



A collaboration between



**Développer les compétences
socio-émotionnelles dans la pré-adolescence :
Données probantes d'une intervention
dans les collèges publics marocains**

Andreas de Barros

Alejandra Campos Quintero

Paul Glewwe

Nikhil Kumar

Laure Lépine

17 juin 2026

Développer les compétences socio-émotionnelles dans la pré-adolescence : Données probantes d'une intervention dans les collèges publics marocains*

Andreas de Barros[†]

Alejandra Campos Quintero[‡]

Paul Glewwe[§]

Nikhil Kumar[¶]

Laure Lépine^{||}

17 juin 2026

Nous étudions une intervention de soutien socio-émotionnel mise en oeuvre par le gouvernement dans les collèges publics marocains. L'intervention a formé des spécialistes sociaux déjà en poste pour animer des ateliers en petits groupes auprès des élèves de la 1^{ère} et de la 2^{ème} année collégiale. Dans le cadre d'un essai randomisé par cluster portant sur 200 collèges, tous participant à la première année d'une réforme plus large, 84 ont été assignés à l'intervention et 116 au groupe de contrôle. L'assignation a augmenté la participation aux ateliers de 43,2 points de pourcentage et la connaissance qu'ont les élèves de leur spécialiste social de 16,9 points de pourcentage. Au cours d'une année scolaire, toutefois, cette augmentation de l'exposition ne s'est pas traduite par des effets moyens détectables sur les résultats principaux pré-spécifiés (compétences interpersonnelles et intrapersonnelles) ni sur les apprentissages globaux, l'abandon

*Ce document présente le rendu de la deuxième phase d'un Rapport pré-enregistré conditionnellement accepté par le processus de revue pré-résultats du Journal of Development Economics. Nous remercions le Ministère de l'Éducation Nationale, du Préscolaire et des Sports du Maroc d'avoir soutenu à la fois cette étude et une évaluation d'impact distincte du programme des collèges pionniers dans les collèges publics. Nous remercions chaleureusement le Ministère pour son généreux financement. Nous sommes reconnaissants à Naila Ebeid, Alice Escande, Naomi Duhau-Marmon et Maxime Sirera, du Behavioural Insights Team (BIT), qui ont conçu l'intervention de soutien socio-émotionnel avec le Ministère et conduit l'analyse diagnostique et la recherche qualitative qui l'ont éclairée. Nous sommes reconnaissants au Morocco Innovation and Evaluation Lab (MEL) et au personnel de J-PAL—Sara Ameziane-Hassani, Youssef Assarsah, Arthur Bluche, Ayoub Chaabi, Florencia Devoto, Najiba El Amrani Mida et Zakaria Mansouri—pour leur excellente coordination de terrain, leur assistance à la recherche et leur direction du projet. Nous remercions Quentin Daviot et Thomas Escande d'EVAL-LAB pour leur excellent soutien dans le processus de collecte de données. La protection des données a été assurée dans le respect des dispositions légales et réglementations locales applicables en matière de recherche sur les sujets humains à l'Université de Californie à Irvine, à l'Université Harvard et à l'Université du Minnesota. L'ordre des auteurs est alphabétique; le statut de co-auteur est partagé à parts égales entre les auteurs.

[†]University of California, Irvine. E-mail : adb@uci.edu.

[‡]Columbia University. E-mail : aac2271@tc.columbia.edu.

[§]University of Minnesota. E-mail : pglewwe@umn.edu.

[¶]Harvard University. E-mail : nikhil_kumar1@g.harvard.edu.

^{||}University Mohammed VI Polytechnic, E-mail : llepine@povertyactionlab.org.

scolaire, la créativité, le bien-être ou les habitudes de travail. Si quoi que ce soit, la maîtrise de soi des élèves a diminué de 0,097 écart-type. Nous observons néanmoins des améliorations dans deux ensembles de résultats secondaires : le redoublement a diminué de 1,7 point de pourcentage, et les résultats des élèves ont augmenté dans certaines matières et sous-domaines : physique, chimie et biologie, production écrite en arabe et en français, et expression orale en arabe. Tous ces résultats tiennent dans le cadre d'une hiérarchie pré-spécifiée de tests d'hypothèses qui contrôle le taux de fausses découvertes. Ainsi, même si le personnel scolaire existant peut accroître l'exposition des élèves au soutien sans personnel supplémentaire ni hausse de rémunération, un protocole universel de faible intensité n'a pas amélioré de manière mesurable les compétences socio-émotionnelles ciblées ; au contraire, les effets positifs ont porté sur des résultats académiques et de progression scolaire.

Mots clés : Education ; compétences socio-émotionnelles ; capital humain ; Maroc.

Classification JEL : I20, I21, I28, O15, O22.

Pré-enregistrement de l'étude : Cette étude a été pré-enregistrée en tant que *Registered Report* et conditionnellement acceptée par le *Journal of Development Economics* dans le cadre de son processus de revue pré-résultats.

Résumé exécutif

Une intervention de soutien socio-émotionnel mise en oeuvre par le gouvernement.

Cette étude évalue une intervention pilotée par le gouvernement et conçue pour développer les compétences socio-émotionnelles des jeunes adolescents dans les collèges publics marocains. L'intervention s'inscrivait dans la première année du programme des collèges pionniers (PCP), la réforme phare du Ministère de l'Éducation Nationale, du Préscolaire et des Sports pour les collèges. Elle a formé et outillé le spécialiste social déjà en poste dans chaque établissement (un fonctionnaire déjà affecté au collège) pour animer quatre ateliers en petits groupes au cours de l'année scolaire auprès des élèves de 1^{ère} et 2^{ème} année collégiale. Les ateliers couvraient la conscience de soi, la reconnaissance et la régulation des émotions, le sentiment d'appartenance, la coopération, l'empathie, l'écoute active et la résolution de conflits ; ils visaient également à accroître la familiarité des élèves avec le spécialiste social de leur établissement et leur confiance envers lui. L'intervention reposait sur le personnel scolaire existant, sans personnel supplémentaire ni hausse de rémunération pour les établissements participants.

Transparence de la recherche et résultats étudiés. L'étude a été pré-enregistrée en tant que *Registered Report* et conditionnellement acceptée par le *Journal of Development Economics* dans le cadre de son processus de revue pré-résultats, ce qui signifie que le dispositif, les hypothèses et le plan d'analyse ont été évalués par les pairs et acceptés avant la collecte des données de fin d'année. Nos résultats principaux pré-spécifiés sont deux familles de compétences socio-émotionnelles : les compétences *interpersonnelles* (pro-socialité et perception des émotions) et les compétences *intrapersonnelles* (sentiment de contrôle, ainsi qu'autorégulation et discipline). Nous avons également pré-spécifié des résultats induits (apprentissage académiques et abandon scolaire) et des résultats additionnels (créativité, climat scolaire et bien-être, et habitudes de travail). Nous avons pré-spécifié deux sous-groupes d'intérêt particulier pour les politiques publiques : les élèves identifiés comme étant à risque d'abandon scolaire et les élèves filles. Tous les tests sont corrigés pour les tests d'hypothèses multiples à l'aide de valeurs q fondées sur le taux de fausses découvertes.

Dispositif de recherche et données. Nous évaluons l'intervention au moyen d'un essai contrôlé randomisé par clusters, pré-enregistré, dans 200 collèges publics participant au programme des collèges pionniers. Parmi eux, 84 collèges ont été assignés aléatoirement à animer les ateliers socio-émotionnels pendant l'année d'étude, et les 116 collèges restants

ont constitué le groupe de contrôle. Étant donné que chaque collège de l'étude mettait déjà en oeuvre le programme des collèges pionniers dans son ensemble, l'essai identifie l'effet *marginal* de l'ajout de la composante socio-émotionnelle au sein d'un système en cours de réforme, et non l'effet d'une programmation socio-émotionnelle par rapport à un statu quo non réformé, ni l'effet global de la réforme elle-même. Nous avons collecté des données primaires d'évaluation et d'entretiens individuels auprès d'un échantillon effectif de 11 132 élèves, complétées par des registres administratifs sur la progression scolaire. Les collèges des groupes traité et de contrôle étaient équilibrés en début d'année sur les caractéristiques observées, avec un léger déséquilibre dans la proportion d'élèves filles et l'âge moyen des élèves ; un test conjoint ne montre aucun déséquilibre systématique d'ensemble.

L'intervention a accru l'exposition des élèves au soutien. L'assignation à l'intervention a produit une augmentation importante et statistiquement significative de l'exposition déclarée des élèves au soutien socio-émotionnel. Les élèves des collèges du groupe traité étaient 43,2 points de pourcentage plus susceptibles de déclarer avoir participé à des ateliers socio-émotionnels (par rapport à 21,7 pour cent dans les collèges du groupe de contrôle) et 16,9 points de pourcentage plus susceptibles de connaître le spécialiste social de leur établissement (par rapport à 43,5 pour cent dans les collèges du groupe de contrôle). L'exposition aux autres composantes du programme des collèges pionniers (remédiation ciblée, enseignement explicite et tutorat) était statistiquement indiscernable entre les collèges des groupes traité et de contrôle, de sorte que la comparaison isole la composante socio-émotionnelle plutôt qu'une exposition différentielle à la réforme dans son ensemble. Nous observons cette adhésion à travers les déclarations des élèves indiquant avoir participé à un atelier avec le spécialiste social. Cette mesure, toutefois, pourrait souffrir d'un biais de rappel et ne donne aucun détail sur le nombre d'ateliers menés dans chaque collège ni sur leur qualité.

Aucun effet moyen détectable sur les compétences socio-émotionnelles. Malgré cette augmentation de l'exposition, nous ne trouvons aucune preuve que l'intervention ait amélioré les compétences socio-émotionnelles des élèves, en moyenne, au cours d'une année scolaire. L'effet estimé sur l'indice de compétences interpersonnelles est de +0,029 écart-type (é.-t.), et l'effet estimé sur l'indice de compétences intrapersonnelles est de +0,012 é.-t. ; les deux sont faibles et statistiquement indiscernables de zéro. Les sous-composantes sous-jacentes suivent le même schéma, à une exception près : la maîtrise de soi des élèves a diminué de 0,097 é.-t.

Aucun effet général sur l'apprentissage ou l'abandon scolaire, à deux ensembles d'exceptions près. Nous ne trouvons aucun effet sur les apprentissages académiques globaux ni sur l'abandon scolaire. L'effet estimé sur un indice agrégé des scores aux tests en arabe, français, mathématiques et sciences est de +0,034 é.-t., et l'effet estimé sur l'abandon scolaire en fin d'année est essentiellement nul (par rapport à un taux d'abandon de 2,6 pour cent dans le groupe de contrôle); les estimations par matière sont faibles et ne survivent pas à la correction pour tests d'hypothèses multiples. Des améliorations apparaissent dans deux ensembles de résultats secondaires qui survivent à cette correction. D'abord, le redoublement a diminué de 1,7 point de pourcentage. Ensuite, les résultats des élèves ont augmenté dans certaines matières et sous-domaines : au sein des sciences, en physique, en chimie et en biologie, et au sein des langues, en production écrite en arabe et en français et en expression orale en arabe. Nous lisons cette baisse du redoublement comme cohérente avec le rôle attendu du spécialiste social, qui consiste à soutenir les élèves susceptibles de redoubler ; les gains langagiers sont plus interprétables, se concentrant sur les composantes de production des évaluations (l'écrit et, en arabe, l'oral) plutôt que sur la compréhension de l'écrit, ce qui est cohérent avec l'accent mis par les ateliers sur la communication et l'expression ; et nous interprétons les gains en sciences avec davantage de prudence, comme des effets dont le canal est plus difficile à expliquer, puisque les ateliers ne ciblaient pas les contenus disciplinaires, et comme des résultats à approfondir dans de futurs travaux. L'intervention n'a pas produit d'améliorations détectables de la créativité des élèves, du climat scolaire et du bien-être, ou des habitudes de travail en dehors de l'école.

Sous-groupes pré-spécifiés. Nous avons pré-spécifié deux sous-groupes d'intérêt pour les politiques publiques, les élèves à risque d'abandon scolaire et les élèves filles. Parmi les élèves à risque, l'effet estimé sur les compétences intrapersonnelles est de 0,144 é.-t. ; il est statistiquement significatif au seuil de 10 pour cent avant correction pour tests d'hypothèses multiples mais ne survit pas à cette correction ($q = 0,281$) ; l'estimation correspondante pour leurs compétences interpersonnelles est essentiellement nulle (0,004 é.-t.). Nous ne trouvons aucun effet détectable pour les élèves filles.

Robustesse. Les résultats sont robustes à une série de vérifications. Les résultats relatifs à l'abandon scolaire sont stables selon différentes définitions de l'échantillon d'inscription, et nous ne trouvons aucune preuve que l'absence d'effets moyens sur les compétences socio-émotionnelles s'explique par un biais de désirabilité sociale dans les réponses des élèves aux enquêtes.

Interpréter les résultats. Plusieurs caractéristiques de l'étude orientent la lecture de ces résultats. Premièrement, parce que chaque collègue de l'étude faisait déjà partie du programme des collèges pionniers, l'essai mesure la contribution additionnelle des ateliers par-dessus un système déjà en cours de réforme, un point de comparaison exigeant plutôt qu'un test par rapport à un statu quo non réformé. Deuxièmement, la dose était faible, à quatre ateliers sur l'année, près du bas de la fourchette que les méta-analyses associent à des effets moyens détectables; et nous ne pouvons pas observer avec quelle fidélité même cette dose a été délivrée, de sorte qu'un résultat nul sur les compétences socio-émotionnelles ciblées peut refléter soit les effets modestes d'ateliers bien menés, soit des insuffisances dans leur mise en oeuvre. Troisièmement, nous avons délibérément mesuré les compétences socio-émotionnelles de manière large, au sein d'un cadre pré-enregistré également utilisé pour évaluer la réforme plus large du Ministère, plutôt qu'au moyen de mesures étroites adaptées aux ateliers. Cette largeur est intentionnelle : elle teste si d'éventuels gains se généralisent aux résultats pertinents pour les politiques publiques, et elle maintient la comparabilité de nos mesures entre études. Son coût est que certains instruments se situent à distance de ce que les quatre séances entraînent le plus directement, et qu'une mesure en fin d'année peut être précoce pour saisir des compétences qui se consolident avec le temps. Les résultats nuls sont donc les plus informatifs pour les compétences larges et pertinentes pour les politiques publiques que nous avons entrepris de mesurer, et moins informatifs sur les compétences étroitement entraînées que le dispositif de recherche n'était pas conçu pour isoler.

Perspectives. Cette étude montre qu'un gouvernement peut mobiliser son personnel scolaire existant pour modifier l'exposition des élèves au soutien socio-émotionnel dans les conditions ordinaires de l'école publique, sans coût de personnel supplémentaire. À la dose déployée ici, ce contact n'a pas fait progresser les compétences socio-émotionnelles que les ateliers ciblaient, et la maîtrise de soi a même, si quoi que ce soit, diminué. Il a toutefois coïncidé avec une baisse du redoublement et une hausse des résultats dans certains sous-domaines des sciences et des langues. Atteindre les élèves de cette manière est faisable et peu coûteux; traduire ce contact en gains sur les compétences ciblées pourrait nécessiter un dispositif plus intensif et plus ciblé. Les prochaines étapes prometteuses incluent une version des ateliers à dose plus élevée, une étude qualitative de la manière dont les séances sont animées afin d'identifier et de corriger les points à renforcer, une attention portée aux raisons pour lesquelles le redoublement et certains sous-domaines académiques se sont améliorés alors que les compétences ciblées n'ont pas évolué, et un enregistrement systématique des ateliers délivrés à mesure que le programme s'étend.

Points clés

- Le gouvernement marocain a formé et outillé les spécialistes sociaux déjà en poste dans les collèges publics pour animer des ateliers socio-émotionnels en petits groupes auprès des élèves de 1^{ère} et 2^{ème} année collégiale, comme l'une des composantes de son programme des collèges pionniers.
- Dans un essai pré-enregistré, randomisé par clusters, portant sur 200 collèges, l'intervention a augmenté la participation déclarée des élèves aux ateliers (de 43,2 points de pourcentage) et leur connaissance du spécialiste social de l'établissement (de 16,9 points de pourcentage).
- Sur une année scolaire, l'intervention n'a produit aucun effet moyen détectable sur les compétences socio-émotionnelles ciblées, ni sur les apprentissages globaux, l'abandon scolaire, la créativité, le bien-être ou les habitudes de travail ; si quoi que ce soit, la maîtrise de soi des élèves a diminué de 0,097 é.-t. (un déclin robuste que nous signalons comme une énigme). Les deux sous-groupes pré-spécifiés (les élèves à risque d'abandon scolaire et les élèves filles) ne présentent de même aucun effet qui survive à la correction pour tests d'hypothèses multiples.
- Deux ensembles de résultats secondaires se sont améliorés et ont survécu à la correction pour tests d'hypothèses multiples : le redoublement a diminué de 1,7 point de pourcentage, et les résultats des élèves ont augmenté dans certains sous-domaines des sciences et des langues (la physique, la chimie et la biologie ; la production écrite en arabe et en français ; et l'expression orale en arabe).
- L'étude montre que le système éducatif existant peut déployer cette intervention sans personnel ni rémunération supplémentaires ; faire progresser les compétences socio-émotionnelles ciblées pourrait toutefois nécessiter un dispositif plus intensif et plus ciblé.

Table des matières

Résumé exécutif	iii
1 Introduction	1
2 Contexte et intervention	3
2.1 Les collèges publics marocains	3
2.2 L'intervention de soutien socio-émotionnel	4
2.3 Théorie du changement	6
3 Dispositif d'étude	7
3.1 Échantillonnage des collèges	7
3.2 Randomisation des collèges	8
3.3 Échantillonnage des élèves	8
3.4 Mesure des résultats	10
3.5 Exposition au programme	11
3.6 Stratégie empirique	12
4 Résultats	13
4.1 Compétences socio-émotionnelles	13
4.2 Apprentissage	14
4.3 Abandon scolaire et redoublement	15
4.4 Résultats additionnels	15
5 Hétérogénéité et robustesse	15
5.1 Effets par sous-groupes	15
5.2 Robustesse et sensibilité	16
6 Conclusion	17
Tableaux	22

Références	27
Annexe A Tableaux et figures supplémentaires	32
Annexe B Mesure	43
B.1 Modèle logistique à deux paramètres (2PL)	43
B.2 Modèle de réponse graduée (GRM)	44
B.3 Analyse en composantes principales de Poisson (ACP de Poisson)	44
B.4 Agrégation des scores latents	44
Annexe C Ajustements pré-spécifiés pour les tests d'hypothèses multiples	48

1 Introduction

Les systèmes éducatifs cherchent de plus en plus à développer des compétences socio-émotionnelles qui façonnent l'apprentissage, la persévérance, les relations sociales et les trajectoires de vie (Deming, 2017). La pré-adolescence s'accompagne de nouvelles exigences scolaires, de l'évolution des relations avec les pairs et d'un risque accru de décrochage scolaire. Pourtant, les politiques permettant de développer ces compétences à grande échelle chez les élèves de cette tranche d'âge restent mal comprises. Les enseignants qui améliorent les résultats aux tests ne sont pas nécessairement ceux qui améliorent les compétences socio-émotionnelles (Jackson, 2018), et une grande partie des données probantes sur les programmes ciblés de développement socio-émotionnel provient de contextes à revenus élevés ou d'interventions menées par des chercheurs et des organisations non gouvernementales. On sait peu de choses sur la capacité des gouvernements à mobiliser le personnel des écoles publiques existantes pour offrir un soutien socio-émotionnel dans le cadre ordinaire de l'enseignement.

Cet article étudie une intervention de soutien socio-émotionnel pilotée par le gouvernement dans les collèges publics marocains. L'intervention s'inscrivait dans la première année du programme des collèges pionniers, une réforme plus large de ces établissements. Elle a formé et outillé des spécialistes sociaux pour animer quatre ateliers de deux heures par an en petits groupes auprès des élèves de 1^{ère} et 2^{ème} année collégiale. Ces ateliers portaient sur la conscience de soi, la reconnaissance et la régulation des émotions, le sentiment d'appartenance, la coopération, l'empathie, l'écoute active et la résolution de conflits. L'intervention visait également à accroître la familiarité des élèves avec le spécialiste social de leur établissement et à instaurer une relation de confiance entre les élèves et le spécialiste.

Nous évaluons l'intervention au moyen d'un essai randomisé par clusters portant sur 200 collèges publics, participant tous au programme des collèges pionniers. Parmi eux, 84 ont été assignés aléatoirement à recevoir l'intervention de soutien socio-émotionnel, tandis que 116 constituaient le groupe de contrôle. Le dispositif identifie donc l'effet marginal de l'affectation à la composante socio-émotionnelle dans ce contexte de réforme. Nos résultats principaux sont les compétences interpersonnelles et intrapersonnelles des élèves. Nous examinons également des résultats induits, notamment les compétences académiques et l'abandon scolaire, ainsi que des résultats additionnels liés à la créativité, au bien-être et aux habitudes de travail.

L'intervention a modifié l'exposition au soutien socio-émotionnel. Les élèves des collèges du groupe traité étaient 43,2 points de pourcentage plus susceptibles de déclarer avoir participé à des ateliers socio-émotionnels, par rapport à une moyenne du groupe de

contrôle de 21,7 pour cent. Ils étaient également 16,9 points de pourcentage plus susceptibles de connaître le spécialiste social de leur établissement, par rapport à une moyenne du groupe de contrôle de 43,5 pour cent. L'exposition aux autres composantes du programme des collèges pionniers était équilibrée entre les collèges du groupe traité et ceux du groupe de contrôle. Ces résultats indiquent que l'intervention a créé un contraste substantiel dans les interactions des élèves avec les spécialistes sociaux et les ateliers. Dans le même temps, environ un élève sur trois du groupe traité n'a pas déclaré avoir participé à un atelier. Nous considérons cette adhésion incomplète comme une caractéristique de l'étude des effets d'intention-de-traiter d'une réforme déployée à grande échelle par le gouvernement.

Malgré cette augmentation de l'exposition, nous ne trouvons aucune amélioration moyenne détectable des résultats principaux pré-spécifiés au cours de la première année. Plus précisément, l'effet estimé sur les compétences interpersonnelles est de 0,029 écart-type (é.-t.), et celui sur les compétences intrapersonnelles est de 0,012 é.-t., deux estimations faibles et statistiquement indiscernables de zéro. Si quoi que ce soit, un sous-domaine des compétences intrapersonnelles a diminué de 0,097 é.-t. (la maîtrise de soi des élèves). Nous ne trouvons pas non plus d'effets détectables sur les compétences académiques globales ni sur les taux d'abandon scolaire. L'effet estimé sur les résultats aux tests est de 0,034 é.-t., et l'effet estimé sur l'abandon scolaire en fin d'année est approximativement nul. Les résultats additionnels, notamment la créativité, le climat scolaire et le bien-être, et les habitudes de travail, ne montrent pareillement aucun effet détectable.

Néanmoins, des améliorations apparaissent dans deux ensembles de résultats secondaires : le redoublement a diminué de 1,7 point de pourcentage, et les résultats des élèves ont augmenté dans certaines matières et sous-domaines, notamment la physique, la chimie et la biologie, la production écrite en arabe et en français, et l'expression orale en arabe. Toutes ces estimations découlent d'une hiérarchie pré-spécifiée de tests d'hypothèses qui contrôle le taux de fausses découvertes, alors même que les scores globaux aux tests et l'abandon scolaire ne montrent aucun changement détectable.

Ces résultats contribuent à trois corpus de littérature. Premièrement, nous enrichissons une littérature restreinte sur les interventions socio-émotionnelles en milieu scolaire dans les pays à faibles et à moyens revenus. Des méta-analyses et revues documentent les effets des programmes socio-émotionnels dans les contextes à revenus élevés (Wilson et al., 2025 ; Cipriano et al., 2023), et des études apparentées dans des contextes à revenus intermédiaires évaluent des interventions telles que les programmes axés sur la mentalité de croissance (*growth-mindset*) ou les curricula socio-émotionnels (Alan et al., 2019 ; Ganimian, 2020 ; Santos et al., 2022). Les données probantes provenant de contextes à faibles revenus restent moins nombreuses et proviennent souvent de programmes mis en oeuvre par des

chercheurs ou des ONG en Amérique centrale, en Inde et en Zambie (Dinarte-Diaz et al., 2024; Edmonds et al., 2023; Dhar et al., 2022; Ashraf et al., 2020). À notre connaissance, il s'agit de la première évaluation expérimentale à grande échelle d'une intervention de développement des compétences socio-émotionnelles pilotée par le gouvernement et mise en oeuvre dans les écoles publiques d'un pays à revenu faible ou intermédiaire de la tranche inférieure.

Deuxièmement, cet article contribue à la recherche sur la mise en oeuvre des réformes éducatives par le secteur public. De nombreuses interventions éducatives couronnées de succès dans les pays en développement ont reposé sur des ONG, des partenariats public-privé ou d'autres structures de mise en oeuvre non gouvernementales (Banerjee et al., 2017; Eble et al., 2021; Romero et al., 2020). Les tentatives d'amélioration des écoles publiques en s'appuyant exclusivement sur le personnel et les systèmes administratifs de l'État ont souvent rencontré des difficultés (de Barros et al., 2024). L'intervention étudiée ici n'était pas un projet pilote autonome mené en dehors du système d'éducation publique; elle a été mise en oeuvre dans des collèges publics ordinaires par des spécialistes sociaux.

Troisièmement, les résultats éclairent la conception des dispositifs de soutien socio-émotionnel à l'adolescence. Les travaux existants dans les pays à faibles et moyens revenus portent souvent sur la petite enfance, l'enseignement primaire ou l'adolescence tardive (e.g., Dillon et al., 2017; Barrera-Osorio et al., 2024; Dam et al., 2025; Beaman et al., 2021; Krishnan and Krutikova, 2013). Nous enrichissons ces travaux en étudiant un soutien socio-émotionnel délivré au début du collège, une fenêtre développementale où le désengagement et le risque d'abandon scolaire augmentent, et en caractérisant la manière dont un dispositif large et de faible intensité, mis en oeuvre par le personnel scolaire public existant, se comporte lorsqu'un gouvernement le déploie à grande échelle.

La suite de l'article est organisée comme suit. La section 2 décrit le contexte des collèges publics marocains et l'intervention de soutien socio-émotionnel. La section 3 présente le dispositif d'étude et la stratégie d'analyse. La section 4 rapporte les résultats, la section 5 présente les analyses d'hétérogénéité et de robustesse, et la section 6 conclut.

2 Contexte et intervention

2.1 Les collèges publics marocains

L'enseignement public, placé sous la tutelle du Ministère de l'Éducation Nationale, du Préscolaire et des Sports, constitue la forme la plus répandue d'enseignement primaire et secondaire au Maroc. Bien que la part des inscriptions dans le privé ait augmenté ces

dernières années, en 2019, les établissements publics accueillent encore 83 pour cent des élèves du primaire et 89 pour cent des élèves du secondaire inférieur (The World Bank, 2025). Les taux de scolarisation sont élevés, avec des taux nets de scolarisation de 99,6 pour cent au niveau primaire et de 90,6 pour cent au niveau du secondaire inférieur en 2019 (The World Bank, 2025).

Entre 2018 et 2023, le taux d'achèvement du secondaire inférieur est passé de 64,5 à 74,2 pour cent (d'après les Indicateurs du développement mondial de la Banque mondiale). Cependant, ces taux élevés de scolarisation et d'achèvement masquent de très faibles niveaux de performance scolaire. Dans l'évaluation PIRLS de lecture de 2021, les élèves marocains de 4^{ème} année primaire se classaient avant-derniers (sur 57 pays), et plus de la moitié d'entre eux (59 pour cent) n'atteignaient pas le seuil minimum de compétence (Mullis et al., 2023). Dans l'évaluation TIMSS de mathématiques de 2023, les élèves marocains de 4^{ème} année primaire se classaient anté-penultième (sur 58 pays), et plus de la moitié (54 pour cent) n'atteignaient pas le seuil minimum de compétence (von Davier et al., 2024). Au niveau du secondaire, dans l'évaluation TIMSS de mathématiques de 2023, les élèves marocains de 2^{ème} année collégiale se classaient avant-derniers sur 42 pays, et 64 pour cent des élèves de 2^{ème} année collégiale se situaient en dessous du seuil international le plus bas en mathématiques (von Davier et al., 2024).

Nous avons mené cette étude dans le cadre d'un effort de réforme plus large visant à remédier aux faibles niveaux d'apprentissage dans les collèges publics marocains. Cette réforme est connue localement sous le nom de "programme des collèges pionniers (PCP)". L'intervention sur les compétences socio-émotionnelles en constitue une composante. Le programme a été lancé dans 232 collèges publics marocains en septembre 2024.¹ Le Ministère considère ce programme comme son intervention phare dans les collèges, et a commencé à l'étendre à 535 collèges supplémentaires au cours de l'année scolaire 2025-2026, couvrant environ 32 pour cent des élèves du pays. Les collèges restants devraient être atteints d'ici 2028.

2.2 L'intervention de soutien socio-émotionnel

La phase de diagnostic du programme s'est appuyée sur le modèle socio-écologique (Kilanowski, 2017) pour cartographier les multiples niveaux de facteurs susceptibles de contribuer à l'abandon scolaire au Maroc, qu'ils soient individuels, relationnels, communautaires ou sociétaux. L'intervention de soutien socio-émotionnel elle-même

1. Les 232 collèges se sont portés volontaires pour participer à la première année de la réforme. Le lancement dans les collèges fait suite au lancement du programme CP dans les écoles primaires l'année précédente. Nous évaluons l'effet global de (l'ensemble des interventions proposées par) la réforme dans les deux types d'établissements dans d'autres travaux de recherche.

s'ancre dans le cadre fondateur des compétences de vie (*life skills*) de l'Organisation mondiale de la Santé (World Health Organization, 1993, 2003) et dans le cadre CASEL pour l'apprentissage social et émotionnel (Durlak et al., 2015). Son lien hypothétique avec l'abandon scolaire se comprend au mieux à travers le modèle de participation-identification de Finn (1989), qui conçoit le décrochage scolaire comme un processus cumulatif alimenté par le déclin de l'engagement comportemental et émotionnel, fournissant ainsi une justification directe au ciblage du sentiment d'appartenance des élèves, de leur sentiment d'auto-efficacité et de leur relation avec un adulte de confiance.

Concrètement, l'intervention consiste à former et outiller des spécialistes sociaux pour qu'ils animent quatre ateliers de deux heures par an en petits groupes auprès des élèves de première et deuxième année du secondaire inférieur (1^{ère} et 2^{ème} année collégiale).² Ces ateliers visent à promouvoir la connaissance du rôle du spécialiste social, à construire une relation de confiance avec les élèves, et à développer les compétences intrapersonnelles et interpersonnelles dans le but d'améliorer le bien-être individuel, la capacité d'adaptation et le climat scolaire. Les ateliers sont également conçus pour accorder une attention particulière aux élèves à risque nécessitant un soutien individualisé et pour les orienter vers un suivi individuel.

Le membre du personnel chargé de mettre en oeuvre l'intervention de soutien socio-émotionnel est le spécialiste social de l'établissement. Douze spécialistes sociaux à travers le Maroc ont été sélectionnés pour recevoir une formation puis former à leur tour les spécialistes sociaux des collèges assignés aléatoirement à l'intervention. Les spécialistes sociaux ont reçu des diapositives de formation et des guides de séance détaillés, et les élèves ont reçu des livrets pour chaque atelier, ainsi qu'un cahier pour consigner leurs ressentis et émotions.

Chaque atelier porte sur un thème spécifique lié aux compétences intra- ou interpersonnelles. Le premier atelier se concentre sur la conscience de soi, abordant l'auto-évaluation (forces et faiblesses) et la mentalité de croissance (*growth-mindset*). Le deuxième atelier porte sur l'écoute active et la communication empathique, ainsi que sur la gestion des difficultés et conflits interpersonnels. Le troisième atelier porte sur l'identification de ses propres émotions et la régulation du stress. Le quatrième atelier porte sur la coopération, la collaboration et le sentiment d'appartenance des élèves.

2. Le spécialiste social est un fonctionnaire formé après un concours ouvert aux titulaires d'un diplôme en psychologie, philosophie ou sciences sociales. Sa mission consiste à assurer le bien-être des élèves à l'école et à identifier les élèves ayant besoin de soutien ou d'attention. Un spécialiste social est affecté à un établissement et y travaille au total 24 heures par semaine.

2.3 Théorie du changement

En termes de *besoins*, l'intervention visait à répondre à deux défis connexes communs à de nombreux pays à faibles et moyens revenus. Premièrement, des observations qualitatives initiales dans les collèges marocains ont mis en évidence un environnement largement peu propice au développement des compétences socio-émotionnelles, caractérisé par une forte prévalence du harcèlement et de l'intimidation, des relations conflictuelles entre élèves et enseignants, des sentiments d'insécurité, ainsi qu'une détresse et un inconfort généralisés parmi les élèves. Deuxièmement, bien que des personnels existent (en l'occurrence, des spécialistes sociaux), les services susceptibles de promouvoir les compétences socio-émotionnelles semblent inefficaces, avec un soutien limité aux élèves reposant uniquement sur la participation volontaire, des spécialistes sociaux affectés à des tâches administratives (ou uniquement aux cas les plus graves et disciplinaires), un manque de clarté quant aux responsabilités, et une absence d'outils et de formation. Pris ensemble, le contexte de l'étude semble révéler un besoin sérieux d'intervention et une forte probabilité qu'elle produise un contraste significatif avec le statu quo.

Les *intrants* offerts par l'intervention sont des sessions de formation pour les spécialistes sociaux travaillant dans les collèges ainsi que des guides de mise en oeuvre de l'intervention (décrits dans la section précédente), conformément à une instruction gouvernementale enjoignant les spécialistes à s'y conformer. Les autres ressources allouées aux collèges sélectionnés dans le cadre de la réforme en cours sont maintenues constantes; l'intervention est mise en oeuvre avec les ressources existantes, sans personnel supplémentaire ni hausse de rémunération pour les collèges bénéficiaires.

Les *produits* attendus sont que les spécialistes sociaux animent quatre ateliers de deux heures par an en petits groupes auprès des élèves de première et deuxième année du secondaire inférieur et que les élèves participent à ces ateliers. Les *résultats intermédiaires* attendus, les mécanismes et les résultats secondaires sont que le bien-être des élèves s'améliore, que leurs habitudes de travail progressent et que les élèves développent une pensée plus créative. L'intervention peut également entraîner des changements dans la participation des élèves aux activités extrascolaires; or, dans la mesure où ces activités ne favorisent pas les compétences académiques ou socio-émotionnelles, cet effet pourrait avoir des conséquences préjudiciables sur les variables en aval.

Les *résultats* attendus sont que les élèves améliorent leurs compétences interpersonnelles (ou sociales) et leurs compétences intrapersonnelles. Les premières renvoient à des effets positifs sur la perception des émotions et la pro-socialité. Les secondes renvoient à des effets positifs sur le sentiment de contrôle des élèves (notamment la mentalité de croissance

(*growth-mindset*), le lieu de contrôle et le sentiment d'auto-efficacité) ainsi que sur leur autorégulation et leur discipline (notamment la maîtrise de soi et la persévérance).³

Enfin, l'*impact* attendu est que les élèves améliorent leurs compétences académiques (en arabe, français, mathématiques et sciences) et deviennent moins susceptibles d'abandonner l'école.

3 Dispositif d'étude

3.1 Échantillonnage des collèges

Notre échantillon comprend 200 collèges publics. Nous avons constitué cet échantillon en deux étapes, à partir de données administratives couvrant l'ensemble des collèges publics, dont les 232 collèges pionniers du Maroc. Premièrement, en utilisant la méthodologie de "post-double selection" (PDS) (Belloni et al., 2014), 30 variables ont été identifiées comme prédictives soit des résultats aux tests, soit de la participation à la réforme CP. Deuxièmement, à partir de ces 30 variables sélectionnées, un appariement par plus proche voisin selon la distance de Mahalanobis (sans remise) a été utilisé pour identifier 100 paires de collèges pionniers très similaires les uns aux autres (en termes de distance de Mahalanobis).⁴

Le tableau 1 présente une analyse de la représentativité de l'échantillon, tant par rapport à la population des collèges marocains qu'aux autres collèges pionniers du pays, ainsi que des tests d'équilibre. Les colonnes (1) à (3) comparent les 200 collèges publics de l'échantillon de l'étude aux 2 344 collèges publics restants au Maroc.⁵ Les colonnes (4) à (6) présentent la représentativité au sein du programme de réforme en comparant les 200 collèges pionniers de l'échantillon de l'étude aux 32 autres collèges pionniers du Maroc. Dans l'ensemble, les collèges de l'échantillon de l'étude comptent légèrement plus d'enseignants, sont moins souvent ruraux, accueillent davantage d'élèves, et scolarisent des élèves ayant obtenu des résultats moyens légèrement plus élevés aux examens de fin de primaire et de collège.

3. Nous nous concentrons ici sur les résultats attendus issus de nos échanges avec le Ministère avant le pré-enregistrement de l'étude. Il s'agit d'objectifs larges, qui s'inscrivent également dans le processus plus général de la réforme des "collèges pionniers" et de son évaluation, présentée dans des travaux connexes (de Barros et al., 2026). Nous reconnaissons que les concepteurs de l'intervention ont pu, pour leur part, se concentrer sur un ensemble plus restreint de résultats plus proximaux.

4. Nous avons initialement envisagé d'assigner 100 collèges pionniers à l'intervention de soutien socio-émotionnel, dans le cadre d'un essai stratifié par paires. Cependant, comme précisé lors de la revue et de l'enregistrement de phase 1 du protocole d'étude, nous avons abandonné ce plan. La sous-section suivante décrit notre assignation aléatoire de 84 collèges pionniers à l'intervention, telle que pré-spécifiée.

5. Au moment de la recherche de collèges de contrôle appariés pour chacun des collèges pionniers, nous disposions de données sur 2 544 collèges publics.

3.2 Randomisation des collèges

Sur les 200 collèges pionniers, nous en avons assigné aléatoirement 84 à recevoir l'intervention de soutien socio-émotionnel (les 116 restants ne devant pas recevoir cette intervention). Plus précisément, pour mettre en oeuvre une stratégie de randomisation stratifiée, nous avons commencé par calculer une composante principale à partir de la note moyenne d'obtention du diplôme primaire, du nombre d'enseignants, du nombre d'élèves et de la proportion d'enseignantes dans les 200 collèges. Nous avons ensuite utilisé cette composante principale pour classer les collèges selon un schéma alterné de triplettes et de paires (84 groupes au total). Enfin, au sein de chaque groupe, nous avons assigné aléatoirement un collège à l'intervention de soutien socio-émotionnel.

Les colonnes (7) à (9) du tableau 1 présentent l'équilibre entre les deux groupes de collèges inclus dans cette étude. Les caractéristiques des collèges sont équilibrées entre les groupes traité et de contrôle, avec un léger déséquilibre dans la proportion d'élèves filles et l'âge moyen des élèves. Le test conjoint d'équilibre est marginal (colonne (9); $p = 0,101$). Fait rassurant, les tests d'équilibre au niveau des élèves présentés ci-après ne montrent aucune différence systématique entre les deux groupes.⁶

3.3 Échantillonnage des élèves

L'unité d'analyse de l'étude est l'élève. Dans chacun des collèges de l'étude, les enquêteurs disposaient d'une liste de priorité d'élèves sélectionnés aléatoirement pour chaque niveau. Ces listes ont été constituées avant le début de la collecte de données de début d'année, à partir des registres d'inscription. Elles permettaient des remplacements aléatoires en cas d'absence d'élèves le jour de l'évaluation de début d'année.

Les élèves ont été échantillonnés pour passer des évaluations dans une seule matière (Arabe, Français, mathématiques ou sciences). Pour les évaluations écrites en Arabe et en Français, ainsi que pour les évaluations orales individuelles en Arabe et en Français et les entretiens portant sur les compétences socio-émotionnelles, nous avons sous-échantillonné jusqu'à 12 élèves par collège (6 élèves de 1^{ère} année et 6 élèves de 2^{ème} année).⁷ Pour les évaluations écrites en mathématiques et en sciences, nous avons sous-échantillonné jusqu'à 18 élèves par collège (tous en 1^{ère} année).⁸

6. Tous les tests F conjoints des tableaux 1, 2 et A1 reportent des valeurs p d'inférence par randomisation, suivant Kerwin et al. (2024).

7. Nous avons également échantillonné et évalué jusqu'à 6 élèves de 3^{ème} année, mais étant donné que l'intervention sur les compétences socio-émotionnelles ne concerne que les niveaux inférieurs, nous les excluons de nos données. Nous avons également précisé cette décision lors de la revue de phase 1 de notre plan d'analyse.

8. Pour ces deux matières, nous n'avons pas inclus les élèves de 2^{ème} ou 3^{ème} année, car la réforme CP n'incluait pas encore d'interventions en mathématiques et en sciences pour ces niveaux.

Plus précisément, les élèves de chaque niveau et filière (APIC ou ASCG) ont été répartis aléatoirement en quatre groupes, un par matière.⁹ Au sein de chaque groupe d'élèves échantillonnés pour une matière, les élèves ont été classés aléatoirement dans une liste ordonnée. Cette liste ordonnée aléatoirement garantissait que chaque niveau et matière disposait du nombre requis d'élèves échantillonnés pour les évaluations et les entretiens, et que le nombre d'élèves de chaque filière sur la liste était proportionnel au nombre d'élèves inscrits dans cette filière pour cette matière. Ainsi, une liste de priorité aléatoire d'élèves a été générée, comprenant des élèves de "haute" priorité que les enquêteurs devaient chercher à évaluer/interviewer en premier, avant de descendre dans la liste vers les élèves de "basse" priorité.

L'échantillon effectif de l'étude comprend 11 132 élèves (6 483 dans le groupe de contrôle et 4 649 dans le groupe traité). Ce sont tous les élèves ayant passé l'évaluation de début d'année et dont nous avons cherché à assurer le suivi pour l'évaluation de fin d'année. Parmi ces élèves, 2 138 ont été évalués en Arabe, 2 148 en Français, 3 428 en mathématiques et 3 418 en sciences.

Le tableau 2 examine l'équilibre en début d'année parmi les élèves échantillonnés, en termes de résultats d'apprentissage, de compétences socio-émotionnelles et de caractéristiques individuelles, entre les collèges du groupe de contrôle et ceux du groupe traité. Comme attendu, compte tenu de l'assignation aléatoire de l'intervention de soutien socio-émotionnel et de l'échantillonnage aléatoire décrit ci-dessus, les élèves des collèges du groupe de contrôle sont généralement similaires à ceux des collèges du groupe traité, et les tests F conjoints confirment l'absence de différences systématiques entre les deux groupes.¹⁰ Individuellement, deux caractéristiques de début d'année diffèrent au seuil de 10 pour cent : les élèves du groupe traité obtiennent un score supérieur de 0,083 é.-t. à l'évaluation d'arabe en début d'année et sont légèrement plus jeunes. Notre spécification d'intention-de-traiter tient compte des caractéristiques des élèves et de leurs résultats de début d'année (section 3.6), et aucune de ces différences ne survivrait à la correction pour tests d'hypothèses multiples que nous appliquons à nos effets de traitement, de sorte que ces légers déséquilibres n'affectent pas nos conclusions. L'attrition entre le début et la fin de l'année a été faible, à environ 5 pour cent au total, avec un léger déséquilibre d'environ un point de pourcentage entre les groupes traité et de contrôle ; la section 5.2 montre que nos estimations sont robustes à cette attrition différentielle.

9. APIC désigne "Année Secondaire Collégiale Parcours International" et ASCG désigne "Année Secondaire Collégiale Général". Il s'agit des deux filières académiques dans lesquelles les élèves du collège sont orientés, APIC désignant la filière à vocation internationale et ASCG la filière générale.

10. Le tableau A1 examine l'équilibre en début d'année parmi l'échantillon sans attrition de 10 526 élèves et confirme que les élèves sans attrition dans les deux groupes randomisés présentent des caractéristiques similaires à début d'année.

3.4 Mesure des résultats

Notre cadre de mesure s'aligne étroitement sur la théorie du changement décrite ci-dessus.¹¹ Les résultats principaux pré-spécifiés correspondaient à deux familles de compétences socio-émotionnelles. L'indice de compétences interpersonnelles combine la pro-socialité et la perception des émotions. La pro-socialité est mesurée à l'aide de la sous-échelle pro-sociale du Questionnaire sur les forces et les difficultés (Goodman, 1997). La perception des émotions est mesurée à l'aide de l'échelle PAGE (Perceiving AI-Generated Emotions), qui demande aux élèves de reconnaître des émotions dans des expressions faciales (Weidmann and Xu, 2024).

L'indice de compétences intrapersonnelles combine le sentiment de contrôle et l'autorégulation/discipline. Le sentiment de contrôle inclut la mentalité de croissance (*growth-mindset*), le lieu de contrôle et le sentiment d'auto-efficacité. L'autorégulation et la discipline comprennent la discipline de travail et la persévérance, ainsi que la maîtrise de soi (Duckworth and Quinn, 2009; Huillery et al., 2025). Ces mesures capturent les croyances des élèves quant à leur propre capacité d'action, leur aptitude à persévérer face aux difficultés, et leur capacité à gérer leur attention et leur comportement.

Nous avons également pré-spécifié des résultats induits. Les compétences académiques sont mesurées par des évaluations écrites en arabe, français, mathématiques et sciences. En arabe et en français, les élèves ont également passé des évaluations orales individuelles. Le résultat d'apprentissage principal agrège les scores standardisés entre matières. Les résultats de progression scolaire sont mesurés à partir de registres administratifs et comprennent l'abandon scolaire ou l'exclusion en fin d'année scolaire, la non-réinscription au début de l'année scolaire suivante, le redoublement et l'exclusion.

Des résultats additionnels captent les mécanismes potentiels et des domaines de développement plus larges. Ceux-ci comprennent la créativité, le climat scolaire et le bien-être, et les habitudes de travail. La créativité est mesurée à l'aide de tâches inspirées des Tests de pensée créative de Torrance (Torrance, 1968). Le climat scolaire et le bien-être combinent des mesures du sentiment d'appartenance, du harcèlement et du stress perçu (Cohen et al., 1983). Les habitudes de travail sont mesurées par le fait que les élèves déclarent consacrer plus de 30 minutes par jour aux devoirs après l'école.

11. Comme dans cette description de la théorie du changement, et dans notre pré-enregistrement de familles de résultats larges, nous reconnaissons que les concepteurs de l'intervention auraient pu choisir, pour leur part, un cadre de mesure composé d'un ensemble plus restreint de mesures de résultats proximaux correspondant plus étroitement au contenu des ateliers. Bailey et al. (2020) avertissent que le "surajustement" (*overalignment*) entre le contenu d'une intervention et les mesures de résultats constitue une préoccupation répandue dans les évaluations en éducation; pour des données probantes connexes, voir de Barros and Lubozha (2026).

Enfin, nous mesurons l'adoption et la mise en oeuvre à partir des déclarations des élèves en fin d'année. Les principaux résultats d'exposition sont la participation des élèves aux ateliers socio-émotionnels et leur connaissance du spécialiste social de l'établissement. Nous mesurons également l'exposition aux autres composantes du programme des collèges pionniers (PCP), notamment la remédiation ciblée, l'enseignement explicite et le tutorat.

Les réponses au niveau des items sont agrégées à l'aide de modèles de théorie de réponse à l'item (TRI) choisis en fonction de la structure de chaque instrument (Jacob and Rothstein, 2016). Les items binaires sont scorés à l'aide de modèles logistiques à deux paramètres, et les items catégoriels ordonnés à l'aide de modèles de réponse graduée. Lorsque plusieurs scores latents sont combinés en un indice de famille, nous utilisons des moyennes pondérées par l'inverse de la covariance. Les scores de fin d'année sont standardisés par rapport à la distribution du groupe de contrôle en fin d'année. L'annexe B fournit des détails sur la mesure et le scoring.¹²

Le tableau en annexe B1 rapporte la fiabilité conditionnelle moyenne de chaque instrument en fin d'année. La précision est élevée pour les évaluations académiques (0,78–0,90) et pour la plupart des composantes socio-émotionnelles, la perception des émotions constituant la principale exception (0,51); comme chaque instrument entre dans l'analyse en tant que résultat et non en tant que variable explicative, cette erreur de mesure gonfle les erreurs-types mais ne biaise pas les estimations d'effet du traitement.

3.5 Exposition au programme

Le panneau A du tableau 3 montre que l'intervention a considérablement accru l'exposition des élèves au soutien socio-émotionnel. Dans les collèges du groupe de contrôle, 21,7 pour cent des élèves déclaraient avoir participé à des ateliers socio-émotionnels. L'assignation au groupe traité a augmenté cette proportion de 43,2 points de pourcentage. L'intervention a également accru la connaissance des élèves de leur spécialiste social de 16,9 points de pourcentage, par rapport à une moyenne du groupe de contrôle de 43,5 pour cent. Les deux estimations sont statistiquement significatives au seuil de un pour cent.¹³

Le panneau B du tableau 3 montre quant à lui que les collèges des groupes traité et de contrôle présentaient une exposition similaire aux autres composantes du programme des collèges pionniers (PCP). Nous ne trouvons pas de différences significatives dans la réception déclarée par les élèves de la remédiation ciblée, de l'enseignement explicite ou

12. Pour des articles connexes sur la mesure des compétences socio-émotionnelles dans les pays à faibles revenus, voir Laajaj and Macours (2021) et Danon et al. (2024).

13. Le panneau A des tableaux A2 et A3 montre que l'intervention a conduit à une augmentation similaire de l'exposition au soutien socio-émotionnel parmi les élèves à risque d'abandon scolaire et parmi les élèves filles, respectivement.

du tutorat. Cette absence d'effet conforte l'interprétation des estimations ci-après comme reflétant l'effet de la composante de soutien socio-émotionnel plutôt qu'une exposition différentielle à l'ensemble du programme de réforme.¹⁴

3.6 Stratégie empirique

Notre stratégie d'identification repose sur l'assignation aléatoire des collèges aux deux groupes expérimentaux au sein de chaque strate de randomisation. Nous estimons l'effet d'intention-de-traiter (ITT) de la composante de soutien socio-émotionnel sur les résultats d'intérêt en comparant les collèges assignés aléatoirement à recevoir l'intervention de soutien socio-émotionnel à ceux assignés aléatoirement à ne pas la recevoir. Pour tous les résultats, nous utilisons la spécification empirique suivante :

$$Y_{igr}^{t=1} = \lambda_r + \beta_1 T_{sr} + \delta' X_{igr}^{t=0} + \epsilon_{igr} \quad (1)$$

Ici, $Y_{igr}^{t=1}$ est le résultat de fin d'année pour l'élève i de niveau g dans le collège s et la strate de randomisation r . Le terme d'effets fixes λ_r absorbe les effets de strate de randomisation ; pour les résultats issus d'enquêtes, nous absorbons en outre les effets fixes matière, et pour les résultats d'abandon scolaire, nous n'absorbons que les effets niveau-par-strate. T_{sr} est l'indicateur de traitement indiquant l'assignation aléatoire d'un collège à l'intervention de soutien socio-émotionnel, et ϵ_{igr} est le résidu. Le coefficient d'intérêt est β_1 , qui capture l'effet d'intention-de-traiter de l'assignation à l'intervention de soutien socio-émotionnel parmi les collèges du programme des collèges pionniers (PCP). Pour accroître la précision, toutes les spécifications incluent $X_{igr}^{t=0}$ comme covariables. Mesurées à début d'année ($t = 0$), $X_{igr}^{t=0}$ est un vecteur de contrôles de début d'année sélectionnés par une procédure Lasso à sélection post-double (PDS) sur les caractéristiques des élèves et des collèges, en neutralisant les effets fixes de strate et, lorsqu'elle est disponible, $Y_{igr}^{t=0}$ (le résultat d'intérêt de l'élève à début d'année).¹⁵ Toutes les inférences sont clusterisées au niveau du collège.

Nous examinons également les effets parmi deux sous-groupes pré-spécifiés : les élèves identifiés comme étant à risque d'abandon scolaire et les élèves filles. Comme pré-spécifié, nous interagissons l'indicateur de traitement avec l'indicateur de sous-groupe et reportons l'effet global de chaque sous-groupe.

14. Le panneau B des tableaux A2 et A3 montre que l'intervention n'a pas entraîné d'exposition différentielle aux composantes du programme de réforme CP plus large parmi les élèves à risque d'abandon scolaire et parmi les élèves filles respectivement.

15. Nous remercions Jacobus Cilliers de nous avoir signalé Cilliers et al. (2024) et recommandé cette approche. Suivant Cilliers et al. (2024), nous utilisons le paramètre de pénalité "plug-in" par défaut de la commande `pdslasso` de Stata (sans validation croisée).

Nous corrigeons pour les tests d'hypothèses multiples (THM) et reportons des valeurs q affinées corrigées par le taux de fausses découvertes (TFD), calculées à l'aide de la procédure de Anderson (2008), suivant l'approche par familles de résultats hiérarchisées de Vivalt et al. (2024). La famille principale de résultats (K0) comprend les compétences interpersonnelles et intrapersonnelles. La deuxième famille de résultats (K1) comprend les compétences académiques et l'abandon scolaire. La troisième famille (K2) comprend la créativité, le climat scolaire et le bien-être, et les habitudes de travail. Les tests au niveau des composantes et des sous-groupes sont corrigés au sein des familles de tests pré-spécifiées pertinentes. L'annexe C fournit des détails supplémentaires sur la procédure de tests d'hypothèses multiples pré-spécifiée.

4 Résultats

4.1 Compétences socio-émotionnelles

Le tableau 4 rapporte les effets sur les deux résultats socio-émotionnels primaires. Dans le panneau A, nous constatons que l'assignation à l'intervention a conduit à une variation faible et statistiquement non significative de l'indice de compétences interpersonnelles de +0,029 é.-t.. Lorsque nous examinons les composantes individuelles de cet indice, l'estimation pour la pro-socialité est de -0,007 é.-t. et celle pour la perception des émotions de +0,058 é.-t. L'estimation pour la perception des émotions est statistiquement significative au seuil de 5 pour cent avant correction pour tests d'hypothèses multiples mais ne reste pas statistiquement significative après correction ; la pro-socialité n'est pas statistiquement significative aux seuils conventionnels.

Dans le panneau B, nous constatons que l'assignation à l'intervention n'a produit aucun changement détectable dans l'indice de compétences intrapersonnelles. Lorsque nous nous concentrons sur les composantes individuelles de cet indice, le sentiment de contrôle a augmenté de 0,045 é.-t. et l'autorégulation et la discipline ont diminué de 0,027 é.-t., aucun de ces résultats n'étant statistiquement significatif.

Ces résultats indiquent que, malgré des augmentations de l'exposition déclarée aux ateliers et aux spécialistes sociaux, l'intervention n'a pas produit d'améliorations moyennes détectables des compétences socio-émotionnelles des élèves au cours de la première année. Un sous-domaine fait exception : parmi les mesures d'autorégulation et de discipline, la maîtrise de soi des élèves a diminué de 0,097 é.-t. (tableau en annexe A4), une estimation statistiquement significative qui survit à la correction pour tests d'hypothèses multiples ($q = 0,003$). Les autres sous-domaines de compétences socio-émotionnelles sous-jacents

suivent le schéma plus général : les estimations sont de faible magnitude et ne restent pas statistiquement significatives après correction pour tests d'hypothèses multiples.

4.2 Apprentissage

Le tableau 5 rapporte les effets sur les compétences académiques et l'abandon scolaire. Dans le panneau A, nous constatons que l'assignation à l'intervention a conduit à une augmentation faible et statistiquement non significative des scores aux tests de 0,034 é.-t. Lorsque nous nous concentrons sur les matières individuelles, les scores ont augmenté de 0,023 é.-t. en arabe, 0,070 é.-t. en français et 0,067 é.-t. en sciences, tandis que les scores en mathématiques ont diminué de 0,033 é.-t.; aucune de ces estimations par matière n'est statistiquement significative après correction pour tests d'hypothèses multiples.

Au sein des sciences, l'augmentation estimée de 0,123 é.-t. en physique-chimie est statistiquement significative et le reste après correction ($q = 0,016$), tandis que l'augmentation de 0,033 é.-t. en sciences de la vie ne l'est pas. En décomposant davantage l'évaluation en sciences (tableau en annexe A5), les gains apparaissent en physique (0,148 é.-t.), en chimie (0,135 é.-t.) et en biologie (0,095 é.-t.), chacun survivant à la correction pour tests d'hypothèses multiples.

Parmi les sous-domaines des langues (tableau en annexe A5), les estimations positives se concentrent sur la production écrite. L'assignation à l'intervention a augmenté les scores de production écrite de 0,124 é.-t. en arabe et de 0,132 é.-t. en français, chacun statistiquement significatif au seuil de 1 pour cent avant correction et d'ampleur similaire dans les deux évaluations de langue; l'estimation en arabe reste statistiquement significative après correction pour tests d'hypothèses multiples ($q = 0,044$), et l'estimation en français reste significative au seuil de 10 pour cent après correction ($q = 0,095$). L'expression orale en arabe présente un gain comparable (0,118 é.-t.) qui reste également statistiquement significatif après correction ($q = 0,044$), mais l'estimation correspondante en français est faible et statistiquement non significative (0,047 é.-t.), de sorte que nous interprétons le résultat oral avec davantage de prudence. Ces estimations portent sur les composantes de production (à l'écrit et, en arabe, à l'oral) plutôt que sur la compréhension de l'écrit, de nature réceptive, un schéma cohérent avec des ateliers centrés sur la communication et l'expression mais aussi avec un engagement accru des élèves dans les tâches de production à réponse ouverte.

4.3 Abandon scolaire et redoublement

Dans le panneau B du tableau 5, nous ne trouvons aucune preuve que l'intervention a réduit l'abandon scolaire en fin d'année. L'effet estimé sur notre mesure principale d'abandon scolaire est une hausse de 0,1 point de pourcentage de la proportion d'élèves n'ayant pas poursuivi l'année suivante, que ce soit volontairement ou en raison d'une exclusion, par rapport à une moyenne du groupe de contrôle de 2,6 pour cent. Les estimations pour l'abandon scolaire seul, la non-réinscription l'année scolaire suivante et l'exclusion sont également faibles et statistiquement non significatives après correction pour tests d'hypothèses multiples. L'assignation à l'intervention a réduit le redoublement de 1,7 point de pourcentage, par rapport à une moyenne du groupe de contrôle de 9,6 pour cent, et cette estimation reste statistiquement significative au seuil de 5 pour cent après correction pour tests d'hypothèses multiples.

4.4 Résultats additionnels

Le tableau 6 examine les résultats additionnels. Nous constatons que l'assignation à l'intervention n'a pas conduit à des améliorations détectables de la créativité, du climat scolaire et du bien-être, ou des habitudes de travail. La créativité a diminué de 0,079 é.-t., tandis que le climat scolaire et le bien-être ont augmenté de 0,014 é.-t., et la proportion d'élèves consacrant plus de 30 minutes par jour aux devoirs après l'école a augmenté de 0,1 point de pourcentage. L'estimation pour la créativité est statistiquement significative au seuil de 10 pour cent avant correction pour tests d'hypothèses multiples mais ne reste pas statistiquement significative après correction; les autres estimations ne sont pas statistiquement significatives aux seuils conventionnels. Ces estimations n'appuient pas l'hypothèse selon laquelle l'intervention a conduit à des changements mesurables du bien-être, de la créativité ou des habitudes de travail en dehors de l'école.

5 Hétérogénéité et robustesse

5.1 Effets par sous-groupes

Le tableau 7 rapporte les analyses de sous-groupes pré-spécifiées pour les élèves identifiés comme étant à risque d'abandon scolaire au début de la période d'étude et pour les élèves filles. Dans le panneau A, nous constatons que l'assignation à l'intervention a augmenté les compétences intrapersonnelles des élèves à risque de 0,144 é.-t. Cette estimation est statistiquement significative au seuil de 10 pour cent avant correction pour tests

d'hypothèses multiples mais ne reste pas statistiquement significative après correction ($q = 0,281$). L'estimation correspondante pour les compétences interpersonnelles est de 0,004 é.-t. et n'est pas statistiquement significative. Nous ne trouvons aucune preuve d'effets sur l'apprentissage, l'abandon scolaire, la créativité, le climat scolaire et le bien-être, ou les habitudes de travail des élèves à risque.

Dans le panneau B, nous ne trouvons aucune preuve d'effets pour les élèves filles. L'assignation à l'intervention a augmenté les compétences interpersonnelles de 0,043 é.-t., diminué les compétences intrapersonnelles de 0,014 é.-t. et augmenté l'apprentissage de 0,064 é.-t. L'effet estimé sur la mesure globale d'abandon scolaire est essentiellement nul. L'estimation pour l'apprentissage est statistiquement significative au seuil de 10 pour cent avant correction pour tests d'hypothèses multiples mais ne reste pas statistiquement significative après correction; les autres estimations sont faibles et statistiquement non significatives, et les estimations sur les résultats additionnels sont pareillement faibles et non significatives.¹⁶

5.2 Robustesse et sensibilité

Étant donné que l'échantillon des élèves initialement inscrits varie selon le mois de sa constitution, nous évaluons la robustesse des résultats relatifs à l'abandon scolaire à des définitions alternatives de l'échantillon d'inscription. Le tableau A7 montre que les effets estimés sur l'abandon scolaire sont stables selon différentes définitions de l'échantillon d'inscription.

Nous examinons également si le biais de désirabilité sociale affecte les estimations des effets sur les compétences socio-émotionnelles, suivant Dhar et al. (2022). Le tableau A8 ne montre aucune preuve que les effets sur les compétences socio-émotionnelles diffèrent de manière significative entre les élèves ayant des scores de désirabilité sociale à début d'année au-dessus et en dessous de la médiane. La différence d'effets est de 0,022 é.-t. pour les compétences interpersonnelles et de $-0,056$ é.-t. pour les compétences intrapersonnelles, et

16. En plus de l'investigation pré-enregistrée des effets du programme parmi les élèves à risque et les élèves filles, nous avons mené une analyse exploratoire à l'aide de forêts causales (Wager and Athey, 2018) pour déterminer si les effets de traitement variaient selon un ensemble plus large de caractéristiques de début d'année des élèves et des collègues, en évaluant la règle de priorisation implicite à l'aide de l'effet de traitement moyen pondéré par le rang (RATE) de Yadlowsky et al. (2025). Le seul résultat dont le RATE différait significativement de zéro était les compétences intrapersonnelles, et l'estimation était négative (RATE = $-0,095$, erreur-type 0,029) : les élèves que la forêt classe comme les plus susceptibles de bénéficier de l'intervention présentaient des effets de traitement inférieurs à la moyenne. Ceci concorde sur le fond avec l'analyse de sous-groupes pré-spécifiée, même si les estimations ponctuelles diffèrent de signe : dans cette dernière, l'estimation intrapersonnelle pour les élèves à risque est positive mais ne survit pas à la correction (0,144 é.-t., $q = 0,281$), et la forêt ne parvient pas non plus à isoler un groupe plus réactif, de sorte qu'aucune des deux n'établit l'existence d'un sous-groupe aux effets supérieurs à la moyenne. Nous n'interprétons donc pas cette analyse comme identifiant un sous-groupe bénéficiaire, dans un sens comme dans l'autre.

aucune des deux estimations n'est statistiquement significative. Nous prenons cela comme une preuve que, plus largement, nos résultats ne sont vraisemblablement pas affectés par un biais de désirabilité sociale dans les réponses des élèves aux enquêtes.

Enfin, nous évaluons si le léger déséquilibre d'attrition entre les collègues des groupes traité et de contrôle pourrait expliquer nos résultats. Le tableau en annexe A9 rapporte les effets du traitement sur les deux indices socio-émotionnels et sur l'indice global des scores aux tests selon trois approches de l'attrition : les bornes de Lee (2009), les bornes resserrées de Behaghel et al. (2015) fondées sur l'effort de suivi des enquêtes, et une pondération par probabilité inverse à partir des covariables de début d'année. Dans les trois approches, les estimations sont proches de celles de notre analyse principale, ce que nous prenons comme une preuve que, plus largement, les résultats que nous présentons ne sont pas affectés par une attrition différentielle.

6 Conclusion

Cet article a étudié une intervention de soutien socio-émotionnel délivrée par le gouvernement dans les collèges publics marocains. L'intervention a formé et outillé les spécialistes sociaux existants pour animer des ateliers en petits groupes auprès des élèves de 1^{ère} et 2^{ème} année collégiale. Dans un essai randomisé par cluster portant sur 200 collèges, l'assignation à l'intervention a augmenté la participation déclarée des élèves aux ateliers de 43,2 points de pourcentage et leur connaissance du spécialiste social de l'établissement de 16,9 points de pourcentage.

Malgré cette augmentation de l'exposition déclarée, nous ne trouvons aucun effet moyen détectable sur les résultats principaux pré-spécifiés (compétences interpersonnelles et intrapersonnelles), ni sur l'apprentissage global, l'abandon scolaire, la créativité, le bien-être ou les habitudes de travail. Si quoi que ce soit, la maîtrise de soi des élèves a diminué de 0,097 é.-t. Néanmoins, des améliorations apparaissent dans deux ensembles de résultats secondaires : le redoublement a diminué de 1,7 point de pourcentage ($q = 0,032$), et les résultats ont augmenté dans certains sous-domaines des sciences et des langues, à savoir la physique, la chimie et la biologie, la production écrite en arabe et en français, et l'expression orale en arabe. Toutes ces estimations découlent d'une hiérarchie pré-spécifiée de tests d'hypothèses qui contrôle le taux de fausses découvertes.

Nous lisons la baisse du redoublement comme cohérente avec le renforcement, par le programme, du rôle du spécialiste social dans le soutien aux élèves susceptibles de redoubler, le type d'attention individualisée que l'intervention était censée encourager. Puisque l'intervention ne ciblait pas les contenus disciplinaires, nous traitons les gains

en physique, en chimie et en biologie comme des estimations inexplicables, qui appellent des recherches supplémentaires. Les gains langagiers sont plus interprétables : les scores de production écrite augmentent de 0,124 é.-t. en arabe et de 0,132 é.-t. en français, et l'expression orale en arabe de 0,118 é.-t., se concentrant sur les composantes de production du langage plutôt que sur la compréhension de l'écrit, de nature réceptive, un schéma cohérent avec l'accent mis par les ateliers sur la communication et l'expression ; nous les interprétons comme une preuve que l'intervention a pu renforcer les compétences langagières productives des élèves. Nous n'avons pas d'explication comparable pour le déclin de la maîtrise de soi, que nous laissons comme une énigme pour de futurs travaux.

Nous concluons avec trois réserves. Premièrement, chaque collègue de l'étude participe au programme des collèges pionniers plus large ; notre dispositif identifie donc l'effet marginal de la composante socio-émotionnelle au sein d'un système en cours de réforme. Il ne capture pas l'effet d'une programmation socio-émotionnelle par rapport à un statu quo non réformé (et il ne renseigne pas non plus sur l'efficacité globale de la réforme). Si d'autres composantes de la réforme répondaient déjà à certains des besoins ciblés par les ateliers, la contribution marginale des ateliers observée ici pourrait sous-estimer ce qu'un soutien socio-émotionnel apporterait dans des collèges hors réforme ; si la réforme a au contraire créé des conditions favorables au fonctionnement des ateliers, elle pourrait la surestimer.

Deuxièmement, les quatre ateliers par an reflètent la dose que le Ministère a choisi de déployer à grande échelle ; les méta-analyses des programmes socio-émotionnels universels trouvent généralement des effets moyens détectables surtout parmi les programmes comportant un plus grand nombre de séances de pratique explicite des compétences (Cipriano et al., 2023). Nous n'avons par ailleurs qu'une visibilité limitée sur la fidélité avec laquelle cette dose a été délivrée, puisque nous observons la participation déclarée des élèves mais pas la qualité des séances effectivement tenues ; un résultat nul sur les compétences socio-émotionnelles ciblées peut donc refléter soit les effets modestes d'ateliers bien menés à cette dose, soit des insuffisances dans leur mise en oeuvre, et nos données ne permettent pas de distinguer les deux. Des travaux futurs pourraient examiner la relation dose-réponse de l'intervention, y compris une version des ateliers à dose plus élevée ; une recherche qualitative pourrait évaluer la manière dont les séances sont conduites en pratique.

Troisièmement, nous avons mesuré les compétences socio-émotionnelles de manière large, au sein d'un cadre de résultats pré-enregistré qui tient également compte des efforts de réforme plus larges du Ministère (de Barros et al., 2026), plutôt qu'au moyen d'un ensemble restreint de mesures proximales adaptées au contenu des ateliers. Cette largeur est délibérée : elle teste si les gains de compétences se généralisent à des résultats d'intérêt général, prémunit contre des effets mécaniques dans des instruments construits autour de

l'intervention, et maintient la comparabilité de nos mesures avec la littérature. Son coût est que certains instruments se situent à distance de ce que les ateliers entraînent le plus directement.¹⁷ Le calendrier de mesure ajoute à cette distance, puisque nous observons les élèves à la fin de la même année scolaire, lorsque les compétences socio-émotionnelles peuvent refléter une acquisition initiale plutôt que le changement qui ne deviendrait visible qu'à des suivis de plus long terme. Nous lisons donc ces résultats comme étant les plus informatifs pour les grands domaines de compétences que notre cadre était conçu pour capturer, et moins informatifs pour les compétences plus proximales que les ateliers entraînent le plus directement.

Dans l'ensemble, l'étude montre qu'un gouvernement peut mobiliser le personnel scolaire existant pour modifier l'exposition des élèves au soutien socio-émotionnel dans les conditions ordinaires de l'école publique, en augmentant la participation aux ateliers et la familiarité avec le spécialiste social sans personnel supplémentaire ni hausse de rémunération. À la dose et sous la forme déployées ici, ce contact n'a pas amélioré de manière détectable les compétences socio-émotionnelles que les ateliers ciblaient, même s'il a coïncidé avec des améliorations sur deux ensembles de résultats induits pré-spécifiés, le redoublement et les résultats dans certains sous-domaines des sciences et des langues, qui survivent à la correction pour tests d'hypothèses multiples. Atteindre les élèves avec un soutien socio-émotionnel par le canal des collègues ordinaires et du personnel existant semble ainsi faisable, mais traduire ce contact en gains sur les compétences socio-émotionnelles ciblées pourrait nécessiter un dispositif plus intensif, et potentiellement plus ciblé, que celui testé ici.

17. Par exemple, la mesure de perception des émotions demande aux élèves de lire les émotions d'autrui plutôt que d'identifier les leurs, ce que les ateliers considèrent comme le fondement de la régulation; le sentiment d'auto-efficacité tient lieu de l'estime de soi, plus large, que les séances cultivent; et plusieurs compétences ciblées, dont l'écoute active, l'empathie et la résolution de conflits, ne sont pas mesurées du tout, tandis que la relation entre les élèves et le spécialiste social n'est saisie que par le fait qu'un élève connaît ou non le spécialiste.

Tableau 1 : Échantillon d'établissements, représentativité et tests d'équilibre

	Représentativité (globale)						Représentativité (CP)			Tests d'équilibre								
	Hors étude		Étude		Différence		Autres CP		CP de l'étude		Différence		Contrôle		Traitement		Différence	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)									
Nombre d'enseignants	29.61 [15.20]	32.38 [14.35]	2.77*** (1.06)	27.12 [12.53]	32.38 [14.35]	5.25** (2.41)	32.88 [14.25]	31.69 [14.56]	-1.10 (1.08)									
Rural (%)	49.27 [50.01]	37.00 [48.40]	-12.27*** (3.57)	65.62 [48.26]	37.00 [48.40]	-28.62*** (9.10)	36.21 [48.27]	38.10 [48.85]	1.76 (6.69)									
Effectif total	774.30 [457.18]	913.21 [466.13]	138.90*** (34.22)	834.62 [457.39]	913.21 [466.13]	78.58 (86.48)	941.09 [487.00]	874.70 [435.60]	-60.21 (41.64)									
Élèves filles (%)	47.30 [4.90]	47.09 [2.90]	-0.215 (0.229)	48.04 [3.53]	47.09 [2.90]	-0.948 (0.650)	47.46 [2.99]	46.58 [2.72]	-0.903** (0.409)									
Âge	14.60 [0.365]	13.94 [0.218]	-0.665*** (0.017)	14.58 [0.260]	13.94 [0.218]	-0.642*** (0.048)	13.96 [0.203]	13.91 [0.235]	-0.053* (0.028)									
Score à l'examen de fin de primaire	6.85 [0.285]	6.94 [0.280]	0.093*** (0.021)	6.79 [0.229]	6.94 [0.280]	0.150*** (0.045)	6.94 [0.256]	6.94 [0.311]	0.001 (0.037)									
Score à l'examen de fin de collège	8.95 [1.46]	9.14 [1.36]	0.189* (0.101)	8.75 [1.09]	9.14 [1.36]	0.386* (0.214)	9.07 [1.31]	9.22 [1.43]	0.143 (0.174)									
Nombre d'établissements	2344	200	2544	32	200	232	116	84	200									
Test F joint (p-value)										0.101								

Notes. Ce tableau présente l'échantillon d'établissements de l'étude. "Étude" désigne les 200 collèges inclus dans l'échantillon effectif de l'étude. "Hors étude" désigne l'ensemble des autres collèges publics du Maroc. "CP de l'étude" désigne les 200 collèges pionniers inclus dans l'étude. "Autres CP" désigne les 32 collèges pionniers restants dans le pays. "Contrôle" désigne les 116 établissements non assignés à recevoir l'intervention d'appui socio-émotionnel, et "Traitement" désigne les 84 établissements assignés à recevoir cette intervention. "Différence" indique la différence ajustée par régression. Les écarts-types sont indiqués entre crochets ; les erreurs standard sont indiquées entre parenthèses et sont robustes à l'hétéroscédasticité, l'unité d'observation étant l'établissement. * p < 0,10, ** p < 0,05, *** p < 0,01.

Tableau 2 : Tests d'équilibre

	Nombre d'observations		Test d'équilibre	
	Contrôle	Traitement	Moyenne du groupe contrôle	
	(1)	(2)	(3)	(4)
Panneau A : Compétences socio-émotionnelles				
Compétences interpersonnelles	2509	1777	-0.368 [1.22]	-0.008 (0.047)
Perception des émotions	2509	1777	-0.288 [0.952]	-0.007 (0.036)
Compétences intrapersonnelles	2509	1777	1.90 [0.796]	0.001 (0.027)
Sentiment de contrôle	2509	1777	1.29 [0.675]	0.002 (0.022)
Auto-régulation et discipline	2509	1777	2.07 [1.00]	0.000 (0.035)
Test F joint (p-value)				0.997
Panneau B : Compétences académiques et créativité				
Apprentissage (scores aux tests empilés)	6483	4649	-0.583 [1.01]	0.037 (0.036)
Arabe	1256	882	-0.586 [1.06]	0.083* (0.048)
Français	1253	895	-0.499 [1.01]	0.025 (0.064)
Mathématiques	2002	1426	-0.355 [1.08]	0.043 (0.045)
Sciences	1972	1446	-0.867 [0.808]	0.027 (0.036)
Physique-Chimie	1972	1446	-0.454 [0.786]	-0.006 (0.032)
Sciences de la vie	1972	1446	-1.27 [0.954]	0.057 (0.041)
Créativité	2509	1777	-0.320 [1.04]	0.029 (0.054)
Test F joint (p-value)				0.826
Panneau C : Caractéristiques des élèves				
Âge	6483	4649	12.64 [1.06]	-0.039* (0.023)
Score à l'examen de fin de primaire	6483	4649	7.10 [1.17]	0.032 (0.037)
% Filles	6483	4649	48.80 [49.99]	0.209 (0.839)
% Attrition	6483	4649	5.92 [23.61]	-0.965* (0.515)
Test F joint (p-value)				0.446

Notes. Ce tableau décrit l'échantillon de 11 132 élèves de l'étude et présente les tests d'équilibre. "Contrôle" désigne les 116 établissements du groupe contrôle. "Traitement" désigne les 84 établissements du groupe traité. "Différence" indique la différence ajustée par régression entre les établissements du groupe traité et ceux du groupe contrôle, en contrôlant pour les effets fixes de niveau et de groupe de randomisation, dans la colonne (4). Les résultats aux échelles inversées ont été retournés de sorte que des scores plus élevés correspondent à des résultats souhaitables. Les écarts-types sont indiqués entre crochets ; les erreurs standard sont indiquées entre parenthèses et sont regroupées au niveau de l'établissement. * p < 0,10, ** p < 0,05, *** p < 0,01.

Tableau 3 : Exposition au programme

	Moyenne du groupe contrôle	Effet ITT
	(1)	(2)
Panneau A : Exposition au programme		
L'élève a participé aux ateliers socio-émotionnels	0.217 [0.412]	0.432*** (0.023)
Connaît le spécialiste social	0.435 [0.496]	0.169*** (0.032)
Panneau B : Exposition aux autres interventions		
L'élève a bénéficié du TaRL	0.956 [0.205]	-0.002 (0.009)
L'élève a bénéficié de l'enseignement explicite	0.885 [0.319]	0.007 (0.013)
L'élève a participé à des activités extrascolaires	0.666 [0.472]	0.005 (0.019)
L'élève a participé à un programme de soutien scolaire	0.503 [0.500]	-0.012 (0.018)
L'élève à risque a participé à un programme de soutien scolaire	0.674 [0.470]	0.034 (0.034)

Notes. Ce tableau décrit l'exposition des élèves au programme dans les 200 collèges de l'étude, parmi l'échantillon d'analyse composé d'élèves non attrités ; la disponibilité des résultats varie selon la mesure d'exposition. "Contrôle" désigne les 116 collèges pionniers n'ayant pas reçu les ateliers socio-émotionnels. "Effet ITT" indique la différence ajustée par régression entre les établissements du groupe traité et ceux du groupe contrôle, en contrôlant pour les effets fixes groupe-de-randomisation, dans la colonne (2). Les écarts-types sont indiqués entre crochets ; les erreurs standard sont indiquées entre parenthèses et sont regroupées au niveau de l'établissement. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$.

Tableau 4 : Effets d'intention-de-traiter sur les compétences socio-émotionnelles

	Moyenne du groupe contrôle	Effet ITT
	(1)	(2)
Panneau A : Compétences interpersonnelles		
Global (indice)	-0.000 [1.00]	0.029 (0.030) {1.000}
Prosocialité	-0.000 [1.00]	-0.007 (0.031) {1.000}
Perception des émotions	-0.000 [1.00]	0.058** (0.029) {0.218}
Panneau B : Compétences intrapersonnelles		
Global (indice)	-0.000 [1.00]	0.012 (0.029) {1.000}
Sentiment de contrôle	-0.000 [1.00]	0.045 (0.030) {0.762}
Auto-régulation et discipline	-0.000 [1.00]	-0.027 (0.027) {0.762}

Notes. Ce tableau présente les effets d'intention-de-traiter (ITT) du programme, conformément à l'équation (1), sur les compétences socio-émotionnelles des élèves sans attrition dans les 200 collèges publics ayant passé les examens écrits en arabe ou en français. L'échantillon d'estimation varie selon le résultat en raison de données manquantes propres à chaque variable. La colonne (1) indique la moyenne du groupe contrôle, et la colonne (2) indique l'effet ITT. La "prosocialité" n'ayant pas été évaluée en début d'année, nous n'incluons pas le contrôle de début d'année lors de l'estimation de l'effet sur cette variable. Les écarts-types des moyennes du groupe contrôle sont indiqués entre crochets; les erreurs standard de l'effet de traitement sont indiquées entre parenthèses et sont regroupées au niveau de l'établissement; les q -valeurs sont indiquées entre accolades, calculées à l'aide de la correction de Anderson (2008) pour les tests d'hypothèses multiples (THM) selon la hiérarchie prédéfinie des familles de résultats, et conformément à Vivalt et al. (2024). Les mesures de la famille principale sont indiquées en gras. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$, avant correction pour les THM.

Tableau 5 : Effets d'intention-de-traiter sur les apprentissages et l'abandon scolaire

	Moyenne du groupe de contrôle	Effet ITT
	(1)	(2)
Panneau A : Apprentissages		
Global (scores agrégés)	0.000 [1.00]	0.034 (0.029) {1.000}
Arabe	0.000 [1.00]	0.023 (0.037) {1.000}
Français	-0.001 [1.00]	0.070* (0.039) {1.000}
Mathématiques	0.000 [1.00]	-0.033 (0.042) {1.000}
Sciences	0.001 [1.00]	0.067 (0.046) {1.000}
Physique-Chimie	0.002 [1.00]	0.123*** (0.042) {0.016}
Sciences de la vie et de la Terre	0.001 [1.00]	0.033 (0.048) {0.394}
Panneau B : Abandon scolaire et redoublement		
Abandon scolaire en fin d'année (global : exclusion ou décrochage)	0.026 [0.158]	0.001 (0.002) {1.000}
Abandon scolaire en fin d'année	0.023 [0.150]	0.001 (0.002) {1.000}
Redoublement en fin d'année	0.096 [0.295]	-0.017*** (0.006) {0.032}
Non réinscrit l'année scolaire suivante	0.026 [0.159]	0.006* (0.004) {0.404}
Exclu en fin d'année	0.003 [0.053]	0.000 (0.001) {1.000}

Notes. Ce tableau présente les effets d'intention-de-traiter (ITT) du programme, estimés selon l'équation (1). Le Panneau A rapporte l'effet du programme sur les apprentissages des élèves sans attrition dans les 200 collèges publics de l'échantillon. Le Panneau B présente l'effet du programme sur l'abandon scolaire et le redoublement parmi l'ensemble des élèves présents dans les établissements de l'étude. L'échantillon d'estimation varie selon le résultat considéré. La colonne (1) indique la moyenne du groupe de contrôle, et la colonne (2) rapporte l'effet ITT. Les écarts-types des moyennes du groupe de contrôle sont indiqués entre crochets; les erreurs-types de l'effet de traitement sont indiquées entre parenthèses et sont clusterisées au niveau de l'établissement; les q -valeurs sont indiquées entre accolades, calculées à l'aide de la correction de Anderson (2008) pour les tests d'hypothèses multiples (MHT) selon la hiérarchie de familles de résultats pré-spécifiée, et suivant Vivalt et al. (2024). Les mesures principales de chaque famille sont indiquées en gras. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$, avant correction pour les tests multiples.

Tableau 6 : Effets d'intention-de-traiter sur les autres résultats

	Moyenne du groupe de contrôle	Effet ITT
	(1)	(2)
Créativité	0.000 [1.00]	-0.079* (0.045) {1.000}
Climat scolaire et bien-être	0.000 [1.00]	0.014 (0.035) {1.000}
Habitudes de travail (plus de 30 min/jour de devoirs après l'école)	0.954 [0.210]	0.001 (0.006) {1.000}

Notes. Ce tableau présente les mécanismes potentiels et autres résultats du programme pour les élèves des 200 collèges de l'étude, parmi l'échantillon d'analyse sans attrition. L'échantillon d'estimation varie selon le résultat considéré. La colonne (1) indique la moyenne du groupe de contrôle et la colonne (2) rapporte l'effet ITT. Les résultats inversés ont été retournés de sorte que des scores plus élevés correspondent à des résultats désirables. Les écarts-types des moyennes du groupe de contrôle sont indiqués entre crochets; les erreurs-types de l'effet de traitement sont indiquées entre parenthèses et sont clusterisées au niveau de l'établissement; les q -valeurs sont indiquées entre accolades, calculées à l'aide de la correction de Anderson (2008) pour les tests d'hypothèses multiples (MHT) selon la hiérarchie de familles de résultats pré-spécifiée, et suivant Vivalt et al. (2024). * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$, avant correction pour les tests multiples.

Tableau 7 : Analyse pré-enregistrée des effets par sous-groupe

	Moyenne groupe contrôle	Effet ITT
	(1)	(2)
Panneau A : Élèves à risque		
<i>Compétences socio-émotionnelles</i>		
Compétences interpersonnelles	-0.260 [1.10]	0.004 (0.089) {1.000}
Compétences intrapersonnelles	-0.400 [0.866]	0.144* (0.075) {0.281}
<i>Apprentissages et abandon scolaire</i>		
Apprentissages (scores agrégés)	-0.609 [0.762]	0.015 (0.048) {1.000}
Abandon scolaire en fin d'année (global : exclusion ou décrochage)	0.085 [0.279]	-0.002 (0.008) {1.000}
<i>Résultats supplémentaires</i>		
Créativité	-0.196 [1.06]	-0.050 (0.087) {1.000}
Climat scolaire et bien-être	-0.194 [1.00]	0.002 (0.083) {1.000}
Habitudes de travail (plus de 30 min/jour après l'école)	0.884 [0.320]	0.039 (0.024) {1.000}
Panneau B : Élèves filles		
<i>Compétences socio-émotionnelles</i>		
Compétences interpersonnelles	0.102 [0.971]	0.043 (0.041) {1.000}
Compétences intrapersonnelles	0.065 [1.04]	-0.014 (0.039) {1.000}
<i>Apprentissages et abandon scolaire</i>		
Apprentissages (scores agrégés)	0.169 [0.996]	0.064* (0.034) {0.868}
Abandon scolaire en fin d'année (global : exclusion ou décrochage)	0.016 [0.124]	0.000 (0.002) {1.000}
<i>Résultats supplémentaires</i>		
Créativité	0.051 [1.00]	-0.054 (0.051) {1.000}
Climat scolaire et bien-être	0.148 [0.971]	-0.018 (0.044) {1.000}
Habitudes de travail (plus de 30 min/jour de devoirs après l'école)	0.963 [0.189]	-0.001 (0.008) {1.000}

Notes. Ce tableau présente les effets d'intention-de-traiter (ITT) du programme, estimés selon l'équation (1), pour les résultats par sous-groupes pré-spécifiés dans les 200 collèges publics de l'échantillon. Le Panneau A rapporte les effets pour les élèves à risque, et le Panneau B rapporte les effets pour les élèves filles. L'échantillon d'estimation varie selon le résultat et le sous-groupe considérés. Les lignes sont regroupées selon les familles de résultats des tableaux principaux correspondants : compétences socio-émotionnelles dans le Tableau 4, apprentissages et abandon scolaire dans le Tableau 5, et résultats supplémentaires et mécanismes potentiels dans le Tableau 6. La colonne (1) indique la moyenne du groupe de contrôle et la colonne (2) rapporte l'effet ITT. Les écarts-types des moyennes du groupe de contrôle sont indiqués entre crochets; les erreurs-types de l'effet de traitement sont indiquées entre parenthèses et sont clusterisées au niveau de l'établissement; les q -valeurs sont indiquées entre accolades, calculées à l'aide de la correction de Anderson (2008) pour les tests d'hypothèses multiples (MHT) selon la hiérarchie pré-spécifiée de sous-groupes et de familles de résultats, et suivant Vivalt et al. (2024). * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$, avant correction pour les tests multiples.

Références

- Alan, S., Boneva, T., Ertac, S., 2019. Ever failed, try again, succeed better : Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics* 134, 1121–1162.
- Anderson, M.L., 2008. Multiple Inference and Gender Differences in the Effects of Early Intervention : A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association* 103, 1481–1495. doi :10.1198/016214508000000841.
- Ashraf, N., Bau, N., Low, C., McGinn, K., 2020. Negotiating a Better Future : How Interpersonal Skills Facilitate Intergenerational Investment*. *The Quarterly Journal of Economics* 135, 1095–1151. doi :10.1093/qje/qjz039.
- Bailey, D.H., Duncan, G.J., Cunha, F., Foorman, B.R., Yeager, D.S., 2020. Persistence and Fade-Out of Educational-Intervention Effects : Mechanisms and Potential Solutions. *Psychological Science in the Public Interest* 21, 55–97. doi :10.1177/1529100620915848.
- Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., Walton, M., 2017. From Proof of Concept to Scalable Policies : Challenges and Solutions, with an Application. *Journal of Economic Perspectives* 31, 73–102. doi :10.1257/jep.31.4.73.
- Barrera-Osorio, F., de Barros, A., Filmer, D., 2024. Longterm Impacts of Primary School Scholarships : Evidence from Cambodia. *Journal of Policy Analysis and Management* 43, 10–38. doi :10.1002/pam.22533.
- de Barros, A., Campos Quintero, A., El Amrani Mida, N., Glewwe, P., Kumar, N., Lépine, L., 2026. Beyond Basics : Whole-School Reform and Early Adolescent Development. Working Paper 260. Berkeley, CA. URL : <https://doi.org/10.26085/C3RP5W>.
- de Barros, A., Fajardo-Gonzalez, J., Glewwe, P., Sankar, A., 2024. The Limitations of Activity-Based Instruction to Improve the Productivity of Schooling. *The Economic Journal* 134, 959–984. doi :10.1093/ej/uead099.
- de Barros, A., Lubozha, T., 2026. Targeting Foundational Skills at Scale : Skill Specificity and Transfer. Working Paper 12542. Munich. URL : https://www.ifo.de/sites/default/files/docbase/docs/cesifol_wp12542.pdf.
- Beaman, L., Herskowitz, S., Keleher, N., Magruder, J., 2021. Stay in the Game : A Randomized Controlled Trial of a Sports and Life Skills Program for Vulnerable Youth

- in Liberia. *Economic Development and Cultural Change* 70, 129–158. doi :10.1086/711651.
- Behaghel, L., Crépon, B., Gurgand, M., Le Barbanchon, T., 2015. Please call again : Correcting nonresponse bias in treatment effect models. *Review of Economics and Statistics* 97, 1070–1080.
- Belloni, A., Chernozhukov, V., Hansen, C., 2014. High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives* 28, 29–50. doi :10.1257/jep.28.2.29.
- Cilliers, J., Elashmawy, N., McKenzie, D., 2024. Using Post-Double Selection Lasso in Field Experiments. Working Paper 10931. The World Bank. Washington, D.C. URL : <https://openknowledge.worldbank.org/entities/publication/0cde089d-33ba-4f51-8c03-b25b5114d41a>.
- Cipriano, C., Strambler, M.J., Naples, L.H., Ha, C., Kirk, M., Wood, M., Sehgal, K., Zieher, A.K., Eveleigh, A., McCarthy, M., Funaro, M., Ponnock, A., Chow, J.C., Durlak, J., 2023. The state of evidence for social and emotional learning : A contemporary meta-analysis of universal school-based SEL interventions. *Child Development* 94, 1181–1204. doi : <https://doi.org/10.1111/cdev.13968>.
- Cohen, S., Kamarck, T., Mermelstein, R., 1983. A global measure of perceived stress. *Journal of Health and Social Behavior* 24, 385–396. doi :10.2307/2136404.
- Dam, A., Gray-Lobe, G., Kremer, M., de Laat, J., Morsink, K., 2025. Learning to Work Towards Goals : A Sequential Evaluation of the Effect of Goal-Setting Course on Academic and Soft Skills. Working Paper 25-1344. Annenberg Institute at Brown University. Providence, RI. URL : <https://edworkingpapers.com/sites/default/files/ai25-1344.pdf>.
- Danon, A., Das, J., de Barros, A., Filmer, D., 2024. Cognitive and Socioemotional Skills in Low-Income Countries : Measurement and Associations with Schooling and Earnings. *Journal of Development Economics* 168, 103132. doi :10.1016/j.jdeveco.2023.103132.
- von Davier, M., Kennedy, A., Reynolds, K., Fishbein, B., Khorramdel, L., Aldrich, C., Bookbinder, A., Bezirhan, U., Yin, L., 2024. TIMSS 2023 International Results in Mathematics and Science. Boston College, TIMSS & PIRLS International Study Center. doi :10.6017/lse.tpisc.timss.rs6460.
- Deming, D., 2017. The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics* 132, 1593–1640. doi :10.1093/qje/qjx022.

- Dhar, D., Jain, T., Jayachandran, S., 2022. Reshaping Adolescents' Gender Attitudes : Evidence from a School-Based Experiment in India. *American Economic Review* 112, 899–927. doi :10.1257/aer.20201112.
- Dillon, M.R., Kannan, H., Dean, J.T., Spelke, E.S., Duflo, E., 2017. Cognitive science in the field : A preschool intervention durably enhances intuitive but not formal mathematics. *Science* 357, 47–55. doi :10.1126/science.aal4724.
- Dinarte-Diaz, L., Egana-delSol, P., Martínez Alvear, C., Rojas Alvarado, C., 2024. When emotion regulation matters : The efficacy of socio-emotional learning to address school-based violence in Central America. Working Paper IDB-WP-1585. IDB Working Paper Series. doi :10.18235/0012854.
- Duckworth, A.L., Quinn, P.D., 2009. Development and Validation of the Short Grit Scale (GritS). *Journal of Personality Assessment* 91, 166–174. doi :10.1080/00223890802634290.
- Durlak, J.A., Domitrovich, C.E., Weissberg, R.P., Gullotta, T.P. (Eds.), 2015. *Handbook of Social and Emotional Learning : Research and Practice*. Guilford Press, New York.
- Eble, A., Frost, C., Camara, A., Bouy, B., Bah, M., Sivaraman, M., Hsieh, P.T.J., Jayanty, C., Brady, T., Gawron, P., Vansteelandt, S., Boone, P., Elbourne, D., 2021. How much can we remedy very low learning levels in rural parts of low-income countries? Impact and generalizability of a multi-pronged para-teacher intervention from a cluster-randomized trial in the Gambia. *Journal of Development Economics* 148, 102539. doi :10.1016/j.jdeveco.2020.102539.
- Edmonds, E., Feigenberg, B., Leight, J., 2023. Advancing the Agency of Adolescent Girls. *Review of Economics and Statistics* 105, 852–866. doi :10.1162/rest_a_01074.
- Finn, J.D., 1989. Withdrawing from school. *Review of Educational Research* 59, 117–142. doi :10.3102/00346543059002117.
- Ganimian, A.J., 2020. Growth-Mindset Interventions at Scale : Experimental Evidence From Argentina. *Educational Evaluation and Policy Analysis* 42, 417–438. doi :10.3102/0162373720938041.
- Goodman, R., 1997. The Strengths and Difficulties Questionnaire : A Research Note. *Journal of Child Psychology and Psychiatry* 38, 581–586. doi :10.1111/j.1469-7610.1997.tb01545.x.

- Huillery, E., Bouguen, A., Charpentier, A., Algan, Y., Chevallier, C., 2025. The Role of Mindset in Education : A Large-Scale Field Experiment in Disadvantaged Schools. *The Economic Journal* , ueaf015doi :10.1093/ej/ueaf015.
- Jackson, C.K., 2018. What Do Test Scores Miss? The Importance of Teacher Effects on NonTest Score Outcomes. *Journal of Political Economy* 126, 2072–2107. doi :10.1086/699018.
- Jacob, B., Rothstein, J., 2016. The Measurement of Student Ability in Modern Assessment Systems. *Journal of Economic Perspectives* 30, 85–108. doi :10.1257/jep.30.3.85.
- Kerwin, J., Rostom, N., Sterck, O., 2024. Striking the right balance : Why standard balance tests over-reject the null, and how to fix it .
- Kilanowski, J.F., 2017. Breadth of the socio-ecological model. *Journal of Agromedicine* 22, 295–297.
- Krishnan, P., Krutikova, S., 2013. Non-cognitive skill formation in poor neighbourhoods of urban India. *Labour Economics* 24, 68–85. doi :10.1016/j.labeco.2013.06.004.
- Laajaj, R., Macours, K., 2021. Measuring Skills in Developing Countries. *Journal of Human Resources* 56, 1254–1295. doi :10.3368/jhr.56.4.1018-9805R1.
- Lee, D.S., 2009. Training, Wages, and Sample Selection : Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies* 76, 1071–1102. doi :10.1111/j.1467-937X.2009.00536.x.
- Mullis, I.V.S., von Davier, M., Foy, P., Fishbein, B., Reynolds, K.A., Wry, E., 2023. PIRLS 2021 International Results in Reading. Boston College, TIMSS & PIRLS International Study Center. doi :10.6017/lse.tpisc.tr2103.kb5342.
- Romero, M., Sandefur, J., Sandholtz, W.A., 2020. Outsourcing Education : Experimental Evidence from Liberia. *American Economic Review* 110, 364–400. doi :10.1257/aer.20181478.
- Santos, I., Petroska-Beska, V., Carneiro, P., Eskreis-Winkler, L., Boudet, A.M.M., Berniell, I., Krekel, C., Arias, O., Duckworth, A.L., 2022. Can Grit Be Taught? Lessons from a Nationwide Field Experiment with Middle-School Students. Working Paper 15588. IZA Institute of Labor Economics. Bonn, Germany. URL : <https://www.econstor.eu/handle/10419/265809>.
- The World Bank, 2025. Education Statistics (EdStats). URL : <https://datatopics.worldbank.org/education/>.

- Torrance, E.P., 1968. *Torrance Tests of Creative Thinking*. Personnel, Princeton, NJ.
- Vivalt, E., Rhodes, E., Bartik, A.W., Broockman, D.E., Krause, P., Miller, S., 2024. The Employment Effects of a Guaranteed Income : Experimental Evidence from Two U.S. States. doi :10.3386/w32719.
- Wager, S., Athey, S., 2018. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association* 113, 1228–1242. doi :10.1080/01621459.2017.1319839.
- Weidmann, B., Xu, Y., 2024. PAGE : A Modern Measure of Emotion Perception for Teamwork and Management Research. doi :10.48550/ARXIV.2410.03704.
- Wilson, S.J., Freeman, B., Hedberg, E.C., 2025. Empirical Benchmarks for Effect Size Interpretation and Study Planning with Social and Behavioral Outcomes. *Journal of Research on Educational Effectiveness* , 1–26doi :10.1080/19345747.2024.2427767.
- World Health Organization, 1993. *Life Skills Education for Children and Adolescents in Schools*. Technical Report WHO/MNH/PSF/93.7A.Rev.2. World Health Organization, Division of Mental Health. Geneva.
- World Health Organization, 2003. *Skills for Health : Skills-Based Health Education Including Life Skills*. WHO Information Series on School Health 9. World Health Organization. Geneva.
- Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E., Wager, S., 2025. Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects. *Journal of the American Statistical Association* 120, 38–51. doi :10.1080/01621459.2024.2393466.

Annexes en ligne

A Tableaux et figures supplémentaires

Tableau A1 : Tests d'équilibre (échantillon sans attrition)

	Nombre d'observations		Test d'équilibre	
	Contrôle	Traitement	Moyenne du groupe contrôle	Différence
	(1)	(2)	(3)	(4)
Panneau A : Compétences socio-émotionnelles				
Compétences interpersonnelles	2331	1681	-0.353 [1.21]	-0.012 (0.046)
Perception des émotions	2331	1681	-0.276 [0.946]	-0.010 (0.036)
Compétences intrapersonnelles	2331	1681	1.91 [0.793]	-0.002 (0.026)
Sentiment de contrôle	2331	1681	1.29 [0.675]	0.002 (0.022)
Auto-régulation et discipline	2331	1681	2.08 [0.996]	-0.006 (0.035)
Test F joint (p-value)				0.991
Panneau B : Compétences académiques et créativité				
Apprentissage (scores aux tests empilés)	6099	4427	-0.576 [1.01]	0.040 (0.036)
Arabe	1160	840	-0.585 [1.05]	0.087* (0.049)
Français	1171	841	-0.488 [1.01]	0.037 (0.065)
Mathématiques	1904	1362	-0.349 [1.08]	0.047 (0.046)
Sciences	1864	1384	-0.857 [0.801]	0.033 (0.035)
Physique-Chimie	1864	1384	-0.448 [0.782]	0.002 (0.032)
Sciences de la vie	1864	1384	-1.26 [0.947]	0.060 (0.041)
Créativité	2331	1681	-0.328 [1.03]	0.040 (0.055)
Test F joint (p-value)				0.903
Panneau C : Caractéristiques des élèves				
Âge	6099	4427	12.61 [1.02]	-0.033 (0.022)
Score à l'examen de fin de primaire	6099	4427	7.09 [1.12]	0.042 (0.036)
% Filles	6099	4427	49.37 [50.00]	0.204 (0.861)
Test F joint (p-value)				0.710

Notes. Ce tableau décrit l'échantillon sans attrition de l'étude, composé de 10 526 élèves, et présente des tests d'équilibre. "Contrôle" désigne les 116 collèges du groupe de contrôle. "Traitement" désigne les 84 collèges du groupe traité. "Différence" rapporte la différence ajustée par régression entre les collèges du groupe traité et ceux du groupe de contrôle, en contrôlant pour les effets fixes de niveau et de groupe de randomisation, dans la colonne (4). Les résultats inversés ont été retournés de sorte que des scores plus élevés représentent des résultats souhaitables. Les écarts-types sont indiqués entre crochets. Les erreurs-types sont indiquées entre parenthèses et sont clusterisées au niveau de l'établissement. Les mesures des familles principales sont indiquées en gras. * p < 0,10, ** p < 0,05, *** p < 0,01.

Tableau A2 : Exposition au programme des élèves à risque

	Moyenne du groupe contrôle	Effet ITT
	(1)	(2)
Panneau A : Exposition au programme		
L'élève a participé aux ateliers socio-émotionnels	0.231 [0.422]	0.407*** (0.041)
Connaît le spécialiste social	0.431 [0.496]	0.135*** (0.045)
Panneau B : Exposition aux autres interventions		
L'élève a bénéficié du TaRL	0.913 [0.281]	0.004 (0.016)
L'élève a bénéficié de l'enseignement explicite	0.859 [0.348]	0.014 (0.033)
L'élève a participé à des activités extrascolaires	0.575 [0.495]	0.063 (0.051)
L'élève a participé à un programme de soutien scolaire	0.684 [0.465]	0.042 (0.047)

Notes. Ce tableau décrit l'exposition au programme des élèves à risque d'abandon scolaire dans les 200 collèges de l'étude, parmi l'échantillon d'élèves sans attrition. La colonne (1) indique la moyenne du groupe de contrôle parmi les élèves à risque, et la colonne (2) rapporte l'effet ITT pour les élèves à risque, estimé dans l'échantillon complet comme la somme du coefficient de traitement et de l'interaction traitement-par-élève-à-risque. Les écarts-types sont indiqués entre crochets. Les erreurs-types sont indiquées entre parenthèses et sont clusterisées au niveau de l'établissement. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$.

Tableau A3 : Exposition au programme des élèves filles

	Moyenne du groupe contrôle	Effet ITT
	(1)	(2)
Panneau A : Exposition au programme		
L'élève a participé aux ateliers socio-émotionnels	0.217 [0.413]	0.443*** (0.027)
Connaît le spécialiste social	0.466 [0.499]	0.190*** (0.037)
Panneau B : Exposition aux autres interventions		
L'élève a bénéficié du TaRL	0.962 [0.192]	0.001 (0.011)
L'élève a bénéficié de l'enseignement explicite	0.877 [0.328]	0.019 (0.017)
L'élève a participé à des activités extrascolaires	0.661 [0.474]	0.008 (0.025)
L'élève a participé à un programme de soutien scolaire	0.469 [0.499]	-0.018 (0.023)
L'élève à risque a participé à un programme de soutien scolaire	0.678 [0.470]	-0.061 (0.081)

Notes. Ce tableau décrit l'exposition au programme des élèves filles dans les 200 collèges de l'étude, parmi l'échantillon d'élèves sans attrition. La colonne (1) indique la moyenne du groupe de contrôle parmi les élèves filles, et la colonne (2) rapporte l'effet ITT pour les élèves filles, estimé dans l'échantillon complet comme la somme du coefficient de traitement et de l'interaction traitement-par-élève-fille. Les écarts-types sont indiqués entre crochets. Les erreurs-types sont indiquées entre parenthèses et sont clusterisées au niveau de l'établissement. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$.

Tableau A4 : Effets sur les sous-domaines des compétences socio-émotionnelles et la créativité

	Moyenne du groupe contrôle	Effet ITT
	(1)	(2)
Compétences interpersonnelles	-0.000 [1.00]	0.029 (0.030) {1.000}
Pro-socialité	-0.000 [1.00]	-0.007 (0.031) {1.000}
Perception des émotions	-0.000 [1.00]	0.058** (0.029) {0.218}
Compétences intrapersonnelles	-0.000 [1.00]	0.012 (0.029) {1.000}
Sentiment de contrôle	-0.000 [1.00]	0.045 (0.030) {0.762}
Mentalité de croissance	0.000 [1.00]	0.042 (0.030) {0.480}
Lieu de contrôle	-0.000 [1.00]	0.035 (0.034) {0.480}
Sentiment d'auto-efficacité	-0.000 [1.00]	0.057** (0.028) {0.435}
Auto-régulation et discipline	-0.000 [1.00]	-0.027 (0.027) {0.762}
Discipline de travail et persévérance	-0.001 [1.00]	-0.007 (0.026) {0.947}
Maîtrise de soi	0.000 [1.00]	-0.097*** (0.027) {0.003}
Créativité	0.000 [1.00]	-0.079* (0.045) {1.000}

Notes. Ce tableau présente les effets d'intention-de-traiter (ITT) du programme, estimés selon l'équation (1), sur les compétences socio-émotionnelles et la créativité des élèves, parmi les 4 012 élèves sans attrition des 200 collèges publics ayant passé les épreuves écrites d'arabe ou de français. La colonne (1) indique la moyenne du groupe de contrôle et la colonne (2) rapporte l'effet ITT. La "pro-socialité" n'ayant pas été mesurée en début d'année, nous n'incluons pas le contrôle de début d'année dans l'estimation de l'effet sur cette variable. Les écarts-types sont indiqués entre crochets. Les erreurs-types sont indiquées entre parenthèses et sont clusterisées au niveau de l'établissement ; les valeurs q sont indiquées entre accolades, calculées à l'aide de la correction de Anderson (2008) pour les tests d'hypothèses multiples (THM), et suivant Vivalt et al. (2024). Les mesures des familles principales sont indiquées en gras. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$, avant correction pour THM.

Tableau A5 : Effets d'intention-de-traiter sur les sous-domaines de l'apprentissage

	Moyenne du groupe contrôle	Effet ITT
	(1)	(2)
Panneau A : Arabe		
Écrit	0.000 [1.00]	0.029 (0.034) [0.766]
Lecture et compréhension	0.000 [1.00]	-0.009 (0.034) [1.000]
Production écrite	0.000 [1.00]	0.124*** (0.043) [0.044]
Oral	-0.000 [1.00]	0.118*** (0.044) [0.044]
Panneau B : Français		
Écrit	-0.001 [1.00]	0.081* (0.042) [0.394]
Lecture et compréhension	-0.001 [1.00]	0.049 (0.039) [0.567]
Production écrite	-0.001 [1.00]	0.132*** (0.049) [0.095]
Oral	-0.001 [1.00]	0.047 (0.044) [0.567]
Panneau C : Mathématiques		
En dessous du niveau scolaire	-0.001 [1.00]	-0.058* (0.032) [0.275]
Au niveau scolaire	-0.001 [1.00]	-0.058* (0.032) [0.275]
Application ou raisonnement	-0.001 [1.00]	-0.074** (0.033) [0.253]
Connaissance	0.001 [0.999]	0.029 (0.034) [0.576]
Algèbre	0.001 [0.999]	-0.048 (0.040) [0.510]
Données et aléatoire	0.000 [1.00]	-0.110** (0.044) [0.253]
Géométrie	-0.000 [1.00]	-0.041 (0.047) [0.576]
Nombres	-0.001 [1.00]	-0.039 (0.043) [0.576]
Panneau D : Sciences		
Application ou raisonnement	0.001 [1.00]	0.081* (0.044) [0.138]
Connaissance	0.002 [0.999]	0.084* (0.045) [0.138]
Biologie	0.001 [0.999]	0.095** (0.042) [0.085]
Physique	0.002 [1.00]	0.148*** (0.042) [0.006]
Chimie	0.002 [0.999]	0.135*** (0.041) [0.007]

Notes. Ce tableau présente les effets d'intention-de-traiter (ITT) du programme sur l'apprentissage des élèves, parmi les 10 526 élèves sans attrition des 200 collèges publics, estimés selon l'équation (1). La colonne (1) indique la moyenne du groupe de contrôle et la colonne (2) rapporte l'effet ITT. Les écarts-types sont indiqués entre crochets. Les erreurs-types sont indiquées entre parenthèses et sont clusterisées au niveau de l'établissement; les valeurs q sont indiquées entre accolades, calculées à l'aide de la correction de Anderson (2008) pour les tests d'hypothèses multiples (THM), et suivant Vivalt et al. (2024). * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$, avant correction pour THM.

Tableau A6 : Effets par sous-groupe sur les apprentissages par matière

	Moyenne du groupe de contrôle	Effet ITT
	(1)	(2)
Panneau A : Élèves à risque		
Arabe	-0.752 [0.838]	0.015 (0.078) {1.000}
Français	-0.651 [0.655]	-0.005 (0.084) {1.000}
Mathématiques	-0.550 [0.789]	-0.016 (0.082) {1.000}
Sciences	-0.553 [0.732]	0.037 (0.079) {1.000}
Panneau B : Élèves filles		
Arabe	0.233 [0.957]	0.018 (0.046) {1.000}
Français	0.161 [0.970]	0.110** (0.047) {0.495}
Mathématiques	0.115 [0.991]	0.003 (0.052) {1.000}
Sciences	0.191 [1.04]	0.091 (0.060) {1.000}

Notes. Ce tableau présente les effets d'intention-de-traiter (ITT) du programme sur les apprentissages des élèves sans attrition dans les 200 collèges publics de l'échantillon, estimés selon l'équation (1). Le Panneau A rapporte les effets pour le sous-groupe d'élèves à risque d'abandon scolaire et le Panneau B rapporte les effets pour les élèves filles. La colonne (1) indique la moyenne du groupe de contrôle, et la colonne (2) rapporte l'effet ITT. Les écarts-types sont indiqués entre crochets ; les erreurs-types sont indiquées entre parenthèses et sont clusterisées au niveau de l'établissement ; les q -valeurs sont indiquées entre accolades, calculées à l'aide de la correction de Anderson (2008) pour les tests d'hypothèses multiples (MHT), et suivant Vivalt et al. (2024). * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$, avant correction pour les tests multiples.

Tableau A7 : Robustesse des effets d'intention-de-traiter sur l'abandon scolaire selon différentes définitions de l'échantillon

	Moyenne du groupe de contrôle	Effet ITT
	(1)	(2)
Panneau A : Élèves inscrits au début du mois de juillet		
Abandon scolaire en fin d'année (global : exclusion ou décrochage)	0.035 [0.183]	0.002 (0.003)
Abandon scolaire en fin d'année	0.030 [0.171]	0.002 (0.002)
Redoublement en fin d'année	0.105 [0.307]	-0.017*** (0.006)
Non réinscrit l'année scolaire suivante	0.032 [0.176]	0.008** (0.004)
Exclu en fin d'année	0.004 [0.067]	-0.000 (0.001)
Panneau B : Élèves inscrits au début du mois d'août		
Abandon scolaire en fin d'année (global : exclusion ou décrochage)	0.027 [0.161]	0.001 (0.002)
Abandon scolaire en fin d'année	0.024 [0.152]	0.001 (0.002)
Redoublement en fin d'année	0.097 [0.296]	-0.017*** (0.006)
Non réinscrit l'année scolaire suivante	0.027 [0.161]	0.007* (0.004)
Exclu en fin d'année	0.003 [0.053]	-0.000 (0.001)
Panneau C : Élèves inscrits au début du mois d'octobre		
Abandon scolaire en fin d'année (global : exclusion ou décrochage)	0.025 [0.156]	0.001 (0.002)
Abandon scolaire en fin d'année	0.023 [0.148]	0.002 (0.002)
Redoublement en fin d'année	0.094 [0.291]	-0.019*** (0.006)
Non réinscrit l'année scolaire suivante	0.027 [0.161]	0.007* (0.004)
Exclu en fin d'année	0.003 [0.050]	-0.000 (0.001)

Notes. Ce tableau présente les effets d'intention-de-traiter (ITT) du programme, estimés selon l'équation (1). La colonne (1) indique la moyenne du groupe de contrôle et la colonne (2) rapporte l'effet ITT. Le Panneau A rapporte l'effet du programme sur les élèves recensés comme inscrits au début du mois de juillet (lorsque l'année scolaire précédente se termine). Le Panneau B rapporte l'effet du programme sur les élèves recensés comme inscrits au début du mois d'août (lorsque la nouvelle année scolaire commence). Le Panneau C rapporte l'effet du programme sur les élèves recensés comme inscrits au début du mois d'octobre (un mois après le début de la nouvelle année scolaire). Les écarts-types sont indiqués entre crochets. Les erreurs-types sont indiquées entre parenthèses et sont clusterisées au niveau de l'établissement. * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$.

Tableau A8 : Effet sur les compétences socio-émotionnelles : robustesse au biais de désirabilité sociale

	Différence	Désirabilité sociale	
		En dessous de la médiane	Au-dessus de la médiane
		(1)	(2)
Compétences interpersonnelles	0.022 (0.069)	-0.007 (0.039) {1.000}	0.015 (0.054) {1.000}
Prosocialité	-0.028 (0.061)	-0.004 (0.040) {1.000}	-0.032 (0.049) {1.000}
Perception des émotions	0.097 (0.067)	0.007 (0.039) {1.000}	0.104** (0.052) {0.217}
Compétences intrapersonnelles	-0.056 (0.062)	0.029 (0.036) {1.000}	-0.027 (0.051) {1.000}
Sentiment de contrôle	-0.022 (0.060)	0.042 (0.036) {1.000}	0.021 (0.050) {1.000}
Auto-régulation et discipline	-0.084 (0.062)	0.011 (0.035) {1.000}	-0.073 (0.049) {1.000}

Notes. Ce tableau présente l'hétérogénéité des effets d'intention-de-traiter (ITT) du programme, estimés selon l'équation (1), sur les compétences socio-émotionnelles des élèves, parmi les 4 012 élèves sans attrition des 200 collèges publics de l'échantillon disposant de scores de désirabilité sociale non manquants en début d'année et ayant passé les examens écrits en arabe ou en français. La colonne (1) rapporte la différence d'effet entre les élèves présentant un biais de désirabilité sociale élevé et ceux présentant un biais faible; les colonnes (2) et (3) rapportent respectivement les effets séparément pour chacun des deux sous-échantillons. La "prosocialité" n'ayant pas été évaluée en début d'année, le contrôle de début d'année n'est pas inclus lors de l'estimation de l'effet sur cette variable. Les erreurs-types sont indiquées entre parenthèses et sont clusterisées au niveau de l'établissement. * p < 0,10, ** p < 0,05, *** p < 0,01.

Tableau A9 : Sensibilité à l'attrition différentielle

	Compétences interpersonnelles	Compétences intrapersonnelles	Apprentissages (scores agrégés)
	(1)	(2)	(3)
Panneau A : Bornes de Lee (2009)			
Borne inférieure			
Treatment	0.020 (0.030) {0.764}	-0.010 (0.030) {1.000}	0.030 (0.030) {1.000}
Borne supérieure			
Treatment	0.020 (0.030) {0.777}	0.020 (0.030) {1.000}	0.030 (0.030) {1.000}
Panneau B : Bornes de Behaghel et al. (2015) : attrition basée sur le nombre de jours de suivi			
Borne inférieure			
Treatment	0.020 (0.030) {0.786}	0.010 (0.030) {1.000}	0.030 (0.030) {1.000}
Borne supérieure			
Treatment	0.020 (0.030) {0.786}	0.010 (0.030) {1.000}	0.040 (0.030) {1.000}
Panneau C : IPW – attrition basée sur les covariables de début d'année			
Treatment	0.030 (0.030) {0.407}	0.000 (0.030) {1.000}	0.030 (0.030) {1.000}

Notes. Effets de traitement estimés. Les variables dépendantes des colonnes (1), (2) et (3) sont les indices de famille des Tableaux 4 et 5. Le traitement capture les effets pour les collèges ayant reçu le traitement. Le Panneau A estime des bornes pour les effets ITT suivant Lee (2009). Le Panneau B estime des bornes resserrées pour les effets ITT suivant Behaghel et al. (2015). Le Panneau C estime un effet ITT par IPW en utilisant les covariables de début d'année comme prédicteurs de l'attrition. Dans le Panneau C, la probabilité d'attrition est estimée par une régression probit d'un indicateur d'attrition sur le statut de traitement et les covariables de début d'année. L'information sur le "nombre de jours de suivi" inclut le nombre de jours de visites dans un collège avant qu'un élève soit enquêté. Toutes les régressions suivent les spécifications des Tableaux 4 et 5, à l'exception des pondérations. Les erreurs-types sont clusterisées au niveau de l'établissement et sont indiquées entre parenthèses; les q -valeurs sont indiquées entre accolades, calculées à l'aide de la correction de Anderson (2008) pour les tests d'hypothèses multiples (MHT), et suivant Vivalt et al. (2024). * $p < 0,10$, ** $p < 0,05$, *** $p < 0,01$, avant correction pour les tests multiples.

Références

- Anderson, M.L., 2008. Multiple Inference and Gender Differences in the Effects of Early Intervention : A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association* 103, 1481–1495. doi :10.1198/016214508000000841.
- Behaghel, L., Crépon, B., Gurgand, M., Le Barbanchon, T., 2015. Please call again : Correcting nonresponse bias in treatment effect models. *Review of Economics and Statistics* 97, 1070–1080.
- Lee, D.S., 2009. Training, Wages, and Sample Selection : Estimating Sharp Bounds on Treatment Effects. *The Review of Economic Studies* 76, 1071–1102. doi :10.1111/j.1467-937X.2009.00536.x.
- Vivalt, E., Rhodes, E., Bartik, A.W., Broockman, D.E., Krause, P., Miller, S., 2024. The Employment Effects of a Guaranteed Income : Experimental Evidence from Two U.S. States. doi :10.3386/w32719.

B Mesure

Les modèles de théorie de la réponse à l’item (TRI) sont utilisés pour estimer la probabilité de répondre correctement à un item en fonction d’un paramètre latent représentant le niveau de l’élève et de paramètres propres à chaque item (Jacob and Rothstein, 2016). Pour agréger les réponses aux questions de tests et d’entretiens (“items”), nous estimons différents types de modèles TRI, chacun sélectionné en fonction de la structure et des caractéristiques de l’instrument considéré.

Cette annexe décrit brièvement chacun des modèles utilisés pour agréger nos items, ainsi que les méthodes employées pour construire des indices à partir des scores TRI. Le Tableau B1 de l’annexe présente les propriétés psychométriques de chaque instrument d’évaluation et d’enquête, notamment le nombre d’items, la discrimination et la difficulté moyennes des items, ainsi que la fiabilité conditionnelle moyenne en fin d’année. Pour estimer les scores de niveau des élèves, chacune de nos méthodes TRI recourt au scoring par espérance a posteriori (EAP). Soland et al. (2024) montrent que, pour le scoring EAP, le fait de ne pas tenir compte de la variation inter-groupes dans les distributions de niveau peut conduire à des estimations biaisées des effets de traitement dans les interventions éducatives. Pour éviter ce biais, nous suivons les recommandations de Soland et al. (2024) et estimons les scores à l’aide d’un modèle TRI à deux groupes. Ainsi, les groupes traité et de contrôle peuvent avoir des distributions de niveau distinctes (c’est-à-dire des moyennes et des variances différentes), tout en supposant des paramètres d’items communs.

B.1 Modèle logistique à deux paramètres (2PL)

Le modèle logistique à deux paramètres (2PL) estime la probabilité de répondre correctement à un item en fonction du niveau de l’examiné (θ) et de la difficulté (b_i) de chaque item. Ce type de modèle ne s’applique qu’aux instruments comportant des items strictement binaires (y compris les questions de test notées correct/incorrect); nous l’appliquons donc aux tests de langue, de mathématiques et de sciences, ainsi qu’à l’instrument de perception des émotions (PAGE).

Le modèle logistique à deux paramètres (2PL) est donné par :

$$P(Y_{ij} = 1 | \theta_j) = \frac{1}{1 + \exp[-a_i(\theta_j - b_i)]} \quad (2)$$

où θ_j est le niveau latent de l’élève j , a_i est le paramètre de discrimination et b_i est la difficulté de l’item i .

B.2 Modèle de réponse graduée (GRM)

Pour les items de type Likert à réponses catégorielles ordonnées (par exemple, 1 = Jamais à 5 = Toujours), nous utilisons le modèle de réponse graduée (GRM), approprié pour des instruments tels que *SDQ* (*prosocialité*).

La probabilité de répondre dans la catégorie k ou supérieure est donnée par :

$$P(Y_{ij} \geq k | \theta_j) = \frac{1}{1 + \exp[-a_i(\theta_j - b_{ik})]}$$

où b_{ik} représente le seuil entre les catégories $k - 1$ et k . Ce modèle capture à la fois la sensibilité de l'item (via a_i) et la difficulté variable des seuils de réponse.

Nous avons également pré-enregistré le modèle de crédit partiel (PCM) par crainte que le modèle de réponse graduée ne puisse pas prendre en charge des items aux catégories ordinales variables ou inégalement espacées. À la mise en oeuvre, le GRM les traite directement (et estime un paramètre de discrimination des items que le PCM omet); nous avons donc estimé tous les indices comportant des catégories de réponse ordinales à l'aide du GRM.

B.3 Analyse en composantes principales de Poisson (ACP de Poisson)

Pour construire l'indice de créativité, nous utilisons une analyse en composantes principales de Poisson (ACP de Poisson). Contrairement à l'ACP standard, qui suppose des variables continues, l'ACP de Poisson modélise de manière plus précise la nature discrète et non négative des données de comptage. Cette approche est adaptée à la mesure de créativité, qui capture le nombre de réponses produites par les élèves.

B.4 Agrégation des scores latents

Les scores latents $(\hat{\theta}_{j1}, \hat{\theta}_{j2}, \dots, \hat{\theta}_{jn})$ sont estimés par la méthode de Bayes empirique, qui calcule les scores à partir des réponses observées des élèves aux items et des paramètres d'items estimés. Pour tenir compte du biais potentiel introduit par la contraction (*shrinkage*), nous suivons Soland et al. (2024) et estimons les traits latents à l'aide d'un modèle TRI à deux groupes. Ce modèle permet aux groupes traité et de contrôle d'avoir des distributions de niveau distinctes, tout en maintenant les paramètres d'items fixes.

Pour comparer les scores entre les évaluations de début et de fin d'année, nous utilisons une procédure standard d'ancrage TRI. Plus précisément, nous estimons les paramètres d'items

à partir du groupe de contrôle en fin d'année, puis maintenons ces paramètres fixes lors du scoring des données de début d'année. Cette approche garantit que les scores sont placés sur une échelle commune et que les évolutions dans le temps sont interprétables.

Après estimation des scores latents individuels $(\hat{\theta}_{j1}, \hat{\theta}_{j2}, \dots, \hat{\theta}_{jn})$ à partir de chaque modèle TRI, nous construisons des indices composites en calculant la moyenne pondérée par l'inverse de la matrice de covariance des scores composants. Plus précisément, nous calculons la matrice de covariance inverse des scores composants, normalisons ces pondérations pour qu'elles somment à un, et les utilisons pour calculer une moyenne pondérée :

$$\text{Index}_j = \sum_{i=1}^n w_i \hat{\theta}_{ji}, \quad \text{où } w_i \propto \frac{1}{\text{Var}(\hat{\theta}_i)}$$

Tableau B1 : Propriétés psychométriques

	Nombre d'items			Théorie classique des tests (TCT)			Théorie de la réponse à l'item (TRI)			
	Effectifs (1)	Exclus (2)	Ancre (3)	% correct (4)	% NR (5)	Alpha de Cronbach (6)	Corr. moyenne item-test (7)	Discrimination moyenne (8)	Difficulté moyenne (9)	Fiabilité conditionnelle moyenne (fin d'année) (10)
Panneau A : Évaluations										
Arabe	75	1	52	45.31	0.056			1.16	0.191	0.897
Français	56	1	36	25.12	4.56			1.22	1.01	0.889
Mathématiques	32	0	11	27.93	7.96	0.787	0.369	0.888	1.18	0.801
Sciences	34	0	10	35.74	2.79	0.765	0.339	0.744	1.02	0.781
Physique-Chimie	19	0	5	37.18	3.37	0.631	0.362	0.740	1.35	0.667
Sciences de la vie	15	0	5	41.14	2.04	0.639	0.408	0.848	0.550	0.652
Panneau B : Autres mesures										
Compétences interpersonnelles										
Prosocialité	5	0						1.34		0.599
Perception des émotions	16	1						0.529		0.505
Compétences intrapersonnelles										
Sentiment de contrôle	13	0						0.840		0.723
Mentalité de croissance	3	0						1.91		0.720
Lieu de contrôle	5	0						1.09		0.434
Sentiment d'auto-efficacité	5	0						1.18		0.627
Auto-régulation et discipline	25	0						0.886		0.851
Discipline de travail et persévérance	17	0						0.998		0.821
Maîtrise de soi	8	0						0.846		0.677
Créativité	7	0								0.791
Bien-être										
Indice de bien-être	18	0						0.995		0.737
Sentiment d'appartenance à l'école	5	0						1.15		0.607
Harcèlement (inversé)	9	0						1.54		0.676
Stress perçu (inversé)	4	0						0.886		0.458

Notes. Échantillon et unité d'observation : 10 514 élèves évalués dans les 200 collèges de l'étude. Ce tableau présente les propriétés de mesure des instruments d'évaluation et d'enquête auprès des élèves inclus dans la collecte de fin d'année de l'étude. "Ancre" désigne les items administrés à la fois en début et en fin d'année, utilisés pour relier les deux vagues sur une échelle commune. "NR" désigne les non-réponses. La discrimination moyenne et la difficulté moyenne renvoient aux paramètres moyens de discrimination et de difficulté issus du modèle TRI utilisé pour chaque instrument (un modèle logistique à deux paramètres pour les évaluations et un modèle de réponse graduée pour les échelles d'enquête de type Likert). La fiabilité conditionnelle moyenne représente la moyenne des fiabilités individuelles calculées sur l'échantillon de fin d'année, sur la base du niveau de compétence estimé de chaque répondant et de l'erreur-type prédite associée.

Références

- Jacob, B., Rothstein, J., 2016. The Measurement of Student Ability in Modern Assessment Systems. *Journal of Economic Perspectives* 30, 85–108. doi :10.1257/jep.30.3.85.
- Soland, J., Kuhfeld, M., Edwards, K., 2024. How survey scoring decisions can influence your studys results : A trip through the IRT looking glass. *Psychological Methods* 29, 1003–1024. doi :10.1037/met0000506.

C Ajustements pré-spécifiés pour les tests d'hypothèses multiples

En tenant compte d'une hiérarchie pré-enregistrée d'hypothèses de recherche, nous ajustons pour les tests d'hypothèses multiples en calculant les q -valeurs ajustées par le taux de fausses découvertes corrigé. Suivant Vivalt et al. (2024), nous organisons nos hypothèses en niveaux (notés K0, K1 et K2), correspondant à notre ordre de priorité des tests.

K0 : Famille de résultats principaux : compétences intra- et interpersonnelles. Nous calculons les q -valeurs pour ces deux variables (ensembles de compétences), que nous désignons comme nos deux items K0. Ce sont nos résultats de plus haute priorité.

K1 : Famille de résultats finaux : compétences académiques et abandon scolaire. Ce sont nos résultats de deuxième priorité. Nous calculons les q -valeurs pour ces deux variables (résultats), que nous désignons comme nos deux items K1, en combinaison avec nos deux items K0.

K2 : Famille de résultats supplémentaires : créativité, bien-être et habitudes de travail. Ce sont nos résultats de troisième priorité. Nous calculons les q -valeurs pour ces trois variables (résultats), que nous désignons comme nos trois items K2, en combinaison avec nos quatre items K0 et K1.

En plus de ces trois niveaux de tests par famille, nous explorons également les effets au niveau des composantes et sous-composantes sous-jacentes. Par exemple, la "famille" des compétences intrapersonnelles comprend deux "composantes" : un indice de contrôle perçu et un indice d'autodiscipline. À son tour, la composante de contrôle perçu comprend trois sous-composantes : une mesure de l'état d'esprit de développement (*growth mindset*), une mesure du lieu de contrôle et une mesure du sentiment d'auto-efficacité perçu. Pour donner un autre exemple, la "famille" des compétences académiques comprend quatre composantes par matière (arabe, français, mathématiques, sciences), qui peuvent être subdivisées en sous-composantes (par exemple, deux indices d'items de mathématiques capturant respectivement le contenu correspondant au niveau scolaire et le contenu en dessous du niveau scolaire). Les q -valeurs pour les estimations au niveau des composantes sont calculées en utilisant les items de niveau famille jusqu'au niveau de famille considéré (K0, K1 ou K2), en plus des items au niveau des composantes appartenant à la même famille que le résultat. Par exemple, les q -valeurs pour les effets sur les compétences en arabe tiennent compte des quatre comparaisons au niveau famille des niveaux K0 et K1, ainsi que des quatre comparaisons au niveau des composantes (arabe, français, mathématiques et sciences).

Au sein de chaque niveau de famille, et à chaque niveau (famille, composante ou sous-composante), nous explorons les effets de traitement parmi deux sous-groupes : les élèves identifiés comme étant à risque d'abandon scolaire, et les filles. Entre ces deux sous-groupes, nous accordons la priorité aux tests portant sur le sous-groupe des élèves à risque. Les q -valeurs pour ces comparaisons sont calculées en utilisant l'ensemble des items de cette famille, jusqu'au niveau considéré, en plus des tests d'effets hétérogènes. Par exemple, les tests portant sur les effets sur les compétences en arabe parmi les élèves à risque d'abandon scolaire tiennent compte des quatre comparaisons aux niveaux K0 et K1, des quatre comparaisons au niveau des composantes (arabe, français, mathématiques et sciences), ainsi que de quatre tests supplémentaires par sous-groupe (un par matière). Les tests portant sur les effets sur les compétences en arabe parmi les filles tiennent compte des quatre comparaisons aux niveaux K0 et K1, des quatre comparaisons au niveau des composantes (arabe, français, mathématiques et sciences), des quatre tests par sous-groupe pour les élèves à risque d'abandon scolaire, ainsi que de quatre tests supplémentaires par sous-groupe (un par matière).

Nos analyses de la prise en charge de l'intervention et de la fidélité de mise en uvre reposent en grande partie sur des statistiques descriptives, et nous n'en tenons pas compte dans nos ajustements pour les tests d'hypothèses multiples.

Références

Vivalt, E., Rhodes, E., Bartik, A.W., Broockman, D.E., Krause, P., Miller, S., 2024. The Employment Effects of a Guaranteed Income : Experimental Evidence from Two U.S. States. doi :10.3386/w32719.