

Repenser l'évaluation dans les projets STEM : vers une pédagogie processuelle, réflexive et inclusive

Rethinking Assessment in STEM Projects: Toward a Process-Based, Reflective, and Inclusive Pedagogy

Cécile Barbachoux – cecile.barchoux@univ-cotedazur.fr – <https://orcid.org/0000-0001-7254-9914>

INSPE-Université Côte d'Azur - France

Pour citer cet article : Barbachoux, C. (2025). Repenser l'évaluation dans les projets STEM : vers une pédagogie processuelle, réflexive et inclusive. *Évaluer. Journal international de recherche en éducation et formation*, 11(2), 75-106. <https://doi.org/10.48782/e-jiref-11-2-75>

Résumé

Dans un contexte éducatif en mutation, où les projets STEM (Sciences, Technologies, Ingénierie, Mathématiques) s'imposent comme des leviers puissants de transformation des pratiques, cet article interroge les conditions d'une refondation de l'évaluation scolaire à partir d'expérimentations interdisciplinaires menées du cycle 2 au lycée. Loin d'une conception strictement normative ou sommative, nous montrons, à partir de données empiriques variées (journaux d'apprentissage, co-évaluations, portfolios, grilles d'analyse), comment l'évaluation processuelle, réflexive et collaborative permet d'accompagner les élèves dans des parcours interdisciplinaires où la compréhension émerge de la circulation entre plusieurs registres de savoirs. L'analyse s'appuie sur une série de dispositifs expérimentés du cycle 2 au lycée, dans lesquels l'évaluation devient un levier de médiation pédagogique : coconstruction de grilles critériées, journaux d'apprentissage, co-évaluations orales, portfolios numériques. Ces pratiques favorisent non seulement l'acquisition de connaissances et de démarches disciplinaires, mais aussi le développement d'apprentissages transversaux tels que la coopération, la gestion de l'effort, la créativité dans la résolution de problèmes, ou encore la capacité à réguler ses actions et à expliciter ses choix. Elles contribuent à transformer progressivement le rapport des élèves au savoir, à l'erreur et à la réussite, en reconnaissant la pluralité des modes d'expression, des démarches et des formes d'engagement mobilisées par les élèves dans les situations d'apprentissage. L'article propose une lecture progressive et contextualisée de l'évaluation STEM, montrant comment ses modalités peuvent évoluer, du symbolique et narratif à l'école primaire, au critique et métacognitif au lycée, tout en conservant une cohérence fondée sur la reconnaissance des processus. En conclusion, l'évaluation est pensée comme un espace de subjectivation, un lieu de construction du sens et un vecteur de littératie critique face aux enjeux technologiques contemporains. Elle appelle, pour se déployer pleinement, une culture collective d'innovation évaluative, soutenue institutionnellement et articulée aux recherches didactiques et pédagogiques.

Mots-clés

Évaluation processuelle, Projets STEM, Différenciation pédagogique, Métacognition, Interdisciplinarité, Pensée critique, Autonomie de l'élève, Innovation didactique

Abstract

In a shifting educational landscape where STEM projects (Science, Technology, Engineering, and Mathematics) are emerging as powerful drivers of pedagogical transformation, this article explores the conditions for a fundamental rethinking of school assessment, guided by the principles of interdisciplinarity, collaboration, experimentation, and critical thinking. Moving away from a normative and summative approach, it advocates for a process-oriented, reflective, and inclusive form of assessment capable of supporting students through open-ended, differentiated, and engaging learning trajectories. The analysis draws on a series of classroom implementations spanning from early primary to upper secondary education, where assessment is reconceived as a pedagogical mediation tool: co-constructed rubrics, illustrated learning journals, peer-assessment practices, and digital project portfolios. These devices not only support the acquisition of disciplinary knowledge but also foster transversal competencies such as cooperation, perseverance, creativity, and self-regulation. They contribute to a progressive transformation of students' relationship to knowledge, error, and success, by valuing diverse forms of intelligence and legitimizing alternative ways of expressing and evidencing progress. The article proposes a developmental and contextualized perspective on STEM assessment, showing how its modalities can evolve — from symbolic and narrative forms in early education to critical and metacognitive formats in secondary school — while maintaining internal coherence centered on process recognition. In conclusion, assessment is envisioned as a space for subjectivation, a site for meaning-making, and a vector of critical literacy in response to contemporary technological challenges. To fully realize this potential, it calls for a collective culture of evaluative innovation, supported institutionally and grounded in pedagogical and didactic research.

Keywords

Formative assessment, STEM projects, Differentiated instruction, Metacognition, Interdisciplinarity, Critical thinking, Student autonomy, Didactic innovation.

1. Introduction : L'évaluation, angle mort de l'approche STEM

L'approche STEM (Sciences, Technologie, Ingénierie, Mathématiques), présentée par de nombreuses institutions internationales (OCDE, UNESCO) et par plusieurs travaux de recherche comme un levier majeur de transformation pédagogique, reste paradoxalement marquée par un impensé : celui de l'évaluation. Alors même que cette approche mobilise des démarches de projet et des apprentissages transversaux tels que la pensée critique, la collaboration, la résolution créative de problèmes ou l'autorégulation, elle est encore trop souvent accompagnée de dispositifs évaluatifs qui reconduisent des logiques normatives, disciplinaires et essentiellement sommatives.

Or, dans un monde où l'intelligence artificielle, les systèmes complexes et les transformations technologiques appellent des capacités d'adaptation, d'interprétation et de coconstruction du sens, il devient urgent de penser une évaluation à la hauteur des finalités de l'approche STEM.

Les projets STEM, en effet, ne visent pas seulement la maîtrise de contenus disciplinaires, mais une transformation de la posture d'apprenant : l'élève y devient explorateur, concepteur, résolveur de problèmes. Cette posture active ne peut être évaluée efficacement par des contrôles classiques, centrés sur des résultats figés. Il s'agit dès lors d'imaginer des formes d'évaluation formative, réflexive et inclusive, qui valorisent le processus d'apprentissage autant que les productions finales.

Plusieurs travaux (Tytler et al., 2019 ; Becker & Park, 2011) ont montré que l'approche STEM favorise significativement la motivation scolaire, l'engagement cognitif et la différenciation pédagogique. Ces recherches soulignent également que ces bénéfices ne peuvent se déployer pleinement que dans un écosystème éducatif cohérent où l'évaluation joue un rôle structurant (Tytler, Symington, & Cripps Clark, 2019).

Dans cette perspective, l'environnement évaluatif apparaît comme une condition essentielle : il doit reconnaître la valeur des démarches d'investigation, des erreurs fécondes, des hypothèses inachevées, des tentatives divergentes ou encore des coopérations constructives. Sans cette reconnaissance, l'approche STEM risque de reproduire les logiques normatives et disciplinaires qu'elle ambitionne pourtant de dépasser.

Ces travaux mettent ainsi en évidence que l'évaluation, dans un contexte STEM, ne peut se réduire à la mesure des compétences ou à la vérification des résultats. Elle doit permettre d'accompagner un parcours, de documenter des transformations progressives et de valoriser la diversité des contributions que génèrent naturellement des situations d'apprentissage interdisciplinaires.

C'est précisément à partir de ces constats que nous proposons, dans la suite du texte, une conception de l'évaluation STEM résolument processuelle et inclusive, centrée sur les démarches plutôt que sur les seuls produits. Cette approche, présentée et illustrée par trois études de cas, vise à répondre aux exigences identifiées par la littérature : cohérence de l'environnement évaluatif, reconnaissance des processus interdisciplinaires et prise en compte des formes variées d'engagement des élèves.

Cette perspective rejoint les enjeux contemporains de l'école inclusive. Face à l'hétérogénéité des élèves — diversité des rythmes d'apprentissage, des modes d'engagement, des parcours scolaires et des expériences culturelles — l'évaluation dans les projets STEM peut devenir un outil puissant pour différencier les apprentissages et soutenir la progression de chacun.

En mobilisant des dispositifs souples (grilles critériées, auto- et coévaluations, journaux d'apprentissage) et en diversifiant les formes de productions attendues (orales, visuelles, techniques, narratives), elle permet à chaque élève de montrer la qualité de sa démarche, de s'engager selon les ressources qu'il mobilise le mieux et d'avancer dans des situations d'apprentissage étayées, adaptées à ses possibilités actuelles de développement (Vygotsky, 1978).

Ces observations convergent avec les résultats de la méta-analyse de Becker & Park (2011), qui démontrent que les approches intégratives en STEM renforcent significativement la compréhension conceptuelle et la résolution de problèmes, à condition que l'évaluation accompagne les processus de manière différenciée et rende visibles les étapes, les essais et les ajustements des élèves.

Pourtant, la mise en œuvre d'une telle évaluation reste souvent marginale. Le système éducatif français, encore largement structuré par une évaluation certificative — entendue, à la suite de Perrenoud (1998), comme un dispositif sommatif visant à établir un bilan institutionnel des acquis et à produire une décision de validation, de sanction ou de promotion — peine à faire une place aux modalités alternatives. Comme le souligne Allal (2007), ce type d'évaluation intervient en fin de parcours, s'appuie sur des critères stabilisés, et laisse peu de place à la prise en compte des démarches, des essais et des ajustements caractéristiques des projets STEM. De nombreux projets innovants se heurtent ainsi à un manque de reconnaissance institutionnelle, l'évaluation devenant l'angle mort d'une pédagogie par ailleurs profondément novatrice.

Dans ce contexte, il devient essentiel d'articuler étroitement les projets STEM et une refondation des pratiques évaluatives. Cette articulation est d'autant plus stratégique que l'évaluation constitue aujourd'hui un levier déterminant pour l'équité, la motivation et la réussite. Si l'on veut que les projets STEM ne soient pas réservés à une élite scolaire déjà en réussite, mais deviennent de véritables dispositifs de démocratisation des savoirs, alors l'évaluation doit en être le moteur pédagogique, et non le frein invisible.

Ce texte propose ainsi d'explorer les conditions concrètes d'une évaluation authentique dans les projets STEM. Dans le sens proposé par Wiggins (1990), l'évaluation authentique repose sur des tâches complexes, contextualisées et proches de pratiques réelles. Dans le cadre de l'approche STEM, elle consiste à documenter non seulement les productions finales, mais aussi les démarches interdisciplinaires, les prises d'initiative et les processus de modélisation ou de conception mobilisés par les élèves (Tytler et al., 2019 ; Becker & Park, 2011). Plus largement, ce travail distingue plusieurs registres évaluatifs : l'évaluation certificative, au sens de Perrenoud (1998) et Allal (2007), qui renvoie à une logique sommative de bilan institutionnel ; et l'évaluation différenciée (Suchaut, 2010 ; Perrenoud, 1998), qui adapte les critères et les supports aux profils d'élèves afin de valoriser la diversité des formes d'engagement.

En articulant ces dimensions, l'objectif est de montrer que penser l'évaluation autrement, c'est permettre aux projets STEM de déployer tout leur potentiel inclusif, réflexif et formateur — en particulier dans une école confrontée à la complexité du monde contemporain.

2. Les projets STEM comme levier d'une évaluation formative et inclusive

L'un des apports majeurs de l'approche STEM réside dans sa capacité à transformer non seulement les pratiques pédagogiques, mais également les postures évaluatives. Parce qu'ils sont ancrés dans des démarches actives, collaboratives et interdisciplinaires, les projets STEM constituent un terrain privilégié pour mettre en œuvre des évaluations formatives, différenciées et centrées sur les processus d'apprentissage.

Dans un système éducatif où l'évaluation demeure encore largement associée à la sélection, à la certification ou à la notation chiffrée, les projets STEM introduisent un véritable changement de paradigme : il ne s'agit plus seulement d'apprécier ce qui est su ou réussi, mais de documenter comment les élèves apprennent, progressent, coopèrent et réfléchissent. Cette évolution n'implique pas de renoncer à la certification — qui demeure une fonction institutionnelle essentielle — mais d'articuler plus étroitement cette dimension sommative à des modalités formatives capables d'accompagner les apprentissages en cours (Perrenoud, 1998 ; Allal, 2007 ; Horoks, 2021).

2.1. *Évaluer pour apprendre : une logique centrée sur les processus*

Dans l'enseignement scientifique, technologique, mathématique et d'ingénierie, l'évaluation ne peut se réduire à une vérification ponctuelle des résultats attendus. Les démarches pédagogiques fondées sur les projets STEM engagent les élèves dans des activités complexes mobilisant la modélisation, le raisonnement hypothético-déductif, l'expérimentation, la résolution de problèmes ouverts ou encore la conception de solutions techniquement viables (Thouin, 1999).

Ces tâches ne relèvent pas d'une logique de réponse unique : elles reposent sur des processus cognitifs, méthodologiques et techniques qu'il importe d'observer, d'analyser et de réguler tout au long de l'activité (Allal, 2007 ; Horoks, 2021 ; Perrenoud, 1998). L'évaluation doit alors soutenir le développement progressif des apprentissages disciplinaires en contexte (Black & Wiliam, 1998), en documentant les choix opérés, les stratégies mobilisées, les erreurs rencontrées et les ajustements successifs.

Dans cette perspective, les propositions de Sadler (1989) demeurent centrales : pour être réellement formative, l'évaluation doit être intégrée à la conception du dispositif pédagogique et permettre aux élèves de comparer, ajuster et réguler leurs actions en temps réel. L'enseignant devient ainsi un guide capable d'interroger les cheminements intellectuels, de mettre en évidence les invariants scientifiques traversant les différentes tentatives et d'encourager la réflexivité sur les démarches engagées (Baron & Bruillard, 2003).

2.2. *Des outils pour rendre visibles les démarches et les apprentissages*

La mise en œuvre de cette évaluation formative centrée sur les processus suppose l'usage d'outils adaptés :

- **grilles critériées**, fondées non sur des « compétences » elles-mêmes mais sur les critères permettant d'apprécier la qualité des apprentissages engagés (rigueur du raisonnement mathématique, pertinence d'un modèle, cohérence d'une solution technique) ;
- **cahiers de bord** retraçant l'évolution des idées et les essais successifs ;

- **moments d'explicitation orale ou écrite** des démarches ;
- **coévaluations entre pairs**, fondées sur des critères disciplinaires précis (Lameul & Loisy, 2014 ; Décamp & Ravachol, 2023).

À titre d'exemple, dans un projet où les élèves doivent concevoir une structure capable de résister à une contrainte mécanique (pont en matériaux recyclés, bras robotisé, etc.), l'évaluation du produit fini — la simple solidité de la structure — serait insuffisante. Deux réalisations conformes peuvent résulter de démarches radicalement différentes : l'une fondée sur une intuition empirique, l'autre sur un raisonnement déductif appuyé sur les lois de la mécanique.

C'est en documentant les calculs initiaux, les choix de forme, les simulations et les ajustements successifs que l'enseignant peut juger de la solidité des apprentissages, de leur transférabilité et de leur articulation aux concepts fondamentaux de la discipline (Baron & Bruillard, 2003).

De même, un résultat final non conforme peut révéler des apprentissages de grande qualité : compréhension fine des contraintes du problème, capacité à analyser ses erreurs, proposition d'alternatives techniquement fondées. Loin de diluer l'exigence scientifique, l'évaluation formative permet de rendre visibles des dimensions essentielles des disciplines — structure logique du raisonnement, pertinence des représentations, maîtrise des langages symboliques, articulation théorie-pratique — et d'ancrer les élèves dans une posture active et réflexive vis-à-vis du savoir.

2.3. Stratégies cognitives, feedback et différenciation dans l'évaluation STEM

Les recherches en psychologie cognitive et en sciences de l'éducation soulignent l'existence de deux grandes stratégies d'apprentissage mobilisées par les élèves : la pensée par analogie et la pensée par règles (Gentner, 1983 ; Tardif, 1992). La première repose sur des associations d'expériences antérieures, permettant de transférer une solution d'une situation connue à une situation nouvelle. La seconde s'appuie sur des procédures explicites et formalisées, telles que les algorithmes, les lois scientifiques ou les consignes normées.

Dans le contexte des projets STEM, l'alternance entre différents types de raisonnement – intuitif, analogique, systématique et explicite – mérite d'être reconnue dans les modalités d'évaluation. Le raisonnement intuitif renvoie aux démarches spontanées que les élèves mobilisent pour proposer une première idée ; le raisonnement analogique constitue une forme plus structurée de cette intuition, en transposant une expérience antérieure à un problème nouveau. À l'inverse, le raisonnement systématique repose sur l'application d'une procédure organisée et reproductible, tandis que le raisonnement explicite désigne la capacité à verbaliser et justifier chacune des étapes de la démarche engagée.

Reconnaître ces registres dans l'évaluation permet de valoriser non seulement les solutions produites, mais aussi la nature du cheminement cognitif engagé. Une grille critériée bien construite peut ainsi intégrer des indicateurs relatifs à la forme du raisonnement mobilisé, afin de mieux documenter la diversité des démarches présentes dans les projets STEM.

Par ailleurs, un consensus s'installe autour de la valeur formative de l'erreur, à condition qu'elle soit suivie d'un retour rapide, précis et constructif (Hattie & Timperley, 2007 ; Black & Wiliam, 1998). Loin de sanctionner, l'erreur active une boucle de rétroaction cognitive qui renforce la mémoire de travail et favorise la consolidation en mémoire à long terme. Dans

les projets d'ingénierie ou de modélisation scientifique, les ajustements progressifs, les hypothèses erronées ou les bifurcations constituent des éléments essentiels à documenter et valoriser : ils incarnent une conception non linéaire de l'apprentissage au cœur des approches processuelles.

Certaines approches pédagogiques facilitent concrètement la mise en œuvre d'une évaluation différenciée et formative. La méthode de la classe puzzle (Aronson et al., 1978), en responsabilisant chaque élève sur une partie du contenu à maîtriser et transmettre, favorise l'engagement cognitif et la coopération. Elle s'inscrit pleinement dans la logique des projets STEM, par son accent sur la réciprocité, la coconstruction et la mise en mots des savoirs. De même, les travaux de Suchaut (2010) rappellent que l'efficacité des dispositifs d'enseignement dépend de leur capacité à s'adapter aux rythmes et profils cognitifs des élèves. Intégrée à l'évaluation, cette différenciation permet de reconnaître des démarches souvent invisibles dans les formats normatifs : persévérance, autonomie, stratégie d'essai-erreur, communication, etc.

Ces éléments renforcent ainsi les principes déjà développés dans l'approche STEM : valorisation du processus, reconnaissance de la diversité des formes de raisonnement et intégration de l'évaluation dans une dynamique réflexive, inclusive et durable.

2.4. Une évaluation différenciée, adaptée à la diversité des profils

Dans les disciplines STEM, la mise en œuvre de projets à composante expérimentale, modulaire ou technologique favorise naturellement une différenciation pédagogique attentive à la diversité réelle des élèves. Cette diversité ne renvoie pas à des « styles d'apprentissage » — notion non étayée scientifiquement — mais à la variabilité des expériences scolaires, des rapports au savoir, des façons d'entrer dans la tâche et des ressources mobilisées par chaque apprenant (Baron & Bruillard, 2003). Le caractère ouvert de ces projets — souvent ancrés dans des problèmes complexes ou des objets techniques à concevoir — permet ainsi à chaque élève de mobiliser ses ressources propres, qu'elles soient logico-mathématiques, visuo-spatiales, langagières, techniques ou socio-relationnelles.

Dans ce cadre, l'évaluation ne peut être uniforme ni exclusivement fondée sur des critères normatifs classiques. Elle doit être différenciée, tant dans ses modalités que dans les indicateurs retenus, afin de refléter la variété des démarches, des formes de participation et des contributions individuelles. L'approche STEM étant intrinsèquement interdisciplinaire — mobilisant conjointement savoirs scientifiques, modélisation mathématique, conception technologique et analyse sociale — l'évaluation processuelle joue un rôle central : elle rend visibles non seulement les résultats obtenus, mais aussi les transferts, hybridations et ajustements qui émergent au croisement des disciplines. L'interdisciplinarité devient alors un objet d'évaluation à part entière, perceptible dans la capacité des élèves à articuler des concepts issus de domaines différents, à justifier leurs choix selon plusieurs critères et à naviguer entre des registres de preuve distincts.

Pour représenter cette diversité, l'enseignant peut s'appuyer sur une palette élargie d'indicateurs : créativité mobilisée dans la recherche de solutions, rigueur des raisonnements scientifiques, articulation entre modèles mathématiques et choix techniques, coopération au sein du groupe, autonomie dans la conduite de la tâche, ou encore persévérance face à l'erreur (Perrenoud, 1998 ; Allal, 2007). Cette diversification se traduit également par la variété des productions évaluables — exposé oral, rapport écrit, capsule vidéo, maquette physique, carte mentale, journal de bord réflexif — qui permet d'adapter l'évaluation aux modes d'expression

privilégiés par les élèves et de rendre plus fidèlement compte de la pluralité des ressources mobilisées dans des projets interdisciplinaires (Lameul & Loisy, 2014; Décamp & Ravachol, 2023).

Dans un projet d'ingénierie visant à optimiser une solution technique (filtrage de l'eau, robotique, etc.), certains élèves mobilisent plus aisément des ressources logico-mathématiques (formalisation du problème), d'autres des ressources spatiales et techniques (modélisation 3D, prototypage), d'autres encore des ressources langagières ou socio-argumentatives (présentation orale, justification des choix). Une évaluation unique, centrée sur un livrable écrit normé, passerait inévitablement à côté de cette diversité et invisibiliserait des démarches pourtant essentielles à l'activité scientifique et technologique.

Toutefois, la complémentarité des ressources ne produit pleinement ses effets que si un véritable transfert s'opère entre élèves : ceux qui maîtrisent un domaine doivent pouvoir aider ceux qui rencontrent des difficultés. Cela suppose un cadre socioconstructiviste explicite : le simple travail de groupe ne suffit pas. Le transfert nécessite des consignes précises, un étayage actif de l'enseignant et un pilotage fin des interactions pour que les échanges deviennent effectivement formatifs.

Lorsqu'elle est pensée de cette manière, l'évaluation différenciée rend visibles des démarches souvent marginalisées dans les formats académiques traditionnels — en particulier pour les élèves en difficulté dans les évaluations écrites classiques. Elle soutient ainsi une forme d'équité qui ne renonce pas à l'exigence disciplinaire (Black & Wiliam, 1998 ; ; Thouin, 1999). En reconnaissant la diversité des ressources mobilisées et en organisant les conditions permettant leur circulation au sein du collectif, cette approche contribue à redonner confiance à des élèves fragilisés par les attendus implicites de la scolarité traditionnelle, tout en installant une dynamique didactique où rigueur disciplinaire et reconnaissance des singularités se renforcent mutuellement (Perrenoud, 1998).

2.5. Coévaluer, s'autoévaluer : vers une posture réflexive

Les projets STEM, par leur structure collaborative et itérative, constituent un terrain privilégié pour l'émergence d'une posture réflexive fondée sur des pratiques d'auto-évaluation et de coévaluation intégrées au processus d'apprentissage. Dans ces contextes, l'évaluation cesse d'être un acte exclusivement externe ou descendant : elle devient un espace partagé, dans lequel les élèves participent à la définition des critères de qualité, analysent leurs démarches et régulent leur progression (Allal, 2007 ; Black & Wiliam, 1998).

Au-delà de la simple prise de recul, ces dispositifs permettent aux élèves d'orienter leur progression de manière équilibrée : l'autoévaluation les invite non seulement à identifier les ressources qu'ils mobilisent spontanément, mais aussi celles dans lesquelles ils souhaitent progresser. Cet engagement sur l'ensemble des dimensions du projet — et non sur les seules activités appréciées ou maîtrisées — constitue un levier essentiel pour relever le défi de la complémentarité productive au sein des groupes.

Lorsqu'elle est accompagnée par l'enseignant, cette démarche favorise l'émergence d'apprentissages métacognitifs fondamentaux : autonomie intellectuelle, capacité de régulation de l'activité, transformation du feedback en action. Les travaux de Nicol et Macfarlane-Dick (2006) rappellent d'ailleurs que la rétroaction efficace repose sur trois principes clés : la clarté des critères, l'implication active des élèves dans le jugement, et la possibilité de fixer des objectifs de progrès situés et réalistes.

Concrètement, l'usage de carnets de projet, de journaux d'apprentissage ou de bilans réflexifs permet aux élèves de verbaliser et structurer leur cheminement, d'identifier les obstacles rencontrés, de reconnaître les erreurs comme leviers de compréhension et de formuler des stratégies d'amélioration. Parallèlement, la coévaluation — qu'elle prenne la forme de feedback entre pairs, de grilles d'observation mutuelle ou de revues collectives de projets — engage les élèves dans un dialogue évaluatif structurant. Celui-ci favorise la reconnaissance de l'altérité des démarches et renforce la qualité argumentative de la justification disciplinaire (Lameul & Loisy, 2014; Décamp & Ravachol, 2023).

Cette dimension réflexive est particulièrement féconde dans les disciplines à forte composante procédurale ou technologique, comme les mathématiques appliquées, l'ingénierie ou les sciences expérimentales, où la réussite d'un projet ne dépend pas uniquement du résultat final, mais de la solidité conceptuelle, méthodologique et collaborative du processus engagé (Baron & Bruillard, 2003).

Ainsi, intégrer l'autoévaluation et la coévaluation dans l'ingénierie didactique des projets STEM, c'est former des élèves non seulement capables de réussir, mais aussi de comprendre les conditions de leur réussite, d'identifier leurs obstacles et d'en dégager des régularités transférables. Cette démarche s'inscrit dans les travaux fondateurs sur la métacognition et l'autorégulation (Flavell, 1976 ; Zimmerman, 2002), qui soulignent l'importance, pour les élèves, de pouvoir analyser leurs démarches, ajuster leurs stratégies et développer une compréhension active de leur façon d'apprendre (OCDE, 2019).

Une telle orientation déplace l'évaluation d'une position terminale vers un rôle central dans l'activité cognitive : elle devient un outil de structuration de l'apprentissage, de responsabilisation des élèves et de reconnaissance de la diversité des manières d'entrer dans les savoirs et de résoudre les problèmes.

2.6. L'évaluation comme moteur d'engagement

Lorsqu'elle est inscrite dans une perspective formative, l'évaluation ne joue plus le rôle d'un simple dispositif de mesure des acquis : elle devient un véritable moteur d'engagement cognitif et affectif. Dans les disciplines scientifiques et technologiques, et plus encore dans le cadre de projets STEM, cette fonction dynamisante s'observe dans la manière dont l'évaluation soutient la persévérance des élèves, stimule leur créativité et leur donne les moyens de s'approprier leur propre parcours d'apprentissage (Perrenoud, 1998 ; Black & Wiliam, 1998).

Évaluer non seulement le produit final, mais aussi les étapes du processus — hypothèses formulées, essais successifs, erreurs rencontrées, ajustements opérés — permet de donner du sens à l'effort et d'installer un climat favorable à la prise d'initiative et à l'exploration (Allal, 2007 ; Baron & Bruillard, 2003). Loin d'être stigmatisante, l'erreur devient un levier de compréhension, révélant les zones d'incompréhension et ouvrant la voie à une reconstruction conceptuelle progressive (Baron & Bruillard, 2003).

Dans cette logique, l'évaluation formative n'est pas un dispositif annexe : elle constitue une composante organique du projet pédagogique. Elle accompagne et structure l'activité intellectuelle des élèves tout en servant de support à une régulation didactique fine par l'enseignant (Lameul & Loisy, 2014, 2005). Par exemple, dans un projet de modélisation d'un phénomène physique complexe — transferts thermiques, dynamique des fluides ou propagation d'ondes — les élèves sont explicitement encouragés à explorer plusieurs pistes,

à documenter leurs tentatives et à revisiter leurs raisonnements. L'évaluation, conçue comme outil de régulation et non de sanction, légitime cette démarche exploratoire en valorisant les ajustements progressifs et l'autonomie croissante des apprenants (Décamp & Ravachol, 2023).

Cette approche s'inscrit dans une conception élargie de l'évaluation formative telle que formulée par Allal et Mottier Lopez (2005). Selon ces auteurs, la régulation externe assurée par l'enseignant, les outils ou les feedbacks prend la forme d'un étayage destiné à soutenir le développement des processus d'autorégulation des élèves. L'évaluation formative vise alors une implication accrue des apprenants dans l'autoévaluation, l'évaluation mutuelle entre pairs et l'évaluation conjointe enseignant-élève, permettant une appropriation progressive des critères de qualité.

Ainsi pensée, l'évaluation devient simultanément motrice de motivation, outil de structuration cognitive et vecteur d'autonomisation. À ce titre, elle constitue l'un des leviers les plus puissants pour faire évoluer les pratiques d'enseignement scientifique vers une logique d'apprentissage actif, réflexif et durable (OCDE, 2019). Elle contribue à installer un rapport au savoir fondé non sur la conformité immédiate, mais sur la progression, la compréhension en profondeur et la capacité à piloter sa propre démarche d'investigation.

3. Étude de cas : expérimenter l'évaluation différenciée en classe de CP

L'expérimentation, menée au sein d'une classe de CP dans le cadre d'un mémoire de Master MEEF à l'Université Côte d'Azur, constitue un exemple concret de mise en œuvre de l'approche STEM centrée sur une évaluation formative, inclusive et contextualisée. Ce projet a mobilisé des outils numériques simples (robotique éducative, codage directionnel) au sein d'une séquence interdisciplinaire intégrant l'éducation à la citoyenneté, la géométrie spatiale, la lecture et la production d'écrit.

Mais au-delà des contenus disciplinaires, l'originalité de l'expérimentation réside dans le dispositif d'évaluation mis en place, conçu comme un outil de régulation des apprentissages, de valorisation des progrès et de différenciation pédagogique.

3.1. Une évaluation pensée dès la conception du projet

Dès la phase de conception du projet, l'enseignante-stagiaire a inscrit l'évaluation au cœur de la démarche pédagogique, en intégrant de manière explicite une grille d'apprentissages transversaux conçue en collaboration avec le formateur. Cette anticipation s'inscrit dans une logique de planification à rebours (Wiggins & McTighe, 2005 ; Tomlinson & McTighe, 2006) et témoigne d'un alignement pédagogique étroit entre les objectifs visés, les situations d'apprentissage proposées et les modalités d'évaluation retenues, conformément aux principes d'une ingénierie didactique cohérente (Allal, 2007 ; Lameul & Loisy, 2014).

La grille élaborée ne se limitait pas aux apprentissages procéduraux ou conceptuels classiques : elle visait également des dimensions souvent invisibilisées dans les formats évaluatifs traditionnels, telles que l'autonomie dans la gestion des tâches, la coopération en environnement technique, la persévérance face à des problèmes ouverts, la mobilisation de ressources visuo-spatiales en situation de conception, ou encore l'aptitude à expliciter une démarche algorithmique ou de codage.

Ce dispositif évaluatif assumait une double fonction. D'une part, il permettait à l'enseignante de réguler les apprentissages en temps réel, en identifiant les leviers de progression propres à chaque élève. D'autre part, il introduisait une dimension participative et réflexive, en plaçant les élèves dans une posture active de coévaluation guidée et d'auto-évaluation à différents moments du projet. Les traces produites (grilles complétées par les pairs, journaux d'apprentissage, retours oraux) montrent que ces outils ont permis aux élèves de mieux comprendre les critères de réussite, d'identifier leurs obstacles et de structurer leurs démarches dans le temps.

L'usage partagé de la grille a également favorisé l'émergence d'un langage commun autour des apprentissages mobilisés, renforçant la lisibilité des attentes et permettant de valoriser des profils d'élèves parfois peu reconnus dans les formats normatifs classiques (Perrenoud, 1998 ; Black & Wiliam, 1998).

Ainsi pensée, l'évaluation n'était pas un ajout postérieur à l'activité, mais un élément constitutif du projet. Elle contribuait à la fois à l'engagement des élèves, à la régulation du travail en cours et à l'ancrage des apprentissages dans une logique de construction progressive — conceptuelle, méthodologique et sociale. En valorisant autant les processus que les productions finales, elle participait pleinement à la transformation du rapport au savoir, en reconnaissant la diversité des démarches déployées pour y parvenir.

3.2. Des outils variés et adaptés à l'âge des élèves

Dans une perspective attentive au développement des apprentissages dès le plus jeune âge, l'évaluation mise en œuvre dans le cadre des projets STEM a été conçue pour articuler rigueur pédagogique et accessibilité cognitive. Afin de prendre en compte la diversité des parcours scolaires, des modes d'engagement et des formes d'expression des élèves de CP, plusieurs outils évaluatifs, variés et adaptés à leur âge, ont été mobilisés dans une logique formative et inclusive. Cette pluralité de supports permettait de valoriser les ressources disciplinaires, méthodologiques et transversales mobilisées par les élèves, tout en respectant leurs rythmes de développement et leurs modes privilégiés de communication (Perrenoud, 1998).

3.3. Journaux d'apprentissage : un support central de verbalisation et de réflexivité

Les journaux d'apprentissage illustrés ont constitué l'un des outils majeurs du dispositif. Les élèves y représentaient graphiquement ou par écrit les trajets programmés du robot, commentaient les erreurs rencontrées et proposaient des pistes d'amélioration.

Les extraits présentés en Annexe 1-9.2, extraits 1 et 3, par exemple :

- Élève A : « J'ai changé la flèche et ça a marché »,
- Élève C : « J'ai trouvé l'erreur »,

montrent que ce support permettait aux élèves d'explicitier leurs ajustements, de verbaliser leur raisonnement algorithmique et de structurer progressivement une pensée réflexive (Baron & Bruillard, 2003).

Ces traces illustrent également la manière dont la pensée algorithmique peut être ancrée dans des formes expressives accessibles aux élèves de CP, tout en contribuant à l'émergence d'un langage commun sur les démarches engagées.

3.4. Grilles d'autoévaluation et de coévaluation

Des grilles d'autoévaluation et de coévaluation adaptées à l'âge des élèves complétaient ce dispositif. Leur usage a permis de diversifier les formes de restitution : certains élèves privilégiaient le repérage visuel (« J'ai coché ici parce qu'elle a changé la flèche »), tandis que d'autres formulaient des justifications plus analytiques.

L'usage partagé de ces grilles renforçait la lisibilité des critères de réussite et soutenait la construction d'une posture réflexive. À travers ces outils, les élèves apprenaient à identifier leurs forces, à situer leurs difficultés et à verbaliser les stratégies mobilisées — dimensions essentielles de l'autorégulation (Allal, 2007 ; Black & Wiliam, 1998).

3.5. Le « permis de coder » : reconnaître des ressources souvent invisibles

À l'issue du projet, un « permis de coder » symbolique était délivré, non comme une certification technique, mais comme un outil de valorisation globale de l'activité. Les items proposés — « Je persévère quand le robot ne fait pas ce que je veux », « Je coopère avec mon partenaire », « Je corrige mon programme après une remarque » — rendaient explicites des ressources socio-comportementales souvent invisibilisées dans les évaluations traditionnelles.

Cet outil contribuait à renforcer l'estime de soi des élèves en rendant tangible leur progression et en légitimant la diversité des manières de réussir.

3.6. Coévaluations orales et analyses collectives

La coévaluation orale en binômes permettait aux élèves de valider mutuellement leurs essais programmés selon des critères établis collectivement. Les traces en Annexe 1, par exemple : « *Tu as oublié la flèche qui fait tourner — Regarde, c'est écrit : "Essayer plusieurs solutions"* » montrent comment les élèves mobilisaient activement les critères pour argumenter leurs retours, favorisant un véritable dialogue évaluatif structurant (Lameul & Loisy, 2014).

En complément, les vidéos de séance et les observations qualitatives de l'enseignante servaient d'appuis pour revenir collectivement sur les moments-clés du projet : identification d'erreurs, analyse de réussites, ajustements opérés. Ces retours permettaient de renforcer la capacité des élèves à « penser leur propre pensée », rejoignant ainsi les finalités métacognitives décrites dans la littérature (Décamp & Ravachol, 2023 ; OCDE, 2019).

3.6.1 Une médiation didactique ajustée aux capacités des élèves

En articulant journaux d'apprentissage, grilles, coévaluations et dispositifs symboliques comme le permis de coder, l'évaluation devenait un véritable outil de médiation didactique, à la fois ludique, réflexif et structurant.

Les traces présentées en Annexe 1 — commentaires d'erreurs, schématisations, co-évaluations argumentées — montrent que les élèves étaient progressivement capables de :

- mettre à distance leur action,
- justifier leurs choix,
- expliquer une erreur,
- proposer des ajustements,
- reconnaître la contribution d'un pair.

Ainsi, l'évaluation participait pleinement à la construction d'une culture d'apprentissage dans laquelle la réussite se mesurait autant par la démarche engagée que par le résultat obtenu.

3.7. Des effets mesurables sur l'engagement et la confiance

L'analyse des données empiriques recueillies tout au long du projet s'appuie sur un corpus diversifié : grilles d'observation renseignées en continu, journaux d'apprentissage illustrés, enregistrements de moments collectifs, auto- et coévaluations, ainsi que des entretiens courts d'explicitation menés auprès des élèves. Ces matériaux complémentaires ont permis de documenter finement l'évolution de l'engagement, de la confiance en soi et du rapport à l'activité.

3.8. Identifier les élèves en retrait : des indicateurs objectifs

Les élèves habituellement en retrait ont été identifiés à partir d'indicateurs comportementaux reconnus dans la littérature sur le sentiment d'efficacité personnelle (Bandura, 1997 ; Bouffard-Bouchard, 1990) : faible initiative dans les activités, tendance à renoncer rapidement, évitement des prises de parole, propos auto-dévaluatifs (« je n'y arriverai pas »). Ces indices, croisés avec les observations de l'enseignante et les traces écrites des élèves, ont permis de repérer un petit groupe présentant une faible assurance académique.

Les grilles de suivi montrent que certains élèves initialement caractérisés comme peu participatifs (Annexe 1 : Élève B, case « Engagement : faible ») manifestaient, en fin de projet, une implication plus soutenue (« Coopération : moyen à élevé », « Persévérance : moyen à élevé »). De même, les journaux d'apprentissage laissent apparaître des évolutions dans la capacité à verbaliser les démarches : « J'ai aidé mon partenaire », « J'ai trouvé pourquoi ça ne marchait pas ».

3.9. L'impact des critères pluriels : des espaces différenciés de réussite

L'introduction de critères d'évaluation pluriels — persévérance, coopération, autonomie, capacité à guider un pair — a offert à ces élèves des espaces différenciés de réussite, particulièrement valorisants pour ceux en difficulté dans les formats traditionnels (Perrenoud, 1998). Les coévaluations confirment cette évolution : les premières séances montrent des retours laconiques ou non argumentés, tandis que les dernières révèlent une augmentation notable des justifications explicites fondées sur les critères construits collectivement (voir Annexe 1).

Cette diversification des indicateurs a ainsi permis de rendre visibles des formes d'engagement souvent invisibles dans les évaluations classiques — aide apportée à un pair, constance de l'effort, prise d'initiative — renforçant la légitimité des contributions des élèves habituellement moins reconnus (Black & Wiliam, 1998 ; Décamp & Ravachol, 2023).

3.10. Des transformations attestées par les entretiens (T1/T2)

Douze élèves ont été interrogés à deux moments du projet : au début (T1) et à la fin (T2). L'analyse thématique des entretiens porte sur trois dimensions : l'engagement, la perception de compétence et le rapport à la tâche.

- À T1, plusieurs élèves exprimaient une posture incertaine : « C'est difficile », « Je ne sais pas si j'y arriverai ».
- À T2, les formulations témoignent d'un basculement net : « J'ai appris en m'amusant », « J'ai compris tout seul », « J'ai aidé mon copain ».

Ces verbalisations convergent vers une transformation du rapport au savoir : les élèves se perçoivent comme des acteurs capables d'essayer, d'expliquer, de coopérer et de transformer un retour en action (Allal, 2007 ; Baron & Bruillard, 2003).

3.11. Une reconnaissance explicite des contributions non académiques

La valorisation de ressources non exclusivement académiques — entraide, écoute, initiative, ingéniosité — a joué un rôle déterminant dans l'évolution du statut de certains élèves. Ainsi, des élèves initialement en retrait (Annexe 1, Élève B : « Participation : faible ») sont progressivement devenus des ressources identifiées par leurs pairs : « Il a trouvé la bonne trajectoire », « C'est lui qui a expliqué la correction ».

Les journaux d'apprentissage confirment cette évolution : aux formulations incertaines du début succèdent des traces valorisantes (« J'ai aidé mon copain », « J'ai eu l'idée de mettre deux flèches »). Ce déplacement du regard — du produit vers le processus et les attitudes — a permis de redynamiser des élèves souvent marginalisés par les attendus normatifs.

3.12. Un effet structurel sur l'engagement et le rapport à l'apprentissage

Les effets de l'évaluation formative et différenciée intégrée à un projet STEM dépassent la seule amélioration des performances : ils concernent la qualité de l'investissement personnel, la construction identitaire de l'élève-apprenant et sa capacité à se projeter positivement dans de futures situations d'apprentissage.

En soutenant la réflexivité, en diversifiant les critères de réussite et en légitimant la pluralité des apports, l'évaluation contribue à installer un rapport actif, confiant et durable au savoir — rejoignant les conclusions des travaux sur l'apprentissage authentique et la motivation scolaire (OCDE, 2019 ; Lameul & Loisy, 2014).

3.13. Une transformation de la posture enseignante

L'implantation du projet STEM a induit une transformation notable de la posture professionnelle de l'enseignante, marquant un déplacement significatif du rôle traditionnel de transmission vers une fonction de guidage, de soutien et d'ajustement des apprentissages. Dans ce cadre, l'évaluation n'était plus envisagée comme un moment ponctuel de vérification en fin de tâche, mais comme un outil continu de médiation pédagogique, favorisant le dialogue, la coconstruction et un soutien différencié (Allal, 2007 ; ; Lameul & Loisy, 2014).

Cette évolution s'inscrit dans une conception élargie de l'acte d'enseigner, où le rôle du professeur ne se limite pas à délivrer un contenu mais consiste à orchestrer des situations d'apprentissage complexes, à soutenir l'activité cognitive des élèves et à en réguler les

dynamiques. Elle rejoint la notion de *pedagogical content knowledge* développée par Shulman (1986), selon laquelle l'expertise enseignante émerge de l'articulation contextualisée entre maîtrise disciplinaire, didactique et connaissance fine des élèves.

3.14. Une évolution observable dans les interactions de classe

Les observations recueillies au fil des six séances du projet (environ 45 minutes chacune) montrent une évolution marquée du type d'interventions de l'enseignante. Au début, les gestes professionnels étaient majoritairement explicatifs ou directifs (« Mets ta flèche ici »). Progressivement, ils laissent place à des questions ouvertes invitant à l'explicitation et à la prise d'initiative :

- « Qu'est-ce que ton robot fait maintenant ? »
- « Comment pourrais-tu vérifier ton idée ? »
- « Quelle autre solution pourrais-tu tester ? »

Les vidéos de classe (Annexe 1) montrent également que l'enseignante mobilisait de plus en plus la grille de critères comme support d'échanges :

- « Selon le critère *persévérance*, qu'est-ce que tu pourrais essayer maintenant ? »
- « Selon *essayer plusieurs solutions*, qu'est-ce que tu pourrais tester ? »

Cette évolution traduit un déplacement d'une posture centrée sur l'enseignement vers une posture centrée sur l'apprentissage et la réflexivité des élèves.

3.15. Une gestion fine de l'hétérogénéité

L'observation qualitative continue a permis d'ajuster de manière ciblée les situations d'apprentissage. Au total, une dizaine d'ajustements ponctuels ont été réalisés au cours du projet. Parmi ceux-ci :

- **simplification temporaire du parcours de programmation** pour deux élèves en difficulté lors de la troisième séance ;
- **ajout d'un support visuel supplémentaire** (grille quadrillée agrandie) après avoir constaté que quatre élèves peinaient à anticiper mentalement les déplacements du robot ;
- **constitution de binômes complémentaires** pour favoriser le guidage mutuel ;
- **reformulation individuelle des consignes** pour certains élèves ayant des difficultés de compréhension verbale.

Ces ajustements avaient pour fonction d'encourager, de relancer ou de clarifier selon les besoins : ils témoignent d'une posture réflexive fondée sur l'écoute active, l'attention fine aux interactions et la valorisation des initiatives des élèves (Perrenoud, 1998).

3.16. L'évaluation comme levier de dialogue et d'autorégulation

La transformation de posture se manifeste également dans la manière dont l'enseignante utilise l'évaluation comme levier pédagogique. Les transcriptions d'entretiens et d'échanges de coévaluation montrent qu'elle engage les élèves dans une verbalisation systématique de leurs choix :

- « Pourquoi as-tu choisi cette trajectoire ? »
- « Qu'est-ce que tu pourrais vérifier avant de tester ton idée ? »

Ces pratiques, quasi absentes en début de séquence, deviennent récurrentes en fin de projet. Elles contribuent à développer l'autonomie réflexive et la capacité d'autorégulation des élèves (Décamp & Ravachol, 2023 ; Baron & Bruillard, 2003).

3.17. Une posture enseignante repensée

L'ensemble de ces traces — vidéos, grilles, transcriptions, observations qualitatives — montre que l'enseignante devenait progressivement facilitatrice de parcours singuliers d'apprentissage, soutenant des trajectoires cognitives différenciées et instaurant un espace de reconnaissance des progrès individuels, qu'ils soient conceptuels, méthodologiques ou relationnels (Black & Wiliam, 1998 ; Baron & Bruillard, 2003).

Ainsi, la transformation de la posture enseignante induite par le projet ne relève pas seulement d'une adaptation contextuelle, mais bien d'un déplacement paradigmatique, cohérent avec les exigences contemporaines de l'enseignement des disciplines scientifiques : favoriser l'apprentissage actif, gérer la complexité, reconnaître la diversité, et accompagner l'émergence d'une subjectivité apprenante.

4. Transposer l'évaluation STEM au collège et au lycée : une progressivité ajustée

L'expérimentation menée en cycle 2, centrée sur une évaluation formative, différenciée et processuelle, a mis en évidence des effets tangibles sur le rapport des élèves au savoir, en particulier chez ceux ayant du mal à s'adapter aux normes scolaires traditionnelles. Ces effets reposent sur plusieurs traces empiriques recueillies au fil du projet : évolution positive des notations dans les grilles de suivi (Annexe 1), formulations plus assurées dans les journaux d'apprentissage (passage de « Je n'y arrive pas » à « J'ai trouvé une idée »), ou encore prise d'initiative accrue relevée dans les transcriptions de coévaluations (« C'est lui qui a corrigé le robot », « Elle a expliqué pourquoi ça ne marchait pas »). Ces données, analysées selon une démarche qualitative incluant un codage thématique et une comparaison longitudinale des premières et dernières séances, montrent un déplacement réel de la posture d'apprenant : davantage d'autonomie, une verbalisation plus fréquente des stratégies et une implication renforcée dans la coopération.

Bien que produits en contexte d'enseignement primaire, ces résultats ne doivent pas être interprétés comme spécifiques à l'âge des élèves. Ils révèlent au contraire des principes pédagogiques et didactiques transversaux — valorisation des démarches, reconnaissance des compétences multiples, coconstruction de critères, développement de la réflexivité — dont la transférabilité vers le collège et le lycée apparaît manifeste (Perrenoud, 1998 ; Vinatier, 2013).

4.1. Transposer au collège : structurer la réflexivité et l'autorégulation

La transposition au collège suppose toutefois d'ajuster ces logiques aux caractéristiques cognitives, sociales et disciplinaires propres à cet âge.

On peut y renforcer :

- la consolidation des capacités d'autorégulation ;
- l'appropriation d'un vocabulaire métacognitif plus structuré ;
- l'initiation à une évaluation collaborative encadrée dans le cadre de projets interdisciplinaires.

Les grilles de critères peuvent intégrer des attendus disciplinaires plus explicites : rigueur du raisonnement en mathématiques, qualité de la modélisation en sciences expérimentales, pertinence des choix techniques en technologie ou en ingénierie (Baron & Bruillard, 2003 ; Décamp & Ravachol, 2023).

4.2. Transposer au lycée : formalisation, argumentation et appropriation critique

Au lycée, cette dynamique peut être approfondie à travers :

- une formalisation plus précise des critères d'évaluation ;
- une implication accrue des élèves dans l'auto- et la coévaluation ;
- un travail explicite sur la justification méthodologique et l'argumentation disciplinaire.

L'évaluation devient alors un outil d'appropriation critique des savoirs : elle aide les élèves à analyser la pertinence de leurs démarches, à expliciter leurs stratégies et à situer leurs progrès. Cette approche participe au développement de compétences transversales essentielles — autonomie, collaboration, pensée complexe, capacité à apprendre à apprendre — mises en avant dans les cadres internationaux (OCDE, 2019 ; Black & Wiliam, 1998).

Elle prépare également aux formes d'évaluation plus ouvertes rencontrées ultérieurement : revue de projet, soutenance orale, justification d'une solution technique, revue de code ou analyse réflexive, typiques de certaines formations comme les BUT, les écoles d'ingénieurs ou les filières professionnelles.

4.3. Une progressivité à construire, non un modèle à dupliquer

En définitive, penser l'évaluation STEM dans une logique de progressivité ajustée ne consiste pas à dupliquer les outils du primaire vers le secondaire, mais à reformuler leurs principes structurants en fonction des contraintes cognitives, affectives et didactiques propres à chaque niveau.

Il s'agit d'une démarche d'ingénierie pédagogique exigeante, mais porteuse d'une véritable transformation des pratiques, au service d'une évaluation qui accompagne durablement les apprentissages, soutient l'engagement scolaire et contribue à installer un rapport au savoir fondé sur la compréhension, la participation et la réflexivité — quelle que soit la tranche d'âge des élèves.

4.4. Une logique de mission et de rôles différenciés au collège

Au collège, l'intégration des projets STEM se déploie dans des configurations pédagogiques plus structurées, souvent au sein de démarches interdisciplinaires formalisées portées par les équipes. Dans l'exemple du projet *Ville intelligente et mobilité durable*, les élèves sont invités à concevoir une maquette fonctionnelle de ville connectée intégrant des éléments programmables (feux de signalisation, capteurs), des dispositifs de régulation de flux et une réflexion critique sur les enjeux urbains et sociaux. Ce cadre favorise la mise en place d'une organisation du travail fondée sur des missions différenciées : certains élèves prennent en charge l'algorithmie et la programmation, d'autres travaillent sur les aspects énergétiques, tandis qu'un troisième groupe se concentre sur les enjeux d'aménagement et de circulation. Cette structuration par rôles, pensée comme moteur d'engagement et levier d'organisation, permet d'instaurer une évaluation située, appuyée sur des critères explicités en amont — précision des scripts, robustesse des modèles, cohérence des choix techniques — et nourrie par des moments d'analyse collective des prototypes (Perrenoud, 1998).

Cependant, une telle répartition comporte un risque bien identifié dans la littérature : celui de favoriser une spécialisation précoce, où les élèves auraient tendance à se cantonner aux dimensions qu'ils maîtrisent déjà — par exemple en déléguant les calculs à un camarade ou en évitant les tâches programmatiques. Pour prévenir cet écueil, l'évaluation différenciée est articulée à une exigence de progression transversale : chaque élève est invité à rendre compte, lors de l'autoévaluation, de ses progrès dans l'ensemble des dimensions du projet, y compris celles qui ne relèvent pas directement de sa mission initiale.

Dans cette perspective, l'élève chargé de la programmation est évalué sur la structuration logique de son algorithme et l'intégration des capteurs ; l'élève en charge de l'urbanisme sur la pertinence de la modélisation territoriale et la prise en compte des contraintes environnementales ; le médiateur sur sa capacité à argumenter, à formuler des compromis et à soutenir une décision commune (Baron & Bruillard, 2003 ; Décamp & Ravachol, 2023). Mais tous doivent également expliciter comment ils ont mobilisé et fait progresser des ressources transversales — raisonnement, communication, coopération, compréhension technique — indépendamment du rôle attribué.

La fiche de mission, coconstruite avec l'enseignant, ne constitue donc pas un contrat didactique différentiel au sens de Schubauer-Léoni (1988), mais un outil d'organisation, d'autorégulation et de formalisation de critères communs. Elle rend visibles les contributions individuelles dans un cadre collectif partagé et soutient l'émergence de compétences transversales essentielles : responsabilité, communication interpersonnelle, régulation de l'action collective (Lameul & Loisy, 2014 ; OCDE, 2019).

Enfin, la notation chiffrée est remplacée par une revue de projet collective, au cours de laquelle les élèves présentent, discutent et justifient leurs choix devant leurs pairs. Cette forme d'évaluation publique, fondée sur la transparence, l'argumentation et la réflexion sur la progression, constitue une alternative féconde aux dispositifs traditionnels. Elle installe une dynamique d'apprentissage active et réflexive, tout en renouvelant la perception de l'évaluation, désormais perçue non comme un verdict, mais comme un moment de consolidation, de reconnaissance et d'apprentissage partagé (Black & Wiliam, 1998 ; Allal, 2007).

4.5. Une évaluation critique et réflexive au lycée

Au lycée, les projets STEM offrent l'opportunité d'intégrer une dimension critique et éthique affirmée, en lien avec les capacités cognitives accrues des élèves et les finalités éducatives de formation à la citoyenneté, à la complexité et à la responsabilité. Les projets ne se limitent plus à la résolution de problèmes techniques : ils mobilisent des questions de société engageant le jugement, l'argumentation et la prise de position raisonnée (OCDE, 2019). C'est le cas, par exemple, du projet *Algorithmes éthiques de régulation sociale*, dans lequel les élèves sont amenés à concevoir un algorithme de tri appliqué à des situations fictives mais socialement sensibles — attribution de logements, admissions médicales ou régulation du trafic en période de crise — tout en analysant les biais implicites, les critères de justice mobilisés et les implications sociopolitiques des décisions algorithmiques qu'ils mettent en œuvre.

L'analyse des données recueillies dans ce projet s'appuie sur un corpus de traces empiriques variées (Annexe 1) : grilles d'analyse éthique complétées par les élèves, extraits de débats filmés, pseudo-codes annotés et autoévaluations portant sur la justification des critères choisis. Les verbatim montrent une évolution nette entre les premières et les dernières séances : les justifications initiales étaient essentiellement centrées sur l'efficacité (« optimiser le flux », « aller plus vite »), tandis qu'en fin de projet apparaissent explicitement des références à l'équité, à la proportionnalité ou à la non-discrimination (« éviter de pénaliser ceux qui ont déjà le moins d'avantages », « un critère égalitaire doit être la base »). Ces traces documentent la manière dont les élèves régulent leurs choix, identifient les biais, révisent leurs critères et ajustent leurs décisions à partir des échanges collectifs. Elles mettent en lumière la montée en puissance d'une argumentation éthique plus élaborée et d'une capacité croissante à articuler enjeux sociaux, principes de justice et contraintes techniques.

Dans ce contexte, l'évaluation repose sur une articulation rigoureuse entre plusieurs types de compétences :

- **techniques** : structuration du code, robustesse algorithmique, optimisation ;
- **argumentatives** : pertinence de la problématisation éthique, justification raisonnée de choix opérés sous contraintes ;
- **réflexives** : identification des biais, mise à distance critique, analyse des effets sociopolitiques d'un dispositif numérique (Perrenoud, 1998 ; Baron & Bruillard, 2003).

Cette évaluation ne se limite donc pas à la mesure de la performance technique : elle vise la profondeur de l'engagement intellectuel des élèves dans une situation où savoirs disciplinaires et enjeux de société s'entrecroisent.

Pour structurer cette évaluation complexe, chaque élève constitue un portfolio numérique de projet, intégrant différents types de traces : code commenté, journal de bord retraçant les étapes de la démarche, productions écrites (analyses, prises de position), captations orales (débats, présentations), ainsi que des éléments de réflexion métacognitive. Ce portfolio est évalué à partir d'une grille critériée coconstruite avec les élèves, explicitant les attendus techniques, argumentatifs et réflexifs, et valorisant autant le processus que le produit final (Lameul & Loisy, 2014 ; Décamp & Ravachol, 2023).

Une telle démarche permet de dépasser une vision utilitariste de l'évaluation au profit d'un dispositif intégratif, dans lequel l'élève est reconnu comme sujet pensant, capable de relier les dimensions scientifiques, sociales, politiques et éthiques d'une même problématique. Elle

contribue à l'élaboration d'une posture critique vis-à-vis des technologies — indispensable dans un monde algorithmique — tout en consolidant une culture scientifique exigeante et située. L'évaluation devient alors un levier de formation globale, à la fois intellectuelle et civique, pleinement aligné avec les compétences transversales visées dans l'enseignement secondaire contemporain (Allal, 2007 ; Black & Wiliam, 1998).

4.6. Une progression continue de l'évaluation vers l'autonomie

L'analyse transversale des projets STEM, déployés de l'école primaire au lycée, révèle la possibilité d'une progression continue et cohérente de l'évaluation, pensée non comme une succession de dispositifs disjoints, mais comme un fil conducteur structurant le parcours scolaire des élèves. Cette trajectoire ascendante se caractérise par un ajustement constant des formes d'évaluation aux capacités cognitives, sociales et disciplinaires des apprenants, tout en demeurant fidèle aux principes fondamentaux de l'approche STEM : interdisciplinarité, expérimentation, collaboration et pensée critique (OCDE, 2019).

4.7. Cycle 2 : une évaluation expressive et accessible

Au cycle 2, les données recueillies — journaux d'apprentissage illustrés, enregistrements d'échanges, traces de coévaluation (Annexe 1) — montrent que les élèves recourent principalement à des formes narratives, graphiques et symboliques pour expliciter leurs démarches.

L'analyse thématique des verbatim indique une progression nette entre T1 et T2 dans la capacité à verbaliser des stratégies simples : « j'ai changé la flèche », « j'ai corrigé l'erreur ». L'évaluation formative prend ici la forme d'une mise en mots émergente, soutenue par des supports visuels accessibles, favorisant l'entrée progressive dans la réflexivité (Allal, 2007). Elle introduit également les premiers jalons d'une structuration des apprentissages transversaux.

4.8. Collège : une structuration disciplinaire et collaborative

Au collège, l'évaluation devient plus formalisée, collaborative et ancrée dans la résolution collective de problèmes complexes. Les grilles d'analyse des rôles distribués, les vidéos de séance et les traces de coévaluation montrent que les élèves s'appuient de plus en plus sur des critères coconstruits, utilisés pour justifier leurs choix techniques (« on doit vérifier la boucle avant d'ajouter un capteur »). Les coévaluations orales analysées révèlent une augmentation notable des justifications explicites et une appropriation progressive du langage disciplinaire. L'évaluation, sans modifier les attendus communs, s'adapte aux responsabilités assumées dans le collectif. Elle devient un outil de structuration, de collaboration et de régulation, renforçant à la fois la responsabilisation individuelle et l'appartenance au groupe (Perrenoud, 1998 ; Baron & Bruillard, 2003).

4.9. Lycée : une évaluation critique, réflexive et citoyenne

Au lycée, les traces collectées — portfolios numériques, pseudo-codes annotés, revues argumentées, enregistrements de débats éthiques — indiquent une évolution vers une évaluation critique, réflexive et éthique. Les élèves articulent progressivement critères techniques (efficacité, robustesse) et critères éthiques (équité, non-discrimination, transparence), comme en témoignent les verbatim analysés : « j'ai modifié mon algorithme

car il pénalisait certains usagers ». L'évaluation devient alors un outil d'appropriation critique des savoirs, permettant aux élèves de prendre distance avec leurs productions, de repérer les biais, de justifier leurs choix et d'interroger leurs implications sociales (Décamp & Ravachol, 2023 ; Black & Wiliam, 1998). Cette maturité réflexive est caractéristique de l'entrée dans une conception citoyenne du numérique et des technologies.

4.10. Une dynamique évaluative progressive, moteur d'autonomie

La comparaison longitudinale entre les trois niveaux, appuyée sur des données diversifiées (journaux, grilles, portfolios, verbatim) et analysées selon une méthodologie qualitative rigoureuse (codage thématique, triangulation, comparaison T1–T2), met en évidence une évolution continue de l'évaluation STEM :

- Cycle 2 : supports expressifs, émergence de la verbalisation.
- Collège : structuration disciplinaire, critères formalisés, coopération outillée.
- Lycée : réflexivité critique, articulation technique-éthique, positionnement argumenté.

Cette dynamique évaluative ne constitue pas un simple accompagnement des apprentissages : elle en est le moteur. En installant un rapport actif, situé et évolutif à l'évaluation, les élèves développent progressivement leur autonomie, leur capacité d'autorégulation, et leur aptitude à construire du sens face à des situations complexes et contextualisées. Loin d'être un instrument de mesure, l'évaluation devient ainsi le vecteur d'une pédagogie inclusive et ambitieuse, capable d'accueillir la diversité des parcours, de légitimer des formes variées d'intelligence, et de soutenir une montée en compétence à la fois individuelle et collective (Lameul & Loisy, 2014). Pensée comme un dispositif évolutif au service de l'apprentissage, elle contribue à faire de l'école un laboratoire d'intelligence collective, où chaque élève, guidé par des critères clairs et une réflexivité progressive, peut construire un parcours de réussite singulier, inscrit dans une dynamique collective de production de sens.

5. Une grille d'évaluation générique pour les projets STEM

L'un des leviers majeurs pour faire évoluer les pratiques évaluatives dans les projets STEM réside dans la mise en place de grilles critériées ouvertes, modulables et partagées. Ces outils rendent visibles les apprentissages tout en respectant la diversité des profils d'élèves. Ils s'inscrivent pleinement dans une logique de conception à rebours (*backward design*), où l'identification des objectifs fondamentaux précède l'élaboration des critères d'évaluation (Wiggins & McTighe, 2005 ; Tomlinson & McTighe, 2006).

L'enjeu n'est pas de standardiser les compétences ou de renforcer une logique certificative, mais d'objectiver des critères qualitatifs, de guider les rétroactions et d'alimenter la réflexivité des élèves tout au long du processus d'apprentissage. Les observations réalisées dans les différents projets le montrent :

- **au cycle 2**, la grille de coévaluation sert de support à la justification des démarches (« essayer plusieurs solutions », « expliquer ce que l'on fait ») ;
- **au collège**, les critères deviennent des repères structurants pour l'analyse des prototypes (« selon le critère de robustesse, que faut-il vérifier ? ») ;
- **au lycée**, ils sont intégrés aux portfolios numériques et orientent la justification des choix techniques et éthiques.

Dans cette perspective, la grille ci-dessous constitue un outil générique de régulation continue, formative et contextualisée, mobilisable en amont, pendant et après le projet. Son adaptabilité permet une appropriation souple par les enseignants comme par les élèves, du primaire au lycée.

Tableau 1 : Grille générique d'évaluation formative pour l'approche STEM

Domaine évalué	Critères possibles	Modalités de recueil
Autonomie	Capacité à s'auto-organiser, à planifier, à persévérer face aux difficultés.	Observation, journal de bord, entretiens réflexifs.
Coopération	Participation active, qualité des interactions, régulation du groupe, soutien aux pairs.	Coévaluation, captations vidéo, feedbacks entre pairs.
Créativité et modélisation	Originalité et pertinence des idées, capacité à modéliser un phénomène ou un système, itération.	Analyse de prototypes, carnet d'essais/erreurs, maquettes.
Rigueur méthodologique	Structuration des démarches, respect des étapes, cohérence des choix, qualité de la démarche expérimentale ou algorithmique.	Grilles critériées, revue de projet, observations.
Réflexivité	Explicitation de la démarche, justification des choix, analyse des erreurs, propositions d'ajustements.	Entretiens d'explicitation, auto-évaluation commentée.
Apprentissages techniques	Qualité du codage, précision des calculs, robustesse des objets ou modèles produits.	Tests pratiques, démonstration, présentation technique.
Apprentissages disciplinaires (Maths / Sciences / Technologie / Informatique / SHS)	Pertinence des modèles, qualité du raisonnement scientifique ou mathématique, maîtrise des notions technologiques ou informatiques.	Analyse des productions, résolutions de problèmes, discussions scientifiques.
Articulation interdisciplinaire (spécifique STEM)	Mobilisation de plusieurs savoirs disciplinaires, cohérence des intégrations, justification des approches combinées.	Soutenance, carnet de projet, analyse de prototypes, entretiens.

5.1. Un outil évolutif, collaboratif et structurant

Cette grille a vocation à devenir un **outil évolutif et collaboratif** :

- enrichissable selon les projets menés ;
- adaptable à différents niveaux de classe ;
- coconstruisable avec les élèves pour renforcer l'appropriation.

En valorisant autant le processus que le produit, elle permet de reconnaître des apprentissages souvent invisibilisés par les évaluations classiques : coopération, engagement, capacité à expliciter une démarche, créativité, rigueur dans les essais-erreurs.

Utilisée comme support de dialogue pédagogique, elle favorise l'alignement entre objectifs, démarches et modalités d'évaluation, conformément aux principes de l'alignement constructif (Biggs & Tang, 2011). Elle peut soutenir :

- les concertations entre enseignants,
- les entretiens tripartites (enseignant-élève-famille),
- la lisibilité des apprentissages pour la communauté éducative.

5.2. Vers une culture professionnelle de l'évaluation formative

L'approche évaluative présentée repose sur une articulation étroite entre :

- l'observation des démarches,
- la valorisation des apprentissages en cours de construction,
- la reconnaissance de la diversité des formes d'engagement.

Les dispositifs mobilisés — grilles critériées, journaux d'apprentissage, coévaluations, fiches de mission, portfolios, traces visuelles — constituent autant d'outils permettant de rendre visibles les apprentissages disciplinaires et transversaux, tout en renforçant réflexivité, coopération et autonomie.

L'enjeu ne se limite donc pas à introduire de nouveaux outils, mais à consolider une culture professionnelle partagée de l'évaluation formative, inscrite dans une ingénierie didactique cohérente. Déployée du cycle 2 au lycée, elle constitue un levier durable d'émancipation intellectuelle, permettant aux élèves de :

- objectiver leurs démarches,
- situer leurs difficultés,
- transférer leurs apprentissages,

et s'engager activement dans la complexité des savoirs scientifiques et technologiques.

Ainsi pensée, l'évaluation STEM devient un moteur de transformation pédagogique, incarnant une vision humaniste, inclusive et critique du STEM et préparant les élèves aux enjeux du XXI^e siècle.

6. Apports de l'intelligence artificielle à l'approche évaluative proposée : entre personnalisation, réflexivité et vigilance critique

L'approche évaluative développée dans cet article, fondée sur des principes de réflexivité, de différenciation, de valorisation des démarches et de reconnaissance des intelligences plurielles, entre en résonance directe avec les potentialités offertes par les outils d'intelligence artificielle (IA) dans l'éducation. Si l'IA éducative est encore en phase d'exploration dans de nombreux contextes scolaires, elle présente d'ores et déjà des convergences fortes avec la conception processuelle de l'évaluation développée dans les projets STEM, tels que décrits dans les expérimentations menées du cycle 2 au lycée (Barbachoux, 2021, 2025).

L'approche évaluative développée dans cet article — fondée sur la réflexivité, la différenciation, la documentation des démarches et la reconnaissance des intelligences plurielles — trouve un prolongement naturel dans les potentialités de l'intelligence artificielle éducative. Les expérimentations menées du cycle 2 au lycée, déjà décrites dans nos travaux antérieurs (Barbachoux, 2021, 2025), montrent que les grilles critériées co-construites, les journaux d'apprentissage et les portfolios numériques facilitent l'analyse fine des cheminements intellectuels des élèves. Par exemple, les grilles d'autoévaluation du projet « Ville intelligente » au collège (Annexe 1) montrent une évolution notable dans la manière dont les élèves justifient leurs choix techniques, passant de descriptions factuelles à des argumentations structurées. De même, les portfolios du projet « Algorithmes éthiques » au lycée documentent les révisions successives de l'algorithme et la prise en compte progressive des critères d'équité.

Ce type de traces, déjà existant dans les projets STEM, constitue un terrain propice à une intégration raisonnée de l'IA. En effet, les systèmes de recommandation ou d'analyse automatisée peuvent repérer des régularités dans les démarches des élèves, identifier des profils d'apprentissage et suggérer des ressources adaptées. Là où les traces humaines permettent surtout une analyse qualitative contextualisée, les analyses algorithmiques peuvent apporter une granularité ou une récurrence complémentaire.

Ce rapprochement n'a de sens que si le rôle de l'enseignant est pleinement articulé à ces dispositifs. Dans la perspective vygotkienne, la zone proximale de développement ne renvoie pas à une simple progression individuelle : elle désigne l'espace dans lequel l'apprenant peut accomplir une tâche grâce à l'étayage expert d'un adulte ou d'un pair plus compétent. L'IA ne remplace pas cet étayage, mais peut en renforcer la précision et la réactivité, en détectant des obstacles récurrents ou en signalant des moments où une guidance humaine devient nécessaire.

Pensée ainsi, l'IA éducative rejoint les finalités d'une évaluation conçue comme accompagnement, structuration et reconnaissance des parcours singuliers — et non comme simple mesure normative. Elle permettrait, dans la continuité des expérimentations présentées, de renforcer la capacité des enseignants à situer chaque élève dans son ZPD, en soutenant un étayage ajusté, une différenciation raisonnée et une analyse plus fine des démarches en cours.

Par ailleurs, l'orientation processuelle de l'évaluation défendue dans l'article — qui privilégie les démarches, les tâtonnements, les ajustements et la réflexivité — trouve, dans certains environnements d'apprentissage intelligents, un appui technologique fécond. L'IA permet en effet de tracer les micro-étapes d'une tâche complexe, de repérer les stratégies mises en œuvre

(essais/erreurs, mobilisation d'analogies, modélisation heuristique), et de fournir des feedbacks contextualisés, renforçant ainsi les compétences métacognitives des élèves (Hattie & Timperley, 2007 ; Black & Wiliam, 1998). Cette approche s'inscrit pleinement dans la conception de l'évaluation comme outil de structuration cognitive, développée tout au long de l'article.

Un autre point de convergence essentiel réside dans la dimension critique de l'évaluation, valorisée notamment dans les projets du secondaire traitant des enjeux technologiques et sociétaux. Dans ces contextes, l'IA ne joue pas seulement un rôle d'outil évaluatif : elle devient objet d'analyse et d'évaluation en elle-même. Intégrée à des projets STEM autour de la justice algorithmique, de l'éthique des données ou de la régulation automatique, elle offre un terrain propice à l'exercice de la littérature critique — c'est-à-dire la capacité à interroger les logiques implicites des outils techniques, à questionner leurs effets sociaux, et à développer une posture d'acteur éclairé du monde numérique (OCDE, 2019).

Néanmoins, comme le rappelle l'approche prudente et réflexive adoptée dans l'article, l'usage de l'IA en évaluation ne saurait être aveugle ni neutre. Il ne s'agit pas de déléguer le jugement pédagogique à la machine, mais d'utiliser ces outils comme amplificateurs des capacités d'analyse de l'enseignant et facilitateurs de réflexivité pour l'élève (Décamp & Ravachol, 2023 ; Lameul & Loisy, 2014). Ce positionnement rejoint les exigences éthiques et épistémologiques posées dès la conclusion de l'article : refonder l'évaluation ne consiste pas à automatiser les processus, mais à reconfigurer les régimes de preuve et de reconnaissance pour faire place à la complexité, à la pluralité et à l'émergence.

L'IA, dans cette optique, ne remplace pas l'évaluation humaine : elle élargit ses possibles. Elle peut contribuer à l'ambition formulée dans cet article d'une évaluation interprétative, évolutive et située, si elle est intégrée à une ingénierie pédagogique pilotée par une intention critique et émancipatrice. À cette condition, elle pourra devenir un vecteur d'intelligence collective, au service d'une éducation scientifique et technologique véritablement transformatrice.

7. Conclusion

Réinterroger l'évaluation à l'aune de l'approche STEM ne consiste pas simplement à substituer une méthode à une autre, ni à adapter marginalement les instruments classiques : cela engage une refondation du geste pédagogique lui-même. L'évaluation devient le lieu d'une reconfiguration profonde des postures — celle de l'élève, qui cesse d'être un exécutant pour devenir un sujet épistémique ; celle de l'enseignant, qui se transforme en médiateur de parcours, en éveillé de pensée, en architecte de contextes d'apprentissage.

Les expérimentations menées de l'école au lycée montrent qu'une évaluation, repensée dans ses finalités, ses temporalités et ses formes, est à même de réconcilier exigence et inclusion, rigueur et créativité, performance et subjectivation. En mettant l'accent sur les processus, sur la réflexivité et sur la pluralité des formes d'intelligence — y compris celles souvent invisibilisées par les formats traditionnels — l'évaluation devient non plus le point final d'un apprentissage, mais son moteur, son miroir, et parfois même son déclencheur.

Les trois études de cas présentées dans le manuscrit illustrent la cohérence entre principes théoriques et mise en œuvre concrète : au cycle 2 (robotique), au collège (ville intelligente) et au lycée (algorithmes éthiques), l'interdisciplinarité n'est pas un simple cadre thématique, mais un espace où l'évaluation processuelle rend visibles les déplacements de pensée entre

registres disciplinaires. Les outils mobilisés — grilles coconstruites, rôles différenciés, portfolios réflexifs — permettent précisément de documenter ces circulations et de soutenir les démarches interdisciplinaires engagées. Cette articulation entre approche processuelle, interdisciplinarité et choix méthodologiques constitue l'un des apports centraux du travail révisé.

Toutefois, penser une telle transformation implique de dépasser les ajustements de surface. Il ne s'agit pas seulement d'introduire des outils plus souples, mais de repenser les régimes de preuve eux-mêmes, de se demander ce que nous voulons voir, reconnaître, encourager. Peut-on évaluer l'incertitude assumée d'une hypothèse ? L'audace d'une modélisation ? La capacité à poser une question féconde ? Peut-on valoriser l'émergence de l'inattendu, la bifurcation constructive, ou le doute argumenté ? Ces dimensions, essentielles dans les disciplines scientifiques, méritent d'être intégrées à part entière dans les cadres évaluatifs, y compris au sein de l'institution scolaire.

À l'heure où les défis éducatifs se complexifient sous l'effet des mutations technologiques, écologiques et sociétales, l'évaluation ne peut plus rester à la périphérie de la réflexion didactique : elle doit en devenir l'un des foyers critiques. Non pour se dissoudre dans une pédagogie de l'indulgence, mais pour ouvrir une voie nouvelle : celle d'une évaluation interprétative, située et évolutive, capable d'articuler savoirs, valeurs, contextes et devenir.

Former les élèves à vivre, penser et agir dans un monde régi par des systèmes hybrides — humains, algorithmiques, automatisés — suppose de leur offrir non seulement des outils de lecture du réel, mais aussi des outils d'auto-interprétation. L'évaluation STEM, pensée dans cette optique, devient un dispositif de subjectivation éclairée, un espace où se construit la conscience critique des savoirs, de leurs conditions de production et de leurs effets.

La perspective d'une évolution durable des pratiques évaluatives suppose une rigueur méthodologique forte et une analyse approfondie des données recueillies. Les résultats présentés dans ce travail s'appuient précisément sur cette exigence : observations systématiques, traces écrites et visuelles, auto- et coévaluations, entretiens courts et analyses comparatives entre les premières et les dernières séances. C'est à partir de ces matériaux empiriques que se dessine la possibilité d'une culture collective de l'expérimentation évaluative, fondée sur des critères explicites, des outils partagés, une analyse fine des démarches d'élèves et un travail d'interprétation mené par les équipes pédagogiques.

Une telle culture, soutenue par les institutions et irriguée par la recherche, permettrait à l'évaluation de dépasser sa fonction de mesure pour devenir un véritable levier de compréhension, de transformation et d'engagement — un espace où les élèves apprennent à se reconnaître comme sujets capables de comprendre, de questionner et d'agir avec les autres, dans le monde, et envers eux-mêmes.

8. Bibliographie

- Allal, L. (2007). Régulation des apprentissages : Orientations conceptuelles pour la recherche et la pratique en éducation. Dans L. Allal et L. Mottier Lopez (dir.), *Régulation des apprentissages en situation scolaire et en formation* (p. 7-23). Bruxelles: De Boeck.
- Aronson, E., Blaney, N., Stephan, C., Sikes, J. et Snapp, M. (1978). *The Jigsaw Classroom*. Sage.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bouffard-Bouchard, T. (1990). Influence of self-efficacy on performance in a cognitive task. *Journal of Educational Psychology*, 82(2), 291--297.
- Barbachoux, C. & al. (2021). Evaluer avec l'Intelligence artificielle, cahiers pédagogiques, 2021, <https://www.cahiers-pedagogiques.com/evaluer-avec-l-intelligence-artificielle/>
- Barbachoux, C. (2025). The role of STEM and AI in enhancing attention skills in young students and the evaluations process. *International Journal of Education in Mathematics, Science, and Technology (IJEMST)*, 13(3), 830-849. (2025). <https://doi.org/10.46328/ijemst.4879>
- Baron, G.-L. et Bruillard, É. (2003). *Les technologies dans l'enseignement des sciences et des technologies : regards didactiques*. INRP.
- Becker, K. et Park, K. (2011). Effects of integrative approaches among STEM subjects on students' learning: A meta-analysis. *Journal of STEM Education*, 12(5-6), 23-37.
- Biggs, J. et Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press.
- Black, P. et Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Chevallard, Y. (1991). *La transposition didactique : du savoir savant au savoir enseigné*. La Pensée Sauvage.
- Décamp, N. et Ravachol, D.-O. (2023). L'évaluation et l'enseignement des sciences et des technologies. *Recherches en didactique des sciences et des technologies*, 27, 15-22. <https://doi.org/10.4000/rdst.4589>
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. Dans L. B. Resnick (dir.), *The nature of intelligence* (p. 231-236). Erlbaum.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2), 155-170.
- Hattie, J. et Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112.
- Horoks, J., Kiwan-Zacka, M., Pilet, J., Roditi, E. et Haspekian, M. (2021). Régulation des apprentissages et évaluation formative : quels regards didactiques ? <https://hal.science/hal-01612783v1>
- Lameul G. et Loisy C. (2014), *La pédagogie universitaire à l'heure du numérique*. De Boeck Supérieur.
- Mottier-Lopez, L. (2015). *Évaluations formative et certificative des apprentissages. Enjeux pour l'enseignement*, De Boeck.
- Nicol, D. J. et Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218. <https://doi.org/10.1080/03075070600572090>
- OCDE. (2019). *Cadre de compétences globales de l'OCDE pour les élèves*. <https://www.oecd.org/education/global-competency-for-an-inclusive-world.htm>
- Perrenoud, Ph. (1998). *L'évaluation des élèves. De la fabrication de l'excellence à la régulation des apprentissages*. De Boeck.

- Prain, V., Tytler, R., Hubber, P. et Waldrup, B. (2019). *Constructing representations to learn in science*. Springer
- Rochex, J.-Y. et Crinon, J. (2011). *La construction des inégalités scolaires : Au cœur des pratiques et des dispositifs d'enseignement*. Presses Universitaires de Rennes.
- Schubauer-leoni, M.-L. (1988). Le contrat didactique dans une approche psycho-sociale des situations didactiques. *Interactions didactiques*, 8, 63-75 [Publication interne]. Université de Neuchâtel.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.
- Suchaut, B. (2010). Efficacité pédagogique des classes à cours double à l'école primaire : le cas du cours préparatoire. *Revue française de pédagogie*, 172, 87-96.
- Tardif, J. (1992). *Pour un enseignement stratégique*. Montréal : Logiques.
- Thouin, M. (1999). *Problèmes de sciences et de technologie pour le préscolaire et le primaire*. Éditions MultiMondes.
- Tomlinson, C. A. et McTighe, J. (2006). *Integrating differentiated instruction & understanding by design: Connecting content and kids*. ASCD.
- Tytler, R., Symington, D. et Smith, K. (2019). A framework for conceptualising integrated STEM education. Dans R. Tytler, L. Hobbs, A. Pietsch et R. Campbell (dir.), *Integrated approaches to STEM education* (p. 3-22). Springer.
- Tytler, R., Symington, D. et Cripps Clark, J. (2019). Re-imagining school science: Achieving sustainable reform through a focus on practice. *Research in Science Education*, 49(5), 1423-1446. <https://doi.org/10.1007/s11165-017-9662-2>
- Vinatier, I. (2013). Le travail de l'enseignant. Une approche par la didactique professionnelle. De Boeck
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). ASCD.
- Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research & Evaluation*, 2(2).
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64-70.

9. Annexe 1 : Traces empiriques issues de l'expérimentation en CP

9.1. Exemple de grille d'auto- et coévaluation utilisée par les élèves

Grille simplifiée d'auto- et coévaluation, utilisée par les élèves de CP à plusieurs moments du projet (fin de séance, bilan intermédiaire, bilan final).

Compétence observée	Je réussis souvent	Je réussis parfois	J'ai encore besoin d'aide
Je coopère avec mon binôme	☐	☐	☐
Je me sers correctement du matériel technique	☐	☐	☐
J'explique ma démarche avec des mots	☐	☐	☐
Je cherche une solution avant de demander de l'aide	☐	☐	☐
Je corrige mon travail après une remarque	☐	☐	☐

9.2. Extraits anonymisés de journaux d'apprentissage

Extrait 1 — Élève A (CP) : « Aujourd'hui j'ai essayé de faire bouger mon robot mais il tournait trop. J'ai changé la flèche et ça a marché. Je suis fier. »

Extrait 2 — Élève B (CP) : « J'ai aidé Léa pour le pont. On a mis plus de cubes parce que ça tombait. On a trouvé ensemble. »

Extrait 3 — Élève C (CP) : « Je ne savais plus comment faire. La maîtresse a dit de regarder dans mon cahier. J'ai trouvé l'erreur. »

9.3. Extrait de transcription d'un moment de coévaluation orale

Élève 1 : « Tu as oublié de mettre la flèche qui fait tourner. »

Élève 2 : « Ah oui... mais j'avais mis tout droit. »

Élève 1 : « Regarde la grille, ça dit : "Essaye plusieurs solutions". »

Élève 2 : « D'accord, on essaie. »

9.4. Exemple de production d'élève (représentation visuo-spatiale)

Dessin d'un pont en Kapla avec zones fragiles indiquées (« ça tombe ici ») et flèches de forces (« ça pousse »), accompagné d'une hypothèse : « il faut mettre un pilier ».

9.5. Grille de suivi formative de l'enseignante

Exemple de grille utilisée par l'enseignante pour observer les « compétences transversales » au cours du projet.

Élève	Autonomie	Coopération	Persévérance	Compréhension des consignes	Commentaires
A	Moyen	Élevé	Élevé	Moyen	S'engage mais a besoin d'étayage pour planifier.
B	Faible	Moyen	Moyen	Élevé	Très à l'aise avec le matériel technique.
C	Élevé	Élevé	Moyen	Moyen	Verbalisations pertinentes lors du codage.

9.6. Grille de compétences transversales – Version détaillée

Domaine	Compétence transversale	Comportements / critères observables	Outils de recueil
1. Autonomie – Gestion du travail	Planifier, organiser et mener une tâche de manière autonome	Anticipe les étapes du projet et planifie l'ordre des actions Gère le temps et le matériel de manière responsable Persévère face aux obstacles et sollicite de l'aide à bon escient	Journal de bord · Observation · Entretien d'explicitation
2. Coopération – Communication	S'engager dans un travail collaboratif structuré	Participe aux échanges et écoute activement les pairs Reformule, clarifie, argumente Apporte une aide pertinente et régule les interactions	Coévaluations · Grilles de rôle · Captations vidéo
3. Créativité – Résolution de problèmes	Explorer, proposer et améliorer des solutions	Propose plusieurs pistes de résolution Expérimente, teste, ajuste Justifie les choix techniques ou conceptuels	Carnet de projet · Maquettes · Schémas
4. Rigueur méthodologique	Adopter et expliciter une démarche structurée	Suit un protocole ou construit une méthode Documente les étapes, Vérifie, contrôle Utilise des outils (maths, code, mesures) avec précision	Fiches de protocole · Traces expérimentales · Codages
5. Réflexivité – Autorégulation	Analyser ses démarches, s'auto-évaluer	Identifie erreurs, réussites, obstacles Explique ses stratégies et propose des améliorations Relit et ajuste sa production	Auto-évaluations · Bilans réflexifs · Entretiens
6. Compétences STEM spécifiques	Mobiliser des savoirs scientifiques, techniques, mathématiques ou technologiques	Modélise un phénomène, code une solution, conçoit un système Utilise des représentations symboliques (schémas, équations, algorithmes) Articule théorie et pratique	Livrables · Codes · Prototypes · Maquettes

9.7. Grille de compétences transversales – Version synthétique

Domaine transversal	Compétence visée	Indicateurs observables (critères)	Traces possibles
Autonomie et organisation	Planifier et gérer son activité	Anticipe les étapes du projet Gère le matériel / temps Persévère face aux obstacles	Journal de bord, observation, entretiens
Coopération et communication	Travailler efficacement en groupe	Écoute et reformule Apporte une aide pertinente Argumente ses choix	Coévaluation, captations vidéo, productions orales
Créativité et résolution de problèmes	Proposer, tester, améliorer des solutions	Imagine plusieurs pistes Justifie ses choix Ajuste après essais/erreurs	Carnet de projet, prototypes, traces numériques
Rigueur méthodologique	Mobiliser des démarches scientifiques/techniques	Suit un protocole ou une méthode Utilise des outils de façon fiable Vérifie ses résultats	Fiches de protocole, essais, simulations
Réflexivité et autorégulation	Analyser et améliorer sa démarche	Identifie ses erreurs Explique ses stratégies Propose des améliorations	Auto-évaluations, bilans réflexifs, entretiens
Compétences techniques disciplinaires	Mettre en œuvre des savoirs STEM	Modèle, code, calcule, conçoit, teste Utilise langages symboliques Articule théorie/pratique	Livrables, schémas, codes, maquettes