vincent-Lancrin*, et al.*, *Fostering Students' Creativity and Critical Thinking : What it Means in School*, Éditions OCDE, Paris.

DOI: https://doi.org/10.1787/5c1da38b-fr

Cet ouvrage est publié sous la responsabilité du Secrétaire général de l'OCDE. Les opinions et les arguments exprimés ici ne reflètent pas nécessairement les vues officielles des pays membres de l'OCDE.

Ce document, ainsi que les données et cartes qu'il peut comprendre, sont sans préjudice du statut de tout territoire, de la souveraineté s'exerçant sur ce dernier, du tracé des frontières et limites internationales, et du nom de tout territoire, ville ou région. Des extraits de publications sont susceptibles de faire l'objet d'avertissements supplémentaires, qui sont inclus dans la version complète de la publication, disponible sous le lien fourni à cet effet.

L'utilisation de ce contenu, qu'il soit numérique ou imprimé, est régie par les conditions d'utilisation suivantes : http://www.oecd.org/fr/conditionsdutilisation.