



HAL
open science

Développer, expérimenter et évaluer des dispositifs pédagogiques innovants : l'exemple du Créativlab d'Ampiric et la question du passage à l'échelle

Nicolas Mascret, Thierry Berthet, Pascale Brandt-Pomares, Magali Coupaud, Mathilde Favier, Jonathan Mirault, Johannes C Ziegler

► To cite this version:

Nicolas Mascret, Thierry Berthet, Pascale Brandt-Pomares, Magali Coupaud, Mathilde Favier, et al.. Développer, expérimenter et évaluer des dispositifs pédagogiques innovants : l'exemple du Créativlab d'Ampiric et la question du passage à l'échelle. ANRT. L'évaluation des dispositifs pédagogiques innovants, ANRT, 2023, 978-2-9591124-1-6. halshs-04360207

HAL Id: halshs-04360207

<https://shs.hal.science/halshs-04360207v1>

Submitted on 8 Jan 2024

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

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

DÉVELOPPER, EXPÉRIMENTER ET ÉVALUER DES DISPOSITIFS PÉDAGOGIQUES INNOVANTS : L'EXEMPLE DU CRÉATIVLAB D'AMPIRIC ET LA QUESTION DU PASSAGE A L'ÉCHELLE

Nicolas Mascret^{1,2,5,7}, Thierry Berthet^{1,6,7}, Pascale Brandt-Pomares^{1,2,3,7}, Magali Coupaud^{1,2,3,7}, Mathilde Favier^{1,2,7}, Jonathan Mirault^{1,2,4,7}, Johannes C. Ziegler^{1,4,7}

¹Pôle Pilote Ampiric

²Institut National Supérieur du Professorat et de l'Éducation d'Aix-Marseille, AMU

³Laboratoire ADEF (Apprentissage Didactique Évaluation Formation), UR4671, Marseille

⁴Laboratoire LPC (Laboratoire de Psychologie Cognitive), UMR7290, AMU-CNRS, Marseille

⁵Laboratoire ISM (Institut des Sciences du Mouvement), UMR7287, AMU-CNRS, Marseille

⁶Laboratoire LEST (Laboratoire d'Économie et de Sociologie du Travail), UMR7317, AMU-CNRS, Aix-en-Provence

⁷Fédération SFERE-Provence (Fédération de recherche en éducation), FED 4238, AMU

Le Pôle Pilote Ampiric est une opération soutenue par l'État dans le cadre de l'action « Territoires d'innovation pédagogique » du Programme d'investissements d'avenir et du Plan France 2030, opéré par la Caisse des Dépôts.

Notice biographique des auteurs

Nicolas Mascret

Nicolas Mascret est professeur des universités à l'Inspé d'Aix-Marseille. Il effectue ses recherches en psychologie sociale et en psychologie sociale expérimentale au sein de l'Institut des Sciences du Mouvement de Marseille (ISM, UMR7287, AMU-CNRS), dont il est directeur-adjoint. Elles portent notamment sur la motivation, le stress / l'anxiété et l'acceptabilité des technologies, qui sont étudiés dans le domaine scolaire et dans le domaine du sport et de l'activité physique. Nicolas Mascret est le Responsable Scientifique et Technique d'Ampiric.

Thierry Berthet

Thierry Berthet est politiste, directeur de recherche CNRS et directeur du Laboratoire d'Économie et de Sociologie du Travail (LEST). Ses travaux portent sur les politiques comparées de lutte contre le décrochage scolaire, l'action publique en matière d'orientation, la territorialisation des politiques publiques et les politiques de l'emploi. Ses travaux sur le décrochage scolaire portent notamment sur l'expérience des dispositifs de remédiation par les jeunes décrocheurs et les opérateurs de terrain. Il pilote le Dispositif de Suivi et d'Évaluation d'Ampiric.

Pascale Brandt-Pomares

Pascale Brandt-Pomares est professeure en sciences de l'éducation et de la formation, Vice-Présidente déléguée au Centre d'Innovation Pédagogique et d'Évaluation d'Aix-Marseille Université et Directrice de l'Inspé d'Aix-Marseille. Didacticienne des enseignements technologiques, elle s'intéresse aux processus d'enseignement-apprentissage. Ses travaux portent en particulier sur les rôles et fonctions du numérique dans l'activité enseignante. Elle est co-responsable du programme de recherche Artifacts de formation de l'UR 4671 ADEF et pilote l'Action 1 d'Ampiric sur la formation initiale et continue des enseignants.

Magali Coupaud

Magali Coupaud est ingénieure de recherche pour Ampiric en charge de l'ingénierie de formation et membre du laboratoire ADEF (Apprentissage, Didactique, Évaluation, Formation) d'Aix-Marseille Université. Elle s'intéresse à la didactique de la biologie, et en particulier aux enjeux didactiques de l'enseignement de l'évolution du vivant.

Mathilde Favier

Mathilde Favier est cheffe de projet du Pôle Ampiric. Elle en coordonne le pilotage stratégique, opérationnel, et scientifique. Elle s'interroge sur la mise en œuvre de l'interdisciplinarité en éducation, et la facilitation de conditions propices à un dialogue constructif et apaisé entre les différentes traditions disciplinaires autour des questions d'éducation.

Jonathan Mirault

Jonathan Mirault est ingénieur de Recherche en Expérimentation et Analyse de données au sein d'Ampiric. Il conçoit et met en place des expériences scientifiques, collecte et analyse les données recueillies sur les projets déployés par le pôle pilote. Il se focalise particulièrement sur l'utilisation de méthodologies telles que l'oculométrie, l'électro-encéphalographie, la réalité virtuelle, la chronométrie mentale et la psychophysique.

Johannes Ziegler

Johannes Ziegler est chercheur en sciences cognitives, Directeur de Recherches au CNRS, Directeur du Laboratoire de Psychologie Cognitive, Directeur de l'Institut Convergences « Langage, Communication et Cerveau » à l'Université d'Aix-Marseille. Il coordonne les actions de recherches fondamentales au sein d'Ampiric. Ses recherches portent sur l'apprentissage de la lecture, ses dysfonctionnements et ses bases neurales ainsi que le développement et l'évaluation d'outils d'entraînement et de remédiation. Il est membre du Conseil Scientifique de l'Éducation Nationale (CSEN).

Résumé

Le développement, l'expérimentation et l'évaluation des dispositifs innovants s'inscrivent dans un processus complexe lorsqu'ils sont réalisés selon une démarche scientifique rigoureuse et lorsque l'ambition est d'envisager leur passage à l'échelle. En appui sur le CréativLab du Pôle Pilote Ampiric, les procédures et modalités d'évaluation de dispositifs pédagogiques innovants, issus d'une collaboration entre entreprises, chercheurs et enseignants, seront présentées. Dans un second temps, les enjeux du passage à l'échelle de ces dispositifs seront examinés dans le contexte de l'éducation fondée sur les données probantes, tout en identifiant les limites liées à la pérennisation / généralisation de ces dispositifs et les limites liées à l'influence des prescriptions chez les enseignants.

Mots clés

Innovation - Procédures d'évaluation - Suivi de projet - Éducation fondée sur les données probantes - Niveaux de preuve - Généralisation - Prescriptions

INTRODUCTION

L'efficacité d'un nouveau dispositif pédagogique ne va pas de soi. Ce n'est pas parce qu'il est dit « innovant » qu'il est par définition « plus efficace ». L'évaluation et l'expérimentation sont donc cruciales pour connaître la valeur d'un dispositif dans la vie réelle à l'échelle d'une classe, d'une école, d'une académie ou même d'un pays tout entier.

Pourquoi expérimenter et évaluer des dispositifs pédagogiques innovants ?

Le mot « expérimenter » est souvent utilisé dans un sens assez vague, celui d'essayer, sur un groupe restreint, quelque chose qui sort de l'ordinaire. Cette idée de l'expérimentation dans le quotidien est assez éloignée de l'expérimentation qui consiste à mettre à l'épreuve une hypothèse à l'aide de méthodes susceptibles de contrôler les biais. Expérimenter permet de tester la véracité d'une hypothèse, ou de déterminer l'efficacité d'une méthode pédagogique. Avant chaque expérimentation, les chercheurs comme les enseignants doivent veiller à la pertinence de l'intervention proposée, penser à son éventuel passage à l'échelle et à son adoptabilité, aussi bien qu'à son impact potentiel. La conception expérimentale d'une recherche fait partie intégrante des aspects éthiques, car il n'est jamais éthique de mener une recherche dont on sait à l'avance qu'elle ne permettra pas de tirer la moindre conclusion solide, par exemple parce qu'elle comprend des biais importants, qu'aucun groupe contrôle n'est prévu, ou qu'elle concerne trop peu d'enfants pour être significative. Quelle serait en effet l'alternative à l'expérimentation ? La prise de décision sur la base de simples intuitions, sans savoir si elles sont fondées ? Certains s'opposent à l'idée même d'expérimentation en éducation pour des raisons faussement éthiques. Mais les enseignants testent tous les jours de nouvelles pratiques sur leurs élèves, sans que personne ne s'y oppose. Malheureusement, ces tentatives se font selon une méthodologie qui ne permet pas de tirer de conclusions rigoureuses, de diffuser les avancées et ne permet donc pas de faire avancer la connaissance au bénéfice de tous. Le Pôle Pilote Ampiric (**Aix-Marseille – Pôle d'Innovation, de Recherche, d'enseignement pour l'éducation**) est né de la volonté de rapprocher les acteurs du domaine éducatif, qu'ils soient enseignants, scientifiques ou entrepreneurs dans un seul et même but : améliorer l'éducation des enfants afin de maximiser leur bien-être, leurs connaissances et leurs compétences.

Le Pôle Pilote Ampiric

Financé par le Programme d'Investissement d'Avenir (PIA3) et s'inscrivant dans la stratégie France 2030, Ampiric est un Pôle Pilote porté par Aix-Marseille Université en partenariat avec l'académie d'Aix-Marseille, Avignon Université, l'Université Côte d'Azur, le CNRS, l'Association Nationale de la Recherche et de la Technologie (ANRT), le Réseau Canopé, et la Région Sud. Le Pôle Pilote Ampiric (dénommé Ampiric dans la suite de ce chapitre) vise à améliorer durablement l'apprentissage des savoirs fondamentaux des élèves tout au long de leur parcours scolaire, notamment de ceux qui sont en difficulté. Ampiric se fonde sur le développement de travaux de recherche innovants, pour mieux informer l'évolution de la formation initiale et continue des enseignants, et développer, expérimenter, diffuser des approches pédagogiques innovantes. L'écosystème Ampiric se fonde ainsi sur une meilleure articulation entre la recherche, la formation et le terrain.

Ampiric poursuit cinq objectifs stratégiques :

- Transformer la formation et l'accompagnement des enseignants pour faire évoluer leurs pratiques face aux défis liés aux apprentissages fondamentaux ;
- Favoriser les coopérations recherche-terrain en soutenant les recherches collaboratives ;
- Améliorer la compréhension des processus d'enseignement-apprentissage des savoirs fondamentaux en soutenant des recherches fondamentales, finalisées par les besoins de l'éducation et de la formation ;
- Coproduire des outils et des applications au service des apprentissages fondamentaux au sein d'un CréativLab ;
- Diffuser et valoriser les ressources pédagogiques sur les apprentissages fondamentaux.

Le CréativLab d'Ampiric

Dans le cadre de ce chapitre, nous nous centrerons plus spécifiquement sur le quatrième objectif stratégique afin d'investiguer l'évaluation des dispositifs pédagogiques innovants dans le cadre du CréativLab d'Ampiric. Le CréativLab est un espace de coopération de recherche et d'action dédié aux enseignants, chercheurs et entreprises. Coordonné par l'ANRT, et à travers elle, son groupe de travail sur la « Pédagogie par le jeu », il bénéficie du soutien et de l'accompagnement de tous les partenaires d'Ampiric. En appui sur la recherche, l'objectif principal du CréativLab est de coproduire, expérimenter et labelliser des ressources pédagogiques permettant d'améliorer les apprentissages fondamentaux.

Le fonctionnement du CréativLab est articulé autour des actions suivantes :

- **L'initiation et l'accompagnement de projets collaboratifs.** Chaque année, des projets portés conjointement par des responsables d'entreprise, des chercheurs et des enseignants sont sélectionnés, au service de l'amélioration de l'apprentissage des savoirs fondamentaux.
- **La mise en réseau des acteurs, entre terrain, formation, recherche et entrepreneuriat.** Diverses manifestations sont organisées pour opérationnaliser cette mise en réseau.
- **Une évaluation des ressources produites ou perfectionnées.** Les outils pédagogiques développés ont vocation à être évalués auprès de classes partenaires au sein du plateau expérimental « Jacques Ginesté » d'Ampiric ou directement dans les établissements scolaires.

L'ambition de ce chapitre est donc double : d'une part présenter les modalités d'évaluation des projets mises en œuvre dans le CréativLab d'Ampiric et d'autre part identifier dans quelle mesure cette évaluation spécifique interroge de façon plus transversale toutes les modalités d'évaluation des dispositifs pédagogiques innovants.

L'EVALUATION DES PROJETS DANS LE CREATIVLAB D'AMPIRIC

L'évaluation des projets au sein du CréativLab répond à un double enjeu : accompagner des projets de recherche partenariale, co-portés par des chercheurs et des acteurs du monde socio-économique en lien avec l'éducation (entreprises, associations, parents d'élèves, enseignants), et évaluer les ressources produites ou développées afin

d'envisager un éventuel passage à l'échelle. Deux niveaux d'évaluation des projets sont identifiés : le premier correspond à la procédure de sélection des projets qui intègrent le CréativLab et le second correspond au suivi et à l'évaluation des projets retenus.

La procédure de sélection des projets

Le premier niveau d'évaluation, visant à initier les collaborations, suit une procédure de sélection en deux temps : tout d'abord, les acteurs répondent à un appel à manifestation d'intérêt (AMI) pour soumettre leur projet. Cet appel ne fait pas l'objet d'une sélection particulière, bien que les projets doivent répondre aux objectifs du CréativLab, et plus largement à ceux d'Ampiric, à savoir l'amélioration de l'apprentissage des savoirs fondamentaux des élèves. L'ensemble des projets déposés dans le cadre de cet AMI sont présentés lors d'un « speed-dating » organisé annuellement, afin de permettre aux acteurs du monde socio-économique de rencontrer des acteurs de la recherche et de réfléchir ensemble à une collaboration potentielle. Cette collaboration se concrétise par le dépôt d'un projet commun dans le second temps de l'appel. Les projets déposés font l'objet d'une double évaluation conduite par des chercheurs du périmètre d'Ampiric en fonction des disciplines et thématiques concernées. Cette évaluation porte sur les éléments suivants :

- la qualité scientifique du projet ;
- les perspectives de succès de la recherche partenariale ;
- l'implication de l'entreprise et la pertinence du projet dans son contexte entrepreneurial ;
- l'adéquation avec le laboratoire partenaire et sa pertinence dans le projet ;
- la participation des enseignants et la pertinence du projet au service de l'amélioration de l'apprentissage des savoirs fondamentaux.

Au terme de cette évaluation, les rapporteurs doivent émettre un avis sur l'opportunité de recevoir les projets pour une audition. Cette dernière permet de clarifier certains aspects liés au projet qui n'auraient pas été précisés dans le dossier écrit, et constitue la dernière phase de ce processus de sélection. Les projets doivent recevoir deux avis favorables pour être reçus en audition, lors de laquelle la commission de sélection émet un avis final sur l'opportunité de soutenir et d'accompagner le projet dans le cadre du CréativLab.

Ce processus de sélection très rigoureux a été élaboré pour s'assurer de la mise en place de conditions favorables au bon déploiement des projets, en termes de déroulement et d'atteinte des objectifs. Financés par de l'argent public dans le cadre du Plan France 2030, les projets soutenus doivent en effet répondre aux exigences fixées par les objectifs du Pôle Ampiric relatifs à l'évolution de la professionnalité des enseignants et de l'amélioration de l'apprentissage des élèves. Pour ce faire, la problématique de recherche doit être pertinente et explicitée, et les conditions propices à une collaboration fructueuse entre les chercheurs et leurs partenaires doivent être réunies. Des zones d'ombre sur les modalités de collaboration ou sur l'implication et la motivation d'un partenaire à s'engager dans un projet, ou des incertitudes concernant des questions de méthodologie, de validation des résultats et des impacts de l'outil sur les apprentissages, sont autant de facteurs qui ne permettront pas aux projets d'être retenus. À l'inverse, une méthodologie expérimentale rigoureuse et clairement explicitée, contrôlant un maximum de paramètres et en gardant une situation

écologique (en salle de classe par exemple), l'articulation avec les recherches actuelles, la plus-value de tel ou tel laboratoire en termes de compétences et d'approche scientifique et une collaboration active et pertinente entre la recherche et son partenaire sont autant d'éléments qui permettront aux projets de devenir lauréats.

Le suivi et l'évaluation des projets

Le suivi et l'évaluation des projets retenus forment le deuxième niveau du processus mis en œuvre par le CréativLab. Ainsi il est demandé aux projets lauréats de compléter un bilan annuel intermédiaire et de présenter leurs avancées lors des « speed-datings » des années suivantes. Ce bilan intermédiaire permet de faire le point sur une mise à jour éventuelle des acteurs impliqués, de faire part des différentes avancées du projet au regard de ses objectifs initiaux, d'indiquer si ces derniers ont dû faire l'objet d'une réorientation, de signaler des premiers résultats potentiels, et de partager les difficultés rencontrées dans l'organisation de la collaboration ou tout autre aspect du déploiement du projet. Le suivi des projets permet également de mettre en évidence que les critères d'évaluation ne sont pas forcément identiques tout au long des différentes étapes du développement du projet, comme l'illustre le projet de recherche basé sur le jeu Darwinium (**Encadré 1**). Ces rapports permettent au comité de suivi des projets lauréats de juger du respect des engagements initiaux par les différentes parties, et d'adresser d'éventuels problèmes. Selon les avancées, le comité se réserve le droit de ne plus soutenir des projets qui seraient jugés non conformes aux objectifs énoncés. Ainsi, certains projets ont dû être arrêtés à la suite du départ de certains collaborateurs (chercheurs ou équipes pédagogiques) ou à cause de difficultés techniques ou technologiques dans le déploiement des outils proposés.

Ce processus présente plusieurs avantages. Il permet tout d'abord de garantir la bonne utilisation et/ou l'optimisation des moyens mis à disposition, qu'ils soient financiers, humains ou matériels. Il assure par ailleurs l'adhésion collective de tous les acteurs et le bon déroulement de la collaboration, en garantissant la possibilité de se retirer d'une collaboration qui ne donnerait pas satisfaction. Cela est particulièrement important dans le développement de dispositifs pédagogiques innovants pour deux raisons : la première concerne le temps extrêmement précieux des équipes pédagogiques de terrain, qui doivent pouvoir trouver une utilité aux dispositifs mis en œuvre en relation avec leur pratique professionnelle. La seconde répond au devoir de vigilance que l'on se doit d'avoir à l'égard des élèves impliqués dans le déploiement des dispositifs, afin que ceux-ci ne soient pas délétères pour leur expérience éducative. Enfin, le suivi rigoureux des projets contribue au développement d'une culture commune au regard de critères de réussite constitutifs des projets soutenus. À mesure du déploiement du second niveau de suivi et d'évaluation, ces critères pourront venir constituer la base d'un référentiel visant à permettre la labélisation de certaines ressources et/ou outils dans le cadre du développement d'une filière française spécialisée dans le développement de dispositifs pédagogiques ludiques innovants.

Encadré 1. Développement, expérimentation et évaluation du jeu Darwinium

Le projet de recherche présenté ici repose sur la conception du jeu Darwinium intégré à des ressources pédagogiques destinées aux enseignants de collège, afin d'aborder les concepts de hasard et d'évolution. Environné par Ampiric, il s'inscrit dans le

cadre du projet ANR EGRESS (Educational Game about Randomness and Evolution for Students in Science). Ce jeu place des élèves de collège dans la situation de chercheurs observant l'évolution de populations chimériques placées dans un dôme expérimental et devant rendre compte de cette évolution sous forme verbale narrative (explication écrite et orale de l'évolution) et graphique (représentation de la population). Ce projet s'inscrit dans la continuité de travaux qui ont souligné les difficultés d'appréhension de l'évolution du vivant et des idées de hasard qui y sont liées (Fielder et al., 2019). L'objectif de cet encadré est de mettre en évidence la diversité des choix méthodologiques retenus pour guider les étapes de conception du jeu et mesurer son efficacité.

La conception du jeu « Darwinium » s'inscrit dans un projet plus large de méta-conception du jeu intégré à une intervention didactique en classe (Ardito et al., 2015). Le principe général de ce processus de méta-conception est d'intégrer différents acteurs, comme des experts (didacticiens, spécialistes de l'évolution, ou joueurs experts), des utilisateurs (comme des enseignants, des élèves), dans le processus de conception même. L'activité de conception implique des cycles itératifs de développement avec réalisation de différents prototypes (éléments et mécaniques du jeu différents) et feedbacks (Bonnardel & Didier, 2016). A différentes phases de test du jeu, plusieurs données sont collectées en contexte d'activité de classe ou non : verbatim, récits et dessins d'utilisateurs. La méthodologie est basée sur des analyses sémio-cognitives avec une attention particulière accordée aux spécificités d'une activité créative, aux interactions entre différents experts dans le processus de conception, à la place donnée aux utilisateurs finaux (enseignants et élèves) dans le processus de conception, et au retour d'expérience sur l'effet du jeu par rapport aux différents objectifs.

La deuxième partie du projet vise à mesurer l'impact du dispositif conçu sur le raisonnement évolutionniste de collégiens. L'étude en cours sera menée en situation écologique de classe de troisième lors d'un enseignement des SVT sur un large échantillon. Les mesures seront réalisées selon une méthode de pré-test / post-test, à partir d'un questionnaire fermé (Coupaud et al., 2019). Au sein d'un vingtaine de classes, un groupe contrôle et un groupe test seront constitués. La méthodologie est basée sur des analyses statistiques multivariées permettant une approche comparative des résultats.

Les procédures de sélection des projets dans le CréativLab et de suivi des projets retenus sont des illustrations, au sein d'Ampiric, du processus de développement, d'expérimentation et d'évaluation de dispositifs pédagogiques innovants. Toutefois, bien que spécifiques, ces procédures interrogent au-delà du CréativLab et interpellent la question du passage à l'échelle, transversale à toute évaluation de dispositifs pédagogiques innovants. En effet, la validation des outils ainsi développés doit faire l'objet de différents niveaux de preuve afin de permettre leur passage à une échelle plus large, et à terme aboutir à un transfert éventuel vers le marché. Au-delà du CréativLab d'Ampiric, les enjeux et les limites du passage à l'échelle des dispositifs pédagogiques innovants en contexte d'éducation se doivent alors d'être examinés de façon transversale.

LE PASSAGE À L'ÉCHELLE : ENJEUX ET LIMITES

L'éducation fondée sur les données probantes

Se développant au début du XX^{ème} siècle, les bonnes pratiques scientifiques basées sur les niveaux de preuves voient le jour petit à petit dans le monde de la recherche. De nos jours, c'est un ensemble de pratiques qui tend à se systématiser pour optimiser l'impact de la recherche dans nos sociétés. Le rapport de la Haute Autorité de Santé (HAS) publié en 2013 a défini le niveau de preuve d'une étude comme étant la capacité de celle-ci à répondre à une question précise, en appui sur les quatre critères suivants : i) adéquation du protocole à la question, ii) existence de biais dans la réalisation, iii) adaptation de l'analyse statistique aux objectifs et iv) puissance statistique et taille de l'échantillon.

L'ensemble de ces bonnes pratiques basées sur les niveaux de preuves (*evidence-based practices* en anglais) est conceptualisé de manière globale mais elles se spécifient en fonction du domaine de prédilection de l'étude en question. Ainsi l'approche « éducation fondée sur des données probantes » ou *evidence-based education* apparaît dans les années 90 (Barrett, 1991). Elle réside dans des pratiques et des politiques éducatives préconisant le fondement des études sur les données les plus probantes produites par la recherche (Saussez & Lessard, 2009) afin de limiter l'influence des modes, des croyances et des idéologies (Davies, 1999 ; Slavin, 2002 ; 2008) dans le monde de l'éducation. Cette démarche qui se veut être un outil d'aide à la décision scientifique et rationnelle pour les praticiens et les politiques est fondée sur quatre étapes : i) formuler une question précise, ii) effectuer une revue systématique de recherche, iii) évaluer la qualité des preuves recueillies en se basant sur un système de production de la preuve, l'essai contrôlé randomisé et iv) formuler une réponse claire à la question posée, établir si la solution fonctionne et rendre ces preuves disponibles (Davies, 1999 ; Gough, 2007 ; Oakley, 2004).

L'ambition du passage à l'échelle s'inscrit donc dans un mouvement plus large qui accorde un rôle important à l'utilisation des méthodes scientifiques afin d'aider à faire progresser les pratiques pédagogiques. Ce mouvement est par exemple mis en avant en France par le Conseil Scientifique de l'Éducation Nationale (CSEN) ou en Angleterre par l'*Education Endowment Foundation* (**Encadré 2**).

Encadré 2. Education Endowment Foundation (EEF)

Le secrétaire d'État britannique à l'éducation, Michael Gove, a annoncé fin 2010 son intention de créer un programme d'aide pour améliorer les conditions dans les écoles difficiles en Angleterre et c'est en 2011 que l'EEF est fondé par une association caritative, The Sutton Trust, en partenariat avec Impetus. Le programme EEF est alors doté d'une subvention du ministère de l'éducation de 125 millions de livres sterling. Il emboîte le pas du mouvement « What Works » aux États-Unis et au Royaume-Uni (White, 2019).

Ce fonds lance chaque année de nombreuses expérimentations en éducation ayant pour but l'identification des pratiques et des aménagements efficaces et répondant le mieux possible aux objectifs définis par les acteurs de l'éducation. Avant chaque expérimentation, l'EEF réalise d'abord une revue de la littérature scientifique afin de

savoir si une intervention est justifiée par des connaissances fondamentales ou empiriques. Puis, l'intervention est testée en situation réelle en utilisant principalement des paradigmes en essais contrôlés randomisés (ECR) avec randomisation des groupes témoin et expérimentaux. L'expérimentation est caractérisée par plusieurs étapes :

1/ Dans un premier temps, les enseignants et les chercheurs co-construisent ensemble une intervention qui est à la fois faisable et souhaitable du point de vue des pratiques éducatives des enseignants et de celui de la recherche ;

2/ Dans un deuxième temps, les enseignantes et les chercheurs mènent une étude pilote sur un petit échantillon afin de savoir si les conditions sont réunies pour que l'intervention puisse faire l'objet d'un ECR ;

3/ Ensuite, une première expérimentation est réalisée sur un échantillon plus large de classes, mais en conditions idéales car l'intervention est menée par les chercheurs et enseignants qui ont conçu l'intervention ;

4/ Si cette première expérimentation produit des résultats positifs, l'intervention est alors testée à grande échelle par des enseignants et des expérimentateurs totalement indépendants de l'équipe initiale, après avoir formé le personnel censé la mettre en place dans la réalité.

Lorsque les résultats sont positifs, afin de multiplier les chances d'adoption des bonnes pratiques, EEF produit des boîtes à outils d'indications pratiques pour enseignants, propose des formations, un accompagnement, et la création de réseaux pour la dissémination des résultats et l'échange des pratiques. L'EEF a démontré qu'il est aujourd'hui possible de mener de nombreuses expérimentations en éducation, en utilisant principalement des paradigmes en ECR.

Cependant, les premières tentatives de l'EEF ont souffert d'un taux d'acceptabilité relativement bas, avec une moyenne de 24 % d'élèves qui ont abandonné entre 2011 et 2012, mais, depuis, l'EEF a introduit de nouvelles stratégies pour recruter et retenir les écoles avec par exemple, un développement des communications sur les avantages des ECR dans les écoles par le biais d'événements de recrutement et de diffusion de documents expliquant clairement la conception et l'évaluation des études mais aussi le rôle central des écoles (Dawson et al., 2017).

En termes d'évaluation, l'EEF a nommé, par le biais d'appel d'offres, un panel d'organismes de recherche spécialisés dans l'évaluation de l'éducation, qui se mettraient en concurrence pour évaluer les programmes et seraient ensuite associés aux développeurs. Un point d'honneur a été mis en avant : publier toutes les conclusions, quel que soit le résultat (Nevill, 2016). Certains des premiers essais de l'EEF ont souffert de l'absence d'évaluation de qualité, ce qui signifie qu'ils n'étaient pas en mesure d'expliquer les processus causaux qui sous-tendaient les résultats ou de décrire leur mise en œuvre (Morris et al., 2016). Début 2012, l'EEF a publié des directives strictes sur la sélection des tests dans ses évaluations, les principaux critères étant qu'ils devaient avoir une large validité externe et être fortement corrélés aux performances dans les évaluations nationales à fort enjeu (The EEF, 2012). Grâce aux efforts réalisés, ce programme a généré un grand nombre de preuves qui aident à identifier ce qui fonctionne et ce qui ne fonctionne pas dans les écoles anglaises (pour une revue plus complète, voir Edovald & Nevill, 2021).

Le passage à l'échelle, tel qu'envisagé par les méthodologies qui se basent sur une éducation fondée sur des données probantes, a pour ambition de s'appuyer sur une collaboration inédite entre les enseignants et les chercheurs dans le but de faire émerger une véritable « recherche translationnelle » en éducation (voir Dehaene et Pasquinelli, 2021). Ce terme trouve son origine en médecine. Son objectif premier est de proposer et d'évaluer des solutions aux problèmes concrets que pose la santé, afin que ces solutions soient fondées sur les meilleures preuves disponibles. Cette approche s'applique également au domaine de l'éducation qui constitue un enjeu sociétal aussi important que la santé. Tout comme la recherche médicale, la recherche en éducation doit aider à identifier des pratiques, des aménagements et des interventions efficaces, qui répondent le mieux possible aux objectifs définis par les acteurs de l'éducation. La recherche translationnelle doit aider à évaluer ce qui marche dans la vie réelle, à l'échelle d'une classe, d'une école, d'une académie ou même d'un pays tout entier. Le terme fait donc référence à l'ensemble des étapes qui permettent de passer d'une connaissance scientifique fondamentale à sa mise en application pratique. Ce processus est progressif : du laboratoire à la salle de classe en passant par une série de recherches sur des cohortes de plus en plus grandes d'élèves, jusqu'à la mise en pratique à grande échelle.

Comme le soulignent Dehaene et Pasquinelli (2021) dans la synthèse du CSEN, « En éducation comme en médecine, il n'est raisonnable de tester une nouvelle intervention que lorsque des recherches préalables indiquent qu'elle a de bonnes chances d'être plus efficace que les meilleures pratiques éducatives actuellement disponibles (présomption d'efficacité). La pré-identification des interventions potentiellement efficaces est indispensable pour éviter les coûts inutiles, aussi bien en termes de temps et d'argent que d'effets indésirables ». Le speed dating d'Ampiric est un bon exemple où les idées des partenaires économiques se confrontent aux regards des scientifiques.

Avant le passage à l'échelle d'un dispositif pédagogique innovant, il existe un continuum de méthodes, de plus en plus sophistiquées, fournissant une hiérarchie de niveaux de preuves (**Encadré 3**).

Encadré 3. La pyramide des preuves

L'approche basée sur les preuves a produit une liste de méthodes de recherche permettant d'apporter des preuves en faveur ou à l'encontre d'une intervention, ordonnées selon leur susceptibilité à des biais de différente nature. Dans son rapport sur la "recherche translationnelle", le CSEN (voir Dehaene et Pasquinelli, 2021) fait la proposition suivante :

Preuves de niveau 0 : Témoignages, questionnaires, interactions entre enseignants, parents, apprenants, chercheurs.

Ce riche tissu d'échanges permet de définir les buts des interventions de façon collégiale, mais aussi de recueillir la sagesse collective des personnes impliquées dans le processus éducatif. Même si les résultats de ces échanges ne fournissent pas une preuve de « ce qui marche », ils sont très importants à prendre en compte, car ils sont la principale source

d'hypothèses sur « ce qui pourrait marcher ».

Preuves de niveau 1 : Connaissances d'ordre fondamental obtenues en laboratoire.

Par ordre croissant de fiabilité pour la recherche translationnelle, on distingue :

- Les connaissances issues de recherches de laboratoire éloignées des conditions réelles, ou avec des échantillons limités ;
- Les connaissances issues de recherches de laboratoire avec conditions proches de celles réelles, ou réalisées sur de nombreux sujets ;
- Les résultats convergents de différentes lignes de recherche, avec réplication et confirmation des résultats, sur de vastes échantillons.

Preuves de niveau 2 : Études observationnelles dans les classes.

Sans modification des conditions initiales, ces études examinent la corrélation entre les pratiques et les progrès des élèves. La qualité de ces études varie selon l'utilisation de supports objectifs pour l'observation (grilles, questionnaires, tests) et d'outils statistiques pour contrôler les autres variables qui pourraient expliquer les résultats.

Preuves de niveau 3 : Études expérimentales avec modification contrôlée des conditions d'apprentissage et observation des conséquences.

Par ordre de fiabilité, on distingue :

- Les études qui comparent un pré-test et un post-test (un design expérimental souvent très insuffisant, car tous les élèves s'améliorent avec le temps) ;
- Les études qui comparent un groupe d'intervention et un groupe témoin en pré- et post-test, mais sans les randomiser ;
- Les études randomisées contrôlées, dans lesquelles les élèves ou les classes sont assignés aléatoirement à la condition d'intervention ou de contrôle, et sont comparés en pré- et post-tests.

Preuves de niveau 4 : Synthèses et méta-analyses.

Résultats répliqués et confirmés par plusieurs groupes de recherche indépendants, idéalement réunis dans des analyses systématiques de la littérature (méta-analyses).

Il est important de souligner que toutes ces méthodes différentes ont un rôle à jouer dans la recherche en éducation. Elles ne sont pas en concurrence mais créent des synergies et soutiennent collectivement la découverte de moyens pour améliorer l'éducation à tous les niveaux. Il serait contre-productif d'affirmer que seule une méthode ou un petit nombre de méthodes ont de la valeur dans le domaine de l'éducation fondée sur les preuves. Par exemple, il est souvent utile de commencer par des études observationnelles, avant de mener des études plus définitives (par études randomisées contrôlées) quant à l'efficacité causale des interventions éducatives.

Très souvent, les idées viennent d'abord des préoccupations du terrain, par exemple d'un enseignant qui souhaite tester une nouvelle idée avec ses élèves. Il est toutefois important d'être conscient que ces observations informelles ne fourniront pas de preuve. Si ces observations informelles suggèrent que l'idée est prometteuse, il est important de mener une véritable expérimentation – il ne s'agit pas encore d'un véritable essai randomisé

contrôlé à grande échelle, mais d'une expérience à petite échelle (quelques classes), avec des mesures objectives des compétences cibles avant et après intervention, et autant que possible un groupe contrôle (voir la méthodologie proposée par EEF). Si cette expérimentation produit des effets espérés, il peut être judicieux de chercher à obtenir le niveau de preuve le plus élevé, en menant un essai randomisé contrôlé à une échelle qui dépasse celle d'un seul établissement. Toutefois, le passage à très grande échelle, à savoir le déploiement d'une nouvelle pratique pédagogique auprès de nombreux enseignants posent d'autres problèmes.

Le passage à l'échelle : la question de la pérennisation / généralisation

L'évaluation d'une expérimentation pose de redoutables questions théoriques et pratiques notamment quand il s'agit de décider de l'opportunité de pérenniser et éventuellement généraliser le dispositif expérimenté. Le développement de logiques d'expérimentation à petite échelle de nouveaux dispositifs en vue de leur possible généralisation tend à s'imposer comme une nouvelle technique de *policy making*. L'évaluation de ces innovations expérimentales a donc une responsabilité particulière puisqu'il s'agit au fond de décider *ex-ante* du bien fondé d'un dispositif et soulève des questions inhabituelles en termes évaluatifs dont la plus prégnante est sans nul doute celle de sa pérennisation/généralisation. Leur évaluation suppose donc de prendre en compte un certain nombre de points de vigilance.

Premier de ces points, les expérimentations s'inscrivent le plus souvent dans l'écosystème local qui les porte et elles visent à répondre à des problématiques elles aussi localisées dont il est souvent difficile de mesurer la validité externe pour d'autres écosystèmes. Il est ainsi plutôt rare que les expérimentations soient construites en intégrant dans leur design la perspective de leur généralisation. Il importe donc dans l'évaluation de bien mesurer le degré d'endogénéité de l'expérimentation en question pour analyser sa transférabilité. C'est là souvent un point critique des évaluations randomisées puisqu'il s'agit, dans le protocole expérimental d'évaluation, de neutraliser tous les biais endogènes possibles lors de la constitution des échantillons (Conseil d'État, 2019).

Second point critique, la représentativité de l'échantillon – ou des échantillons test et témoin en cas d'évaluation randomisée - qui supposerait que celui-ci soit représentatif de la population cible d'un processus de généralisation. Ainsi, évaluer la possibilité de généraliser les expérimentations conduites dans le cadre d'un pôle pilote comme Ampiric supposerait d'avoir la capacité de garantir que les populations d'élèves et d'enseignants soient représentatifs de la population cible régionale ou nationale visée.

La conscience, pour les parties-prenantes, de participer à une expérimentation peut également porter en soi un biais et nécessite de s'interroger sur l'effet de la disparition de cet effet potentiel en phase de généralisation. Ainsi, le caractère volontaire généralement en œuvre pour la participation à une expérimentation dans le cadre des projets Ampiric suppose de s'interroger sur l'impact de la disparition de ce caractère en phase de généralisation. Il n'est ainsi pas dit qu'une expérimentation produisant un résultat positif auprès d'enseignants volontaires produise les mêmes effets dans le cadre d'une politique académique. Ce que rappelle également judicieusement Gordon Stobart à propos des

conditions de généralisation d'une expérimentation d'évaluation par les apprentissages localisée à toutes les écoles britanniques lorsqu'il écrit « La manière dont les politiques centralisées sont parfois mises en œuvre empêche que la généralisation de l'expérimentation atteigne ceux pour lesquels le dispositif est pourtant une clé essentielle de leur réussite, à savoir les enseignants de terrain » (2011, p.47).

Comme le rappellent Bérard & Valdenaire (2013, p.5), « le changement d'échelle peut induire des effets dits d'équilibre général qui n'affectent pas un dispositif expérimental ». Dit autrement, rien ne garantit que la généralisation d'un dispositif expérimental, par la montée en généralité qu'elle produit, se traduise par la réplication à l'identique des effets constatés sur une population réduite ; ce qui rejoint la question de la représentativité.

Si l'évaluation d'une expérimentation permet de contrôler des effets de court terme – ceux mesurables dans le temps de la phase expérimentation/évaluation – on peut s'interroger sur l'émergence d'effets de moyen terme car on fait comme si « la généralisation se déroulerait dans les conditions de l'expérimentation sans évolution du contexte, ni des normes, et représentations sociales » (Beaupère, Bosse et Lemistre, 2014). Ces évolutions échappent de fait à une observation temporellement située. Ce questionnement est particulièrement prégnant au regard de la préconisation de pérenniser ou pas le dispositif au moment de son évaluation.

Pour finir, on peut rajouter que l'évaluation d'un dispositif tel qu'Ampiric et en son sein le CréativLab, parmi d'autres actions et projets, engendre un bouquet de difficultés supplémentaires liée sa complexité. En effet, le caractère emboîté et multiniveau des projets et des expérimentations dans l'expérimentation suppose d'adopter une démarche évaluative elle aussi multiniveau, touchant à des domaines éducatifs divers, des populations scolarisées hétérogènes (élèves, étudiant-es, enseignant-es) tant en formation initiale que continue, des territoires différenciés et des problématiques nombreuses. En évaluer la possibilité de pérennisation et plus encore de généralisation représente un challenge évaluatif de premier ordre.

Si la question de la pérennisation / généralisation des dispositifs évalués se pose pour les raisons évoquées plus tôt, celle des limites des prescriptions lors de la confrontation au terrain se pose également.

Le passage à l'échelle : les limites des prescriptions

L'innovation ne se décrète pas. Pas plus en pédagogie qu'ailleurs. Pour advenir elle suppose de faire différemment dans un but d'amélioration des méthodes d'enseignement, des ressources pédagogiques, des apprentissages ... La question de la généralisation de l'innovation dans la classe, du changement là où les résultats de la recherche les plus probants doivent arriver, concerne in fine celle ou celui qui en a la responsabilité : l'enseignante ou l'enseignant. Or, il peut s'avérer qu'arrivée en bout de chaîne, la verticalité de prescriptions trouve parfois des obstacles dans la confrontation au réel. En France, les programmes d'enseignement et leurs documents d'accompagnement ne manquent pas de détails. Pour autant, certains aspects du programme peuvent être plus ou moins traités, tout

comme certaines pratiques, qu'ils ne préconisent pas particulièrement, peuvent largement se répandre. Cette distinction entre prescrit et réel s'inscrit dans l'analyse de l'activité enseignante qui intéresse particulièrement la formation professionnelle des enseignants. À un niveau de master universitaire, celle-ci s'inscrit en faux avec la diffusion de méthodes pédagogiques, voire de recettes, peu étayées sur le plan scientifique qui renverraient l'exercice de la profession d'enseignant à un métier d'exécution loin d'être à la hauteur des enjeux de société en matière d'éducation. L'adossement de cette formation aux savoirs produits par la recherche est donc un enjeu majeur du travail de l'enseignant pour lequel les processus d'enseignement-apprentissage occupent une place centrale. L'ancrage relatif aux savoirs à transmettre et à leur didactique disciplinaire conditionne l'organisation de l'enseignement et induit des processus de transmission-appropriation de savoirs. La discipline joue un rôle majeur qui situe l'enseignement dans un champ épistémologique et dans un format scolaire précis (programmes, niveaux, enseignants...).

Toutefois, bien que lié au travail et à la professionnalité enseignante, cet ancrage disciplinaire ne suffit pas à totalement caractériser le travail de l'enseignant et les processus d'enseignement-apprentissage pour lesquels l'activité de l'enseignant, à l'origine de celle de l'élève, est déterminante. Les ressources, les matériels et les dispositifs pédagogiques sont des médiateurs instruments (Vérillon & Rabardel, 1995 ; Engeström & Sannino, 2010) de l'activité du professeur, activité dont l'efficacité se mesure au fait que les élèves ont appris. En d'autres termes, c'est ce que l'enseignant demande de réaliser qui est porteur de l'activité de l'élève. L'élaboration et le choix des tâches est donc capital (Brandt-Pomares, 2008). Aussi, réunir les conditions pour que les enseignants développent leur activité pour favoriser l'apprentissages des élèves, interroge la question de l'articulation entre le monde de la recherche en éducation et celui des pratiques enseignantes. Elle bute aujourd'hui sur la difficulté à répartir le travail d'ingénierie, de co-construction entre chercheurs et enseignants. Le CréativLab apporte une solution en amont de ce qui peut devenir une innovation pédagogique efficace par un travail d'articulation entre le monde de la R&D et celui de son application.

Sans exclure que, pour leur généralisation, la mise en œuvre des ingénieries développées dans le cadre du CréativLab fasse l'objet d'analyse en contexte écologique, le processus proposé par le CréativLab tient compte de la professionnalité des enseignants et de la complexité de leur tâche. Ce type d'élaboration qui intègre cette dimension professionnelle dès la conception des dispositifs favorise le processus d'appropriation des innovations par les enseignants et leur pilotage de l'enseignement par les apprentissages.

Conclusion

Développer, expérimenter et évaluer des dispositifs pédagogiques innovants est un processus long et complexe. Le CréativLab d'Ampiric en est un exemple particulièrement illustratif et des précautions d'évaluation sont prises pour donner aux projets retenus les meilleures chances de succès, en appui sur une collaboration étroite entre entreprises, chercheurs et enseignants. Mais le développement, l'expérimentation et l'évaluation des projets dans le cadre du CréativLab posent aussi la question du passage à l'échelle, qui est transversale à tous les dispositifs innovants. Ce passage à l'échelle est un enjeu, notamment dans le contexte de l'éducation, mais certaines précautions doivent être prises, tout comme

les limites doivent être connues. Lors de l'évaluation de dispositifs innovants, il s'agit d'être vigilant pour placer le curseur de l'évaluation au bon niveau, permettant de trouver un équilibre entre rigueur de la méthodologie scientifique, échelle envisagée et finalité du dispositif.

Références

Ardito, C., Buono, P., Costabile, M. F., Lanzilotti, R., Piccinno, A., & Zhu, L., « On the transferability of a meta-design model supporting end-user development », in *Universal Access in the Information Society*, 14(2), 2015, p.169-186.

Barrett, B. H., Beck, R., Binder, C., Cook, D. A., Engelmann, S., Greer, R. D., ... & Watkins, C. L., The right to effective education, in *The Behavior Analyst*, 14, 1991, p.79.

Beaupère, N., Bosse, N., & Lemistre, P., « Expérimenter pour généraliser le Portefeuille d'Expériences et de Compétences à l'université : le sens de l'évaluation, in *Formation Emploi*, 126, 2014, p. 99-117.

Berard, J., & Valdenaire, M., « L'expérimentation pour renouveler les politiques publiques ? », in lavedesidees.fr, <https://lavedesidees.fr/L-experimentation-pour-renouveler-les-politiques-publiques.html> (consulté le 2 janvier 2023), 2013.

Bonnardel, N. & Didier, J., « Enhancing creativity in the educational design context: An exploration of the effects of design project-oriented methods on students' evocation processes and creative output », in *Journal of Cognitive Education and Psychology*, 15(1), 2016, p.80-101.

Brandt-Pomares, P. (2008). Searching for information on the internet about the link between task and activity. In J. Ginestié (Ed.), *The cultural transmission of artefacts, skills and knowledge: Eleven studies in technology education*, Rotterdam: Sense Publishers, 2008, p.173-192.

Conseil d'État, « Les expérimentations : comment innover dans la conduite des politiques publiques ? », Collection « Les études du Conseil d'État », 2019.

Coupaud, M., Castéra, J., Armand, H., Brandt-Pomares, P. & Delsérieys, A., « Validité et fiabilité d'un questionnaire visant à situer les conceptions de l'évolution du vivant de collégiens », in *Recherche en Didactique des Sciences et des Technologies*, 20, 2019, p.27-59.

Davies, P., "What is evidence-based education?", in *British Journal of Educational Studies*, 47, 1999, p. 108-121.

Davies, P., *Systematic reviews and the Campbell collaboration. Evidence-based practice in education*, Cornwall: Open university Press, 2004.

Dawson, A., Yeomans, E., & Brown, E. R., "Methodological challenges in education RCTs: reflections from England's Education Endowment Foundation", in *Educational Research*, 60(3), 2018, p.292-310.

Dehaene, S., Pasquinelli, E., *La recherche translationnelle en éducation. Synthèse de la recherche et recommandations*, Document retrieved on <https://www.reseau-canope.fr/conseil-scientifique-de-leducation-nationale-site-officiel/outils-pedagogiques/syntheses-et-recommandations.html>, 2021.

Edoald, T., & Nevill, C., “Working out what works: The case of the Education Endowment Foundation in England”, in *ECNU Review of Education*, 4(1), 2021, p.46-64.

Engeström, Y., & Sannino, A., “Studies of expansive learning: Foundations, findings and future challenges”, in *Educational Research Review*, 5(1), 2010, 1-24.

Fiedler, D., Sbeglia, G., Nehm, R., & Harms, U., “How strongly does statistical reasoning influence knowledge and acceptance of evolution?”, in *Journal of Research in Science Teaching*, 2019, p.1-24.

Gough, D., *Systematic research synthesis. Evidence-based practice in education*, Corwall: Open university Press, 2004.

Gough, D., “Weight of evidence: a framework for the appraisal of the quality and relevance of evidence”, in *Research Papers in Education*, 22, 2007, p. 213-228.

Haute Autorité de Santé, *Niveau de preuve et gradation des recommandations de bonne pratique. État des lieux*, Paris: HAS, 2013.

Morris, S., Edoald, T., Lloyd, C., & Kiss, Z., “The importance of specifying and studying causal mechanisms in school-based randomised controlled trials: Lessons from two studies of cross-age peer tutoring”, in *Education Research and Evaluation*, 22(7-8), 2016, p.422–439.

Nevill, C., EEF blog: Do EEF trials meet the new gold standard? [Blog post]. <https://educationendowmentfoundation.org.uk/news/do-eef-trials-meet-the-new-gold-standard/>, 2016.

Oakley, A. (2004). *Who’s afraid of the randomized controlled trial? Social research methods: a reader*, Londres: Routledge, 2004, p. 261-268.

Saussez, F., & Lessard, C., « Entre orthodoxie et pluralisme, les enjeux de l’éducation basée sur la preuve », in *Revue Française de Pédagogie. Recherches en éducation*, 168, 2009, p.111-136.

Slavin, R., “Evidence-based educational policies: transforming educational practice and research”, in *Educational Researcher*, 31, 2002, p.15-21.

Slavin, R., “What works? Issues in synthesizing educational programs evaluations”, in *Educational Researcher*, 37, 2008, p.5-15.

Stobart, G., « L'évaluation par les apprentissages : d'une expérimentation locale à une politique nationale », in *Revue Française de Pédagogie*, 174, 2011, p.41-48.

The Education Endowment Foundation, EEF guidance on choosing and delivering attainment tests, 2012.

[https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting up an Evaluation/EEF testing criteria and guidance on blinding.pdf](https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/EEF_testing_criteria_and_guidance_on_blinding.pdf)

Vérillon, P., & Rabardel, P., "Cognition and artifacts - a contribution to the study of thought in relation to instrumented activity", in *European journal of psychology of education*, 10(1), 1995, p.77-101.