

**L'évaluation
au service des
apprentissages en
sciences**

L'évaluation au service des apprentissages en sciences

Sommaire

Qu'est-ce qu'évaluer ?	1	Q6. Quels outils de mesure peuvent être mobilisés pour l'évaluation ?	12
Q1. Quelles indications provenant de la recherche pour bien évaluer ?	2	Q7. Quand conduire des évaluations ?	13
Q2. Quels bénéfices pour l'enseignant ?	3	Q8. Comment fournir un retour à la suite de l'évaluation ?	13
Q3. Quels bénéfices pour les élèves ?	3	Q9. Quelles précautions prendre dans la création, l'administration, l'interprétation d'évaluations ?	14
Q4. Que peut-on évaluer en sciences ?	7	Q10. Où trouver des lectures pour justifier les propos de cette page ?	18
Q5. Quelles modalités adopter pour l'évaluation en sciences ?	8		

Qu'est-ce qu'évaluer ?

Évaluer consiste à recueillir des informations pertinentes par rapport à des objectifs donnés.

Parfois, l'évaluation est conçue comme un processus externe à l'apprentissage, visant à l'attester (*évaluation sommative*, avec ou sans notation, éventuellement *certifiante*).

En réalité, l'évaluation doit s'entendre comme un formidable outil d'aide à l'apprentissage, permettant d'anticiper et de diagnostiquer un besoin (*évaluation diagnostique*) et de fournir des retours nécessaires pour perfectionner les acquis ou en prendre conscience (*évaluation formative*). Cependant, cette manière d'envisager l'évaluation peut être contre-intuitive pour les élèves et les enseignants, et doit donc faire l'objet de clarification.

La temporalité de l'évaluation doit également faire l'objet d'une réflexion approfondie. L'évaluation doit être pensée en amont de la mise en œuvre de toute séquence d'enseignement pour faire partie intégrante des apprentissages. Avant d'évaluer, il faut en effet déterminer les objectifs et les enjeux des apprentissages et de l'évaluation, et se questionner sur les méthodes les plus adaptées pour s'assurer que les objectifs sont atteints.

Ci-dessous, vous trouverez des réponses à des questions qui se posent souvent, en lien avec le recours à l'évaluation.



La main à la pâte propose MAPÉVAL, un dispositif pour faciliter l'évaluation des connaissances et des compétences scientifiques (*L'évaluation au service des apprentissages en sciences*). L'outil MAPÉVAL a également pour vocation de faciliter la planification et la clarification des objectifs d'apprentissage, pour l'enseignant comme pour les élèves.

Q1. Quelles indications provenant de la recherche pour bien évaluer?

Parmi les diverses indications qui émergent de la littérature sur la conception et la conduite d'évaluations pédagogiques (au service des apprentissages), on peut retenir les suivantes :

Quelques points d'attention pour une évaluation au service des apprentissages

- ▶ Penser l'évaluation comme faisant partie intégrante des enseignements :
- ▶ l'évaluation est définie lors de la préparation des séquences, en amont de leur mise en œuvre ;
 - elle permet une clarification des objectifs d'acquisition de connaissances et de compétences, de savoir-faire, de capacités ;
 - elle est garante de la cohérence entre ce qui est enseigné, appris et évalué.
- ▶ Définir clairement les critères de réussite ou les attendus, en termes d'objectifs, de compétences ou de connaissances, de savoir-faire, de capacités à faire acquérir.
- ▶ Définir clairement ce qui sera observé chez les élèves (indicateurs) pour confirmer que les attendus sont atteints.
- ▶ Spécifier les outils à utiliser pour réaliser les mesures ou les observations.
- ▶ Utiliser les évaluations comme soutien pour les échanges entre enseignants et élèves, pour permettre à ces derniers de :
 - bien saisir les objectifs d'apprentissage ;
 - suivre leur propre progression ;
 - s'autoévaluer ;
 - comprendre les retours faits par l'enseignant ;
 - gagner progressivement en autonomie via l'autorégulation.
- ▶ Veiller à fournir un retour bienveillant et informatif. Toute évaluation a un impact émotionnel, mais aussi cognitif. Les apprenants doivent recevoir des conseils constructifs sur la manière de s'améliorer, et pas uniquement sur ce qui « va » ou « ne va pas » dans leur prestation.
- ▶ S'assurer de ne pas renforcer des stéréotypes (de genre, par exemple), très présents en sciences, ni les inégalités sociales, lors de la définition des évaluations.
- ▶ Ne pas ignorer les biais qui peuvent affecter les évaluations (biais de comparaison et d'ordre des copies, par exemple), afin que celles-ci soient plus justes.
- ▶ Modifier chez les élèves la perception de l'évaluation, qui doit rester un outil au service des apprentissages. Autrement dit, ne pas concevoir l'évaluation comme un outil de comparaison entre apprenants ou associer résultats et capacités intellectuelles.

En ce qui concerne les étapes clés de l'évaluation (formative), on peut en identifier quatre :

- ▶ Identifier ce que les élèves ont besoin d'apprendre :
 - par rapport aux contenus des curricula qui vont constituer les objectifs d'apprentissage ;
 - par rapport aux prérequis que les élèves doivent posséder pour développer les connaissances et les compétences qui font l'objet de l'apprentissage, mais aussi par rapport aux difficultés que l'on peut rencontrer dans l'apprentissage de certains concepts ou savoir-faire, aux idées fausses courantes...
- ▶ Définir clairement et partager avec les élèves les objectifs d'apprentissage et les critères de succès qui vont permettre de les valider.
- ▶ Vérifier régulièrement la progression d'apprentissage des élèves :
 - fixer un plan d'évaluation avec les moyens à prévoir ;
 - collecter les informations via les moyens établis et les interpréter pour prendre des décisions sur la manière d'intervenir, aider... ;
 - utiliser une variété de formes d'évaluation pour chercher à objectiver les effets des pratiques d'enseignement choisies sur l'apprentissage des élèves.
- ▶ Fournir un feed-back informatif, permettant de se corriger, de savoir ce qui a bien ou moins bien fonctionné et comment mieux étudier ; échanger autour de ce feed-back avec les élèves et les aider à l'utiliser pour mieux apprendre.

Q2. Quels bénéfices pour l'enseignant?

L'utilisation régulière des évaluations permet à l'enseignant :

De suivre la progression des acquis des élèves de manière rigoureuse, c'est-à-dire le niveau de compréhension des connaissances, leur mémorisation et la capacité à récupérer et à réutiliser ces connaissances. L'enseignant se rend ainsi compte de l'efficacité de ses actions pédagogiques. Par là, l'évaluation lui permet d'affiner ses pratiques.

De donner un retour qualitatif aux élèves afin qu'ils puissent s'améliorer. Un retour approprié dynamise de manière importante l'apprentissage (voir Q3).

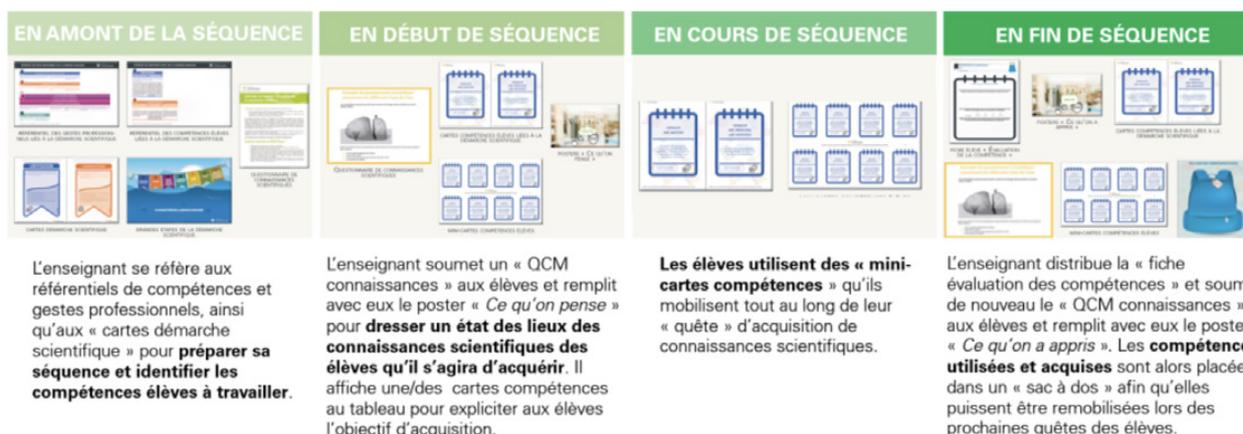
De mettre en place des adaptations ou des activités de remédiation, toujours dans l'optique de favoriser les progrès de tous les élèves pour atteindre les objectifs établis.

Moins souvent évoquée, l'évaluation aide à clarifier les objectifs en termes d'apprentissage, car ce sont eux qui seront évalués.

L'évaluation est un outil structurant puissant. Ceci est particulièrement important dans le domaine de l'enseignement des sciences, où l'on **court** deux risques :

des cours qui se concentrent uniquement sur la transmission et la mémorisation de faits, sans réelle compréhension ;
des cours où les élèves « font », mais ne comprennent pas ce qu'ils font.

Schéma d'utilisation des outils d'évaluation MAPÉVAL



Pour aider l'enseignant à clarifier les objectifs en termes de connaissances et de compétences en sciences (pour soi-même et pour la classe), à suivre la progression des élèves de manière à pouvoir donner un feed-back et à mettre en place des remédiations.

Q3. Quels bénéfices pour les élèves?

L'évaluation permet notamment de :

- ▶ clarifier et expliciter les objectifs d'apprentissage ;
- ▶ proposer des moments formels pendant lesquels les élèves doivent se remémorer les apprentissages (voir Q3.1) ;
- ▶ favoriser l'expression des représentations initiales, de ce que les élèves savent déjà ;
- ▶ prendre conscience de ce que l'on a appris, en allant au-delà du sentiment subjectif de maîtrise ;
- ▶ mobiliser ses connaissances dans des contextes variés.

Par l'explicitation des objectifs de connaissances et de compétences, l'évaluation favorise ainsi :

- ▶ la mémorisation ;
- ▶ la focalisation de l'attention, l'engagement actif ou métacognitif au service de l'autorégulation (le suivi de son propre apprentissage, la compréhension des procédures qui ont bien fonctionné et de celles qui n'ont pas donné le résultat espéré) ;
- ▶ la motivation (si certains principes de l'évaluation sont respectés) ;
- ▶ l'évolution des représentations et l'utilisation de l'erreur comme moteur d'apprentissage ;
- ▶ la généralisation et le transfert des apprentissages.

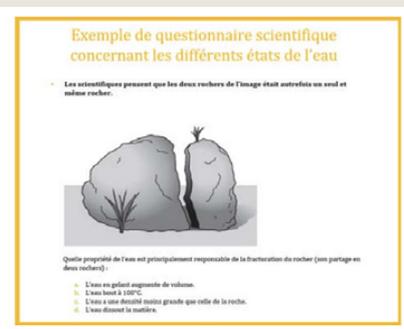
(Pour plus de détails sur les aspects cognitifs, voir questions Q3.1 à Q3.6.)

Q3.1 Quels bénéfices pour la mémorisation (rétention et récupération)?

L'évaluation permet la récupération active en mémoire des contenus appris. La recherche expérimentale a confirmé que la pratique de la récupération en mémoire (des connaissances, des concepts, des compétences) favorise la construction d'un apprentissage durable (rétention) et la capacité de récupérer les informations en mémoire (récupération). Ceci, notamment, si l'on compare l'exercice actif de la récupération en mémoire à d'autres stratégies visant la mémorisation, et plus particulièrement celle qui consiste à relire un même texte plusieurs fois : pour une même quantité de temps passé à relire ou à récupérer activement un souvenir en mémoire (en répondant à des QCM, en se posant des questions et en y répondant...), c'est cette dernière qui produit un effet plus grand et durable sur la mémorisation.

La récupération en mémoire peut être proposée avant l'apprentissage, immédiatement après, puis à distance variable de temps, de façon espacée et répétée (proposée plusieurs fois pour un même contenu d'apprentissage sur une période scolaire). Pour le moment, il n'existe pas d'indication claire sur le laps de temps nécessaire entre l'apprentissage et la phase de récupération en mémoire, ou entre plusieurs phases de récupération en mémoire. L'intervalle doit être suffisamment long pour oublier un peu, mais pas trop, entre une évaluation et une autre (entre un jour et un mois).

Outil MAPÉVAL - Fiche pour l'évaluation des compétences scientifiques et exemple de QCM pour l'évaluation des connaissances



L'évaluation des connaissances est proposée avant et après l'apprentissage. Celle des compétences uniquement après la séance ou un ensemble de séances.

Q3.2 Quels bénéfices apporte le retour qui suit l'évaluation?

Il existe une vaste littérature sur le feedback, qui en atteste les effets positifs importants, à bas coût de mise en œuvre, pour toutes les tranches d'âge et dans toutes les disciplines.

- ▶ Le retour permet de corriger les représentations des élèves, et donc de les faire évoluer.
- ▶ Donner un retour à l'élève permet de corriger la sensation d'avoir appris quelque chose, alors que ce n'est pas le cas, ou celle de manque de confiance, alors que la réponse donnée est correcte.

Donner un retour en relation avec l'évaluation est donc important pour obtenir des effets positifs sur les apprentissages. Voici quelques points d'attention pour des retours efficaces :

- ▶ Les retours sous forme de notes (chiffrées, icônes, couleurs) sans aucune explication ont de faibles effets. Le retour correctif et riche d'informations se montre plus efficace parce qu'il fournit à l'élève l'information nécessaire pour se corriger. Ce type de retour doit contenir les informations concernant les procédures à employer pour accomplir la tâche demandée. Il aide donc non seulement à comprendre quelles erreurs l'élève a commises, mais aussi comment les éviter la prochaine fois. Il permet de faire avancer sa réflexion.

- ▶ Le retour doit être donné dans un délai raisonnable pour que l'élève puisse le relier à la tâche à corriger et l'exploiter.
- ▶ Le retour doit concerner aussi bien les erreurs à éviter que les réponses correctes. En effet, le feed-back n'agit pas uniquement sur le contenu de la réponse, mais aussi sur la sensation métacognitive d'avoir donné ou pas une réponse correcte. Ainsi, il permet de corriger une éventuelle incohérence entre la nature de la réponse (correcte ou pas) et la confiance que l'élève met dans sa réponse (je pense avoir donné la réponse correcte/je ne me sens pas sûr d'avoir donné la réponse correcte).

Q3.3 Quels bénéfices pour la capacité à gérer ses apprentissages?

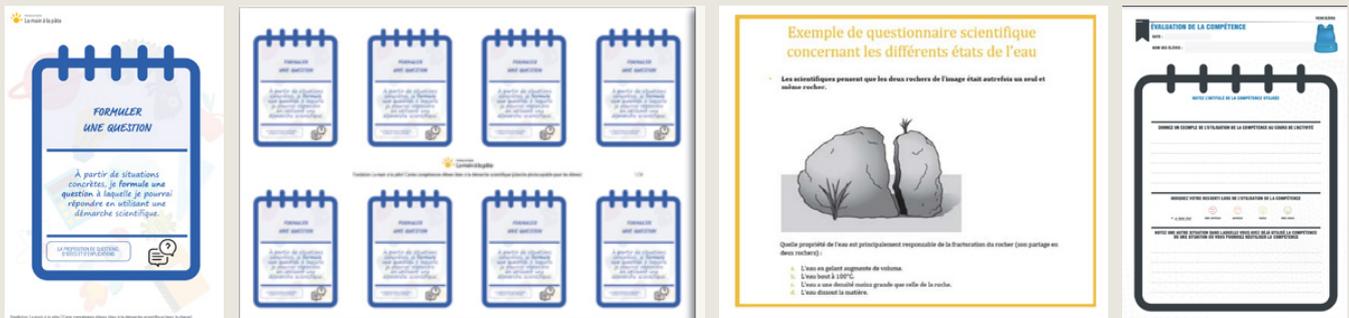
L'évaluation peut jouer un rôle dans l'autorégulation, à plusieurs niveaux.

Lorsqu'elle est proposée en début d'apprentissage, elle permet à l'élève de saisir les objectifs d'apprentissage, et donc de s'engager métacognitivement (de façon réflexive) dans la tâche à venir, aussi bien que de focaliser son attention sur ce qu'il faudra retenir.

Lorsqu'elle est proposée après ou en cours d'apprentissage, l'évaluation permet à l'élève d'analyser par lui-même sa progression, et donc de prendre le contrôle sur celle-ci.

L'évaluation avec retour permet en outre de mieux calibrer la sensibilité métacognitive – définie comme la correcte perception de savoir (d'avoir compris et appris) ou de ne pas savoir (voir Q3.2). Ceci est important pour permettre de doser son effort et de focaliser son attention sur les bons objectifs.

Outil MAPÉVAL - Éléments permettant l'explicitation des compétences et des connaissances scientifiques travaillées, et le suivi par l'élève de ses progrès



Les élèves disposent de cartes qui leur permettent de saisir clairement les objectifs d'apprentissage en termes de compétences. Les connaissances à acquérir sont explicitées par l'enseignant en début de séance et cette explicitation s'accompagne de quiz d'évaluation des connaissances, à proposer en amont de la séance, ce qui permet de focaliser l'attention sur les principaux contenus qui seront à comprendre et à retenir. Le même quiz sur les connaissances est proposé à la fin de l'apprentissage pour permettre à l'élève de se rendre compte de sa progression. Une évaluation finale des compétences permet également de vérifier si la compétence travaillée a été réellement comprise au point de pouvoir l'exploiter, voire de pouvoir la transférer.

La fiche dédiée à l'évaluation des compétences comporte des émojis pour l'autoévaluation du sentiment d'auto-efficacité dans la mise en place de la compétence travaillée. Ceci permet aux élèves de se poser explicitement la question de leur degré de confiance dans leur capacité à utiliser et à réutiliser la compétence.

Q3.4 Quels effets sur la motivation?

L'évaluation peut jouer un rôle positif sur la motivation. Par exemple, elle permet à l'élève de se rendre compte de sa compréhension limitée d'un phénomène (« Je pensais savoir, mais quand j'ai dû répondre, je me suis rendu compte que c'était confus »), ce qui peut stimuler sa curiosité.

Ceci, à condition d'agir positivement sur le sentiment de confiance en soi ou le sentiment d'auto-efficacité de l'élève (croire en sa capacité de mener à bien une tâche). Il s'agit pour cela :

- ▶ De proposer à l'élève des évaluations objectives, de manière à corriger l'évaluation qui passe par les sentiments implicites d'auto-efficacité, de confiance et de réussite (et qui ne sont pas toujours fiables).
- ▶ De lui fournir un retour bienveillant, et notamment de dissocier le feed-back portant sur la personne de celui portant sur l'apprentissage, et de favoriser ce deuxième. Il s'agira de présenter les erreurs comme des étapes indispensables de l'apprentissage et de donner les outils nécessaires pour se corriger, et de permettre ainsi à l'élève de gagner en sentiment d'efficacité (voir Q3.2).
- ▶ De mettre l'élève en condition de succès via des tâches d'une difficulté bien calibrée par rapport à ses compétences.

Q3.5 Quels bénéfices pour la capacité à exploiter les représentations initiales et les «erreurs»?

L'évaluation permet d'exprimer ses conceptions ou ses représentations de façon explicite. Connaître les conceptions et les préconceptions des enfants permet à l'enseignant de les prendre en compte pour proposer des situations adaptées à leur évolution et favoriser l'acquisition de nouvelles connaissances. Les élèves perçoivent ainsi les changements qui s'opèrent au cours même de leurs apprentissages.

L'évaluation permet également d'émettre des hypothèses (ce que l'on pense être la bonne réponse) qui seront éventuellement contredites par les constats faits en cours de séance. Ceci permet donc d'exploiter l'un des mécanismes que notre cerveau utilise pour apprendre, basé sur la prédiction et la correction des erreurs à la lumière des informations collectées.

Outil MAPÉVAL - Éléments permettant de révéler les conceptions initiales des élèves et leur évolution à la suite de l'activité menée en classe



En début de séance, les élèves sont amenés à se rendre compte de leur manque de connaissances et de leurs propres représentations grâce à un quiz et à une discussion sur ce que l'on pense. À la fin de la séance, la proposition du test initial, cette fois avec feed-back, et le poster Ce qu'on a appris permettent d'objectiver le changement de représentations par rapport au début de la séance.

Q3.6 Quels bénéfices pour le transfert des apprentissages à de nouvelles situations?

Arriver à réutiliser, dans un contexte différent, ce que l'on a appris est bel et bien le but d'un apprentissage de qualité. Et pourtant, le transfert est l'un des principaux défis de l'apprentissage. L'évaluation peut constituer un outil d'aide au transfert car elle peut fournir des occasions de s'entraîner à transférer. Il est également possible d'évaluer la capacité de transfert de l'élève, en lui proposant des situations de plus en plus éloignées du contexte initial.

Outil MAPÉVAL - Éléments permettant de conceptualiser les compétences travaillées



La fiche dédiée à l'évaluation des compétences permet aux élèves d'envisager d'autres contextes d'utilisation des compétences travaillées. Elle les invite donc à considérer la généralisation ou le transfert des compétences. Le sac à dos des compétences élèves permet de garder en mémoire les compétences déjà abordées.

La frise des grandes étapes de la démarche scientifique permet de se référer aux grandes catégories de compétences que les élèves auront à maîtriser progressivement au cours de leur scolarité.

Q4. Que peut-on évaluer en sciences?

Dans le domaine de l'apprentissage des sciences, l'évaluation peut porter sur différents objectifs d'apprentissage et de maîtrise :

- ▶ Évaluation des *connaissances* (connaissances en sciences) et de la *compréhension des phénomènes naturels* (savoir expliquer les phénomènes étudiés, en termes scientifiques).
- ▶ Évaluation des *savoir-faire et des compétences* relatives à la *démarche scientifique* (savoir raisonner de façon scientifique, savoir concevoir une investigation scientifique, savoir interpréter scientifiquement données et preuves...).
- ▶ Évaluation des *connaissances relatives à la nature de la science* : connaître la manière dont les connaissances scientifiques sont produites et savoir distinguer ces connaissances d'opinions, par exemple.
- ▶ Évaluation de la *culture ou littératie scientifique* : savoir utiliser les connaissances et les compétences scientifiques dans la vie quotidienne.
- ▶ Évaluation des *attitudes, motivations et valeurs* des apprenants, en lien avec l'utilisation des connaissances et des compétences scientifiques.

(Pour des exemples de tests standardisés concernant ces différents contenus d'évaluation, voir Q10.2)

Outil MAPÉVAL - Éléments permettant d'évaluer les connaissances et les compétences travaillées



Le dispositif MAPÉVAL comporte des évaluations des connaissances, sous forme de quiz, ainsi que des recommandations pour la production de quiz efficaces.



Le dispositif MAPÉVAL comporte également une évaluation spécifique des savoir-faire et des compétences propres à la maîtrise de la démarche scientifique, à la compréhension de la nature de la science et à la capacité à utiliser ces acquis pour examiner opinions et arguments, ou pour imaginer des applications à de nouveaux problèmes.

Q5. Quelles modalités adopter pour l'évaluation en sciences?

Les connaissances, compétences, attitudes et comportements que l'on souhaite favoriser chez les élèves sont typiquement consignés dans les programmes scolaires ou sous forme de standards (nationaux et internationaux, voir Q5.1), ou encore sous forme de référentiels ou de grilles (« rubrics » et « grids » en anglais, voir Q5.2, Q5.2, Q5.3). Les référentiels peuvent venir compléter les programmes scolaires ou être élaborés par un établissement scolaire, un groupe d'enseignants, un organisme de formation, un producteur et fournisseur de ressources pédagogiques...

Les standards et les référentiels servent à leur tour de guides pour la production de grilles critériées d'évaluation et de notation.

Q5.1 Qu'est-ce qu'un standard?

En éducation, la notion de standard revient à décrire ce que les élèves sont censés apprendre et ce que les enseignants sont censés enseigner, c'est-à-dire le curriculum. Les évaluations basées sur les standards mesurent le niveau d'acquisition des connaissances et des compétences. Le résultat d'une évaluation peut être rapporté sous forme de notes ou équivaloir à un passage au niveau suivant. Il existe donc, associée à celle de standard, une notion d'acquisition minimale, de seuil pour déclarer un élève suffisamment compétent dans un domaine. Les standards sont fortement associés aux curricula nationaux dans les différentes disciplines, car ils définissent ce que les élèves sont censés apprendre à un niveau particulier d'éducation.

En France, les curricula ou **programmes scolaires** de l'enseignement des sciences définissent des attendus faisant partie du socle commun de connaissances, de compétences et de culture établi comme objectif pour tous les élèves entre 6 et 16 ans.

Les programmes spécifient les connaissances et les compétences à travailler pour atteindre les objectifs du socle commun, et sont accompagnés de ressources pédagogiques pouvant illustrer des mises en pratique possibles.

Il existe aussi des standards internationaux, définis en sciences par des organismes tels que la **IEA** (pour les standards relatifs aux évaluations **Timss**) et l'**OCDE** (pour les standards des évaluations **PISA**). Cependant, l'objectif de ces standards et de ces évaluations ne devrait pas être confondu avec celui des standards nationaux, car l'objectif principal des évaluations internationales est une évaluation comparative entre systèmes éducatifs.

Outil MAPÉVAL - Connaissances et compétences visées

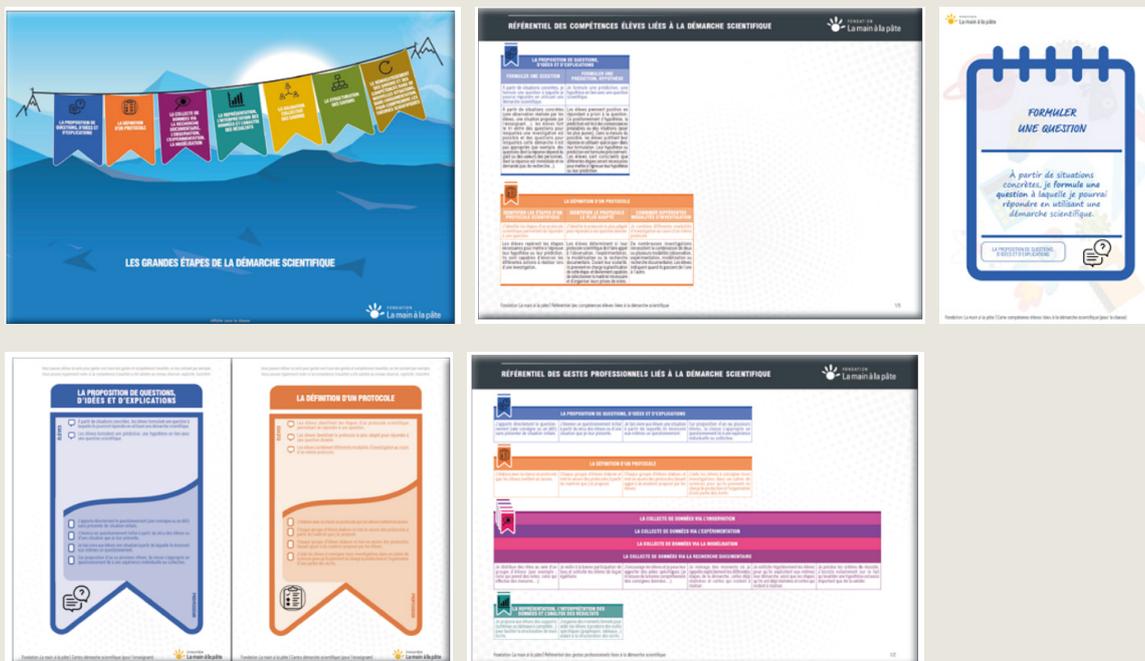


Le dispositif MAPÉVAL se base sur les programmes scolaires français et fait notamment référence à ceux-ci pour tout ce qui concerne l'évaluation des connaissances et les attendus, en termes de compétences scientifiques. Cependant, l'évaluation des compétences se base sur un référentiel détaillé qui inclut également des éléments de culture scientifique et d'esprit critique en lien avec les sciences.

Q5.2 Qu'est-ce qu'un référentiel?

Un référentiel est généralement un ensemble de lignes directrices (une grille de critères), utilisé pour promouvoir une application cohérente des attentes, des objectifs ou des normes d'apprentissage et pour en mesurer la réalisation. Un référentiel peut donc être utilisé à la fois pour déclarer des objectifs attendus et pour servir de base à leur évaluation.

Un exemple de référentiel des compétences scientifiques, inclus dans l'outil MAPÉVAL



La frise des grandes étapes ou savoir-faire de la démarche scientifique permet de se rappeler les grandes catégories de savoir-faire et de compétences à maîtriser progressivement, dans différents contextes et par rapport à différents contenus. Ceux-ci sont regroupés dans un référentiel des compétences liées à la démarche scientifique. Chaque compétence est rapportée sur des supports à mobiliser en classe, afin de rendre les élèves conscients des savoir-faire et des compétences travaillées.

À cela s'ajoutent un référentiel et des cartes relatifs aux gestes professionnels mobilisés par l'enseignant pour favoriser l'acquisition des savoir-faire et des compétences correspondantes.

Q5.3 Qu'est-ce qu'une grille d'évaluation?

Une grille d'évaluation est un outil qui rend explicite une série de critères d'évaluation normés, en lien avec les performances attendues. De cette manière, elle aide les enseignants à appliquer des normes cohérentes lors de l'évaluation. Les grilles peuvent comporter des niveaux de progression (échelle de niveau) et des éléments gradués qui permettent d'aboutir à une note (échelle de notation).

Les grilles d'évaluation peuvent être conçues pour n'importe quel domaine et être utilisées pour évaluer l'apprentissage à tous les niveaux.

Elles constituent un outil communément utilisé dans le cadre des approches éducatives dites fondées sur les résultats ou sur les objectifs*, tout comme dans les approches dites par compétences**.

(Pour des exemples de grilles d'évaluation, voir Q10.2)

*L'éducation basée sur les résultats ou les objectifs

est une théorie éducative qui se fonde sur l'idée de fixer des objectifs à atteindre pour chaque élève. Les objectifs (ou standards ou compétences à atteindre) sont des comportements ou des réponses observables et mesurables. Les objectifs déterminent avec précision le résultat souhaité, au terme d'un processus d'enseignement. Ils permettent donc de concevoir des curricula, mais également des épreuves d'évaluation standardisées, basées sur des listes d'objectifs d'apprentissage, ordonnés selon des taxonomies précises. La plus connue est celle proposée par Benjamin Bloom en 1956. On y constate une hiérarchie d'objectifs de complexité croissante. L'élève est donc progressivement exposé à des problèmes de plus en plus complexes et évalué en relation avec des objectifs de niveau cognitif de plus en plus haut. On peut retrouver ce type d'approche dans les évaluations Timss, qui sont organisées autour de trois grands domaines cognitifs : connaître, appliquer et raisonner. Les méthodes basées sur les résultats ont été adoptées dans les systèmes éducatifs du monde entier, à différents niveaux. Parmi les plus connus, le programme américain, qui a donné lieu en 2016 aux National Science Education Standards. Les avantages des standards ou des objectifs clairement identifiés se réfèrent notamment au fait de pouvoir facilement identifier les buts à atteindre dans le cadre d'un apprentissage et de pouvoir facilement concevoir des épreuves standardisées avec un haut niveau de validité et de fidélité. Ces dispositifs favorisent notamment l'évaluation sommative et certificative. Ce modèle a cependant reçu de nombreuses critiques, notamment du fait que :

- ▶ les objectifs sont, dans la plupart des cas, très nombreux ;
- ▶ les objectifs d'un certain niveau ne sont pas connectés entre eux – chaque objectif fait l'objet d'une évaluation à part et n'a pas de liens explicites avec d'autres objectifs qui concourent pourtant à préparer l'élève à passer à un niveau supérieur de compétence ;
- ▶ les objectifs ne spécifient pas comment atteindre les résultats attendus.

****L'approche par compétences** porte une attention particulière aux référentiels de compétences, qui en constituent un outil essentiel. L'évaluation y est plutôt conçue comme une évaluation de type formatif (mais pas uniquement). Une compétence (« competency ») associe connaissances, savoir-faire ou habileté (« skill ») dans le but de savoir agir. Au centre de cette approche se trouvent l'idée de tâche complexe à accomplir (l'agir) et celle de la capacité à mobiliser, à utiliser et à transférer une compétence pour la réalisation de telles tâches. Par conséquent, l'évaluation ne se limite pas à observer si un objectif est atteint, mais porte sur la maîtrise des ressources nécessaires pour résoudre la tâche. Ces ressources sont variées : les savoirs comprennent les connaissances factuelles, mais aussi celles procédurales, conceptuelles, métacognitives ; aux savoir-faire sont associés aussi des savoir-être ou des attitudes. La notion de compétence incluant en plus celle de transférabilité à de nouveaux

contextes et situations, les critères d'évaluation sont multiples et doivent porter sur des situations réelles, avec leur contexte (les contextes, et donc les situations d'évaluation, devant être variés). Il ne faudrait pas confondre (mais cela arrive souvent dans les programmes ou les curricula) les compétences au sens propre et les connaissances ou les savoir-faire. Par exemple, savoir écrire une formule décrivant une réaction chimique n'est pas une compétence à proprement parler, mais une connaissance, ou alors un savoir-faire. Il est important de définir comment observer si les compétences sont maîtrisées et à quel niveau, et donc de définir des indicateurs et des observables auxquels se référer pour évaluer. Or, une compétence est au fond une potentialité, au sens où elle peut s'exprimer de manière différente en fonction du contexte et qu'elle n'est pas immédiatement visible. Par exemple, un même élève peut faire montre de l'utilisation correcte d'une compétence liée à la démarche expérimentale (comme proposer un protocole expérimental) dans un cadre spécifique, semblable à d'autres cas déjà rencontrés, mais ne pas être capable de définir un protocole en relation avec un contenu non familier. De même, l'élève peut réussir à faire montre de ses compétences lors d'une évaluation sans enjeu, et ne pas réussir à les manifester dans le cadre d'une épreuve notée et avec enjeu. Évaluer des compétences est donc beaucoup plus difficile qu'évaluer des résultats. La maîtrise d'une compétence comportant le fait de savoir la mobiliser dans diverses circonstances, ceci implique pour l'élève de se rendre progressivement moins dépendant du contexte. Il est donc important de multiplier les observations de la même compétence pour un même élève dans des contextes différents. Enfin, la notion de compétence étant destinée au transfert dans la vie réelle, les moyens pour l'évaluer doivent faire référence à des situations complexes et aussi réalistes que possible. Or, ceci est particulièrement difficile lorsque l'on doit concevoir des tâches d'évaluation qui font référence à des situations réelles, mais qui s'adressent à des élèves au début de leurs apprentissages. Pour toutes ces raisons, l'approche par compétences est complexe non seulement dans son évaluation, mais également dans sa mise en pratique.

L'approche par compétences peut être considérée comme un développement de l'approche par objectifs, ou comme une conséquence du passage d'approches de type behavioriste (où ce qui compte est le résultat final observable) à des approches cognitivistes, éventuellement socioconstructivistes. Les bases scientifiques des deux approches pédagogiques restent à établir, ainsi que leur efficacité. Le rôle des grilles est le même dans les deux approches, mais change les formulations des critères utilisés comme référence pour l'évaluation. On peut ainsi retenir des aspects communs aux deux approches, en relation avec les problématiques d'évaluation :

- ▶ l'importance de fixer des objectifs clairs et bien définis ;
- ▶ celle d'identifier, via l'évaluation, le niveau de maîtrise ;
- ▶ avec une visée d'utilisation et de transfert.

Q5.3.1 Quelles sont les composantes d'une grille ou d'un référentiel pour l'évaluation?

Lorsque l'évaluation repose sur des grilles critériées, celles-ci doivent comprendre des critères d'évaluation, des indicateurs – à leur tour renseignés par des observations ou des mesures, éventuellement des échelles de progression (voir tableau ci-dessous pour une explication de ces termes ; voir les encadrés des Q5.3 et Q6. pour des exemples concrets). Il faut souligner qu'une évaluation n'est pas un processus mécanique et que les indicateurs une fois collectés sont soumis à interprétation. Par exemple, s'il y a plusieurs indicateurs portant sur un même critère, l'élève peut valider certains de ces indicateurs uniquement. En outre, l'évaluation ne se termine pas par l'interprétation, mais doit donner lieu à une restitution, à un retour aux élèves.

Critères d'évaluation. Il s'agit des qualités recherchées, en lien avec la tâche visée ou l'objectif à atteindre ; par exemple, le fait de savoir reconnaître un protocole expérimental rigoureux, de savoir en proposer un ou, finalement, de savoir le mettre en œuvre constitue un ensemble de compétences liées à la démarche scientifique.

Indicateurs. Les critères n'étant pas directement observables, il est important de se poser la question de comment on pourra réellement et concrètement en observer/mesurer la réalisation. Il est alors nécessaire d'identifier pour chaque critère des indicateurs qui spécifient de manière non ambiguë comment les attentes se traduisent dans la réalité. Par conséquent, les indicateurs indiquent à l'évaluateur ce qu'il doit observer ou mesurer pour qu'un élève valide un critère. Les indicateurs doivent également être compréhensibles par les élèves afin de rendre explicites les attentes (voir Q3).

On peut faire une distinction générale entre indicateurs observables – comportements, réponses aux questions, qualité du geste – et indicateurs qui peuvent être mesurés – nombre de réponses correctes, taux de réussite à un examen en classe...

Outils d'observation et de mesure. Les indicateurs ne sont pas « autonomes ». L'évaluateur doit collecter des données – et cela peut se faire de différentes manières, à partir de différents types de productions et de tests. Le passage des critères aux indicateurs fournit le squelette et la structure des évaluations. L'indicateur est donc le résultat d'un agrégat d'éléments que les évaluateurs rassemblent. Les interrogations, sous forme de questionnaires (ouverts ou fermés), les observations d'écrits, les observations de classe, les interactions professeurs-élève(s) ou élève(s)-élève(s) constituent autant de modalités de collecte des informations en relation avec la grille d'évaluation et ses indicateurs.

Par exemple, dans le cadre des objectifs liés à la maîtrise progressive de la démarche d'investigation, une compétence fondamentale consiste à savoir collecter des données, via différentes modalités, dont l'observation. Cette dernière doit être préparée, puis menée de façon précise, détaillée, en utilisant les instruments et les moyens les plus adaptés, et de façon à limiter les biais (par exemple, grâce à la multiplication des observations). La compétence « Savoir collecter des données via l'observation » se divise donc en plusieurs sous-compétences spécifiques, à leur tour rendues observables ou mesurables grâce à des comportements (gestes, communication orale ou écrite), à des réponses à des questions... Ainsi, pour renseigner les indicateurs, l'enseignant pourra observer le comportement des élèves pendant les activités de classe et voir si les actions accomplies correspondent à celles décrites dans le référentiel au niveau des indicateurs ou des observables. L'enseignant pourra également consulter les écrits des élèves, dont le cahier d'expériences qui est rempli pendant les investigations, pour vérifier si les éléments recherchés sont présents. Enfin, il pourra vérifier le niveau de maîtrise de la compétence via des interrogations orales ou écrites qui amènent les élèves à s'exprimer de façon plus ou moins libre.

Q6. Quels outils de mesure peuvent être mobilisés pour l'évaluation?

Lors de la planification de l'évaluation des apprentissages, différents outils seront choisis : tests écrits, travaux pratiques... Le choix des outils qui peuvent être mobilisés pour renseigner les indicateurs (voir Q5.3.1), et donc, au final, les critères choisis pour l'évaluation, dépend de plusieurs considérations : leur ergonomie et leur facilité d'usage ; les contenus à évaluer et le contexte ; les limites propres à chaque outil d'évaluation ; la finalité de l'évaluation.

Q6.1 En quoi consistent les tests écrits (avec des questions à choix multiples ou des questions ouvertes)?

Dans le cadre d'un test écrit avec questions à choix multiples (QCM), l'élève choisit sa réponse. Il peut aussi rédiger des réponses courtes ou longues dans ses propres mots pour justifier ses réponses. Ce genre de test permet d'évaluer un large éventail de contenus dans un laps de temps relativement court. Il s'adapte bien à vérifier si les apprenants ont acquis certaines connaissances ou compris des concepts.

Les réponses ouvertes sont plus riches d'informations, mais plus difficiles à évaluer. Elles supposent d'explicitier préalablement les réponses que l'on s'attend à obtenir de la part des élèves. Par exemple, lister les mots qui devront apparaître dans la réponse ou donner des exemples de réponses correctes et incorrectes.

L'association de QCM et de réponses ouvertes est particulièrement favorable à la mémorisation, et donc à l'apprentissage. Cependant, tous les tests écrits – y compris ceux à choix multiples – peuvent être utilisés aussi bien pour évaluer les connaissances que pour l'évaluation des capacités de raisonnement et d'application, comme en témoignent les items des tests IEA-Timss, qui visent à évaluer les trois domaines cognitifs.

Des points de vigilance sont à garder à l'esprit. La plupart des tests écrits impliquent une part de lecture et d'écriture pouvant constituer un obstacle et réduire la validité du test, notamment pour les jeunes enfants ou les élèves ayant des difficultés d'ordre linguistique (voir Q9).

(Pour des exemples de tests standardisés concernant différents contenus d'évaluation, voir Q10.2)

Outil MAPÉVAL - Éléments permettant d'évaluer les connaissances travaillées



Le dispositif MAPÉVAL comporte des évaluations des connaissances, sous forme de quiz, ainsi que des recommandations pour la production de quiz efficaces.

Q6.2 En quoi consistent les tests portant sur les pratiques?

Ce type de tests permet d'évaluer les connaissances relatives à l'application de procédures ou d'un raisonnement lié à ces procédures.

Ces tests présentent l'avantage de réduire l'impact de la maîtrise de la lecture et de l'écriture sur l'évaluation d'une compétence. Par contre, ils restent dépendants des effets de contexte (voir Q9.). En effet, les compétences des élèves se révèlent plus facilement dans le cadre de contextes qui leur sont familiers. Par exemple, les élèves qui obtiennent de bons résultats dans une situation d'investigation ne le feront pas nécessairement dans une autre testant les mêmes compétences, mais proposant un contexte différent.

(Pour des exemples de tests portant sur les pratiques, voir Q10.2)

Q.6.3 En quoi consistent les observations de classe basées sur des grilles d'observation et de notation?

Les observations que l'enseignant mène en classe pendant les activités sont elles aussi des formes d'évaluation. Cependant, sans l'appui d'une grille capable de guider l'observation, elles peuvent présenter de nombreux biais. On utilise alors des grilles, telles que celles décrites en Q5.3.

Les observations basées sur des grilles sont considérées comme plus appropriées pour évaluer les apprentissages cognitifs de haut niveau (selon la taxonomie de Bloom), tels que les capacités d'analyse et de raisonnement (savoir établir des relations, identifier des principes d'organisation d'une situation), de synthèse et d'application (par exemple, produire un travail personnel) ou d'évaluation (par exemple, porter un jugement sur une situation).

Ces grilles sont généralement destinées à être utilisées par l'enseignant (ou le formateur, dans des contextes de développement professionnel). Cependant, dans un contexte d'aide à l'apprentissage, cet outil d'évaluation peut être utilisé pour encourager les élèves à analyser ce qu'ils ont appris et la manière dont ils ont acquis des connaissances et des compétences. Ainsi, la grille d'évaluation peut être utilisée pour l'autoévaluation ou la coévaluation, car elle fournit à l'apprenant des informations précises sur ses acquis et lui permet de cibler les contenus d'apprentissage sur lesquels il doit encore travailler.

(Pour des exemples de grilles critériées utilisées pour l'évaluation avec notation, voir Q10.2)

Q7. Quand conduire des évaluations?

La fréquence des évaluations dépend de leur rôle. Si les évaluations sommatives se concentrent sur des périodes précises, et notamment en fin d'apprentissage, celles formatives, de par leur impact positif sur l'apprentissage, doivent avoir lieu tout au long de celui-ci (voir Q3.).

Pensée comme un outil au service de la pratique de la récupération en mémoire, l'évaluation peut être proposée :

- ▶ Avant l'apprentissage.
- ▶ Immédiatement après.
- ▶ Puis à distance variable de temps, de façon espacée et répétée (proposée plusieurs fois pour un même contenu d'apprentissage sur une période scolaire). La pratique espacée consiste par exemple à proposer des quiz ou d'autres formes d'évaluation concernant des thèmes abordés à distance de semaines. Pour le moment, il n'existe pas d'indication claire sur le laps de temps nécessaire entre l'apprentissage et la phase de récupération en mémoire, ou entre plusieurs phases de récupération en mémoire. L'intervalle doit être suffisamment long pour oublier un peu, mais pas trop, entre une évaluation et une autre (entre un jour et un mois). La pratique espacée et répétée se montre particulièrement efficace, mais ceci ne signifie pas que la récupération « simple » après le cours ne le soit pas.

Q8. Comment fournir un retour à la suite de l'évaluation?

L'évaluation peut être menée par l'enseignant, par les pairs ou par l'élève lui-même (cas d'une autoévaluation). Par exemple, en utilisant des quiz ou des *flashcards* pour s'autoévaluer pendant une révision, l'élève se rend compte tout de suite de ce qu'il a appris et s'il a besoin de réviser certaines connaissances en particulier. Dans ce cas, le contexte est celui d'une évaluation formative et elle peut avoir lieu aussi souvent que possible.

L'évaluation peut utiliser des supports technologiques qui fournissent un retour immédiat à l'élève, sans nécessairement passer par l'enseignant. Ce type d'évaluation comporte des avantages en termes de praticité : l'enseignant peut préparer, à l'avance, une batterie de tests et facilement la faire évoluer d'année en année. On peut personnaliser ces tests en fonction du niveau des élèves. Ils sont disponibles à tout moment durant les heures de cours ou en dehors des moments de classe. Leur longueur peut varier. Cependant, les retours d'information sont plus efficaces quand ils sont donnés par une personne qui argumente et personnalise son retour en fonction des élèves (voir Q3).

Q9. Quelles précautions prendre dans la création, l'administration, l'interprétation d'évaluations?

Aucun outil d'évaluation n'est parfait ! Le recours à une variété d'outils permettra des évaluations plus justes, en termes de validité, de fiabilité et d'équité.

Ceci est particulièrement important dans le cadre des évaluations sommatives, comme en témoigne l'existence d'une discipline destinée à favoriser la production d'évaluations justes : la docimologie. Dans le cas des évaluations sommatives, les tests sont généralement notés à l'aide d'une grille de notation prédéfinie, soit par l'enseignant de l'élève, soit par des correcteurs externes, qui sont souvent des professeurs d'autres écoles, soit par des machines (voir Q8). Certaines évaluations sont présentées à tous les élèves de la même manière. La raison de cette uniformité des procédures est de permettre de comparer les résultats des élèves passant les tests à des endroits différents. L'évaluation à des fins sommatives doit en effet être aussi fiable (exempte d'erreurs et de biais) que possible.

Dans le cas de l'évaluation formative, par définition, l'enjeu est moindre. Cependant, les points de vigilance qui s'appliquent aux évaluations sommatives peuvent être également dans la conception d'évaluations formatives, afin d'éviter de mettre les élèves en difficulté ou de mal interpréter leurs performances.

Dans ce qui suit, les férus de l'évaluation pourront trouver quelques éléments de docimologie pour des évaluations plus valides, fiables et justes.

Un peu de docimologie

Les outils produits pour les évaluations et l'interprétation des résultats doivent répondre aux critères de validité, de fiabilité et d'équité.

- ▶ Validité : permettre d'évaluer ce qu'elles disent évaluer, sans impliquer des connaissances ou des compétences supplémentaires, et en permettant de collecter des informations sur les connaissances et les compétences que l'on veut réellement mesurer.
- ▶ Fiabilité : prendre la mesure des connaissances et des compétences de l'élève avec le moins de biais possible influant sur le résultat. Par exemple, une évaluation fiable donne un résultat similaire, même si elle est conduite par différents évaluateurs.
- ▶ Équité : Une évaluation est équitable lorsqu'elle permet à tout le monde d'exprimer ses capacités de la manière la plus égalitaire et libre possible (sans entraves).

Docimologie = discipline qui vise à établir comment répondre au problème de l'objectivité de l'évaluation, et notamment à la problématique de la note. La discipline remonte au début du XX^e siècle. Le terme apparaît en 1929 sous la plume d'Henri Piéron. Si, au début, la docimologie était surtout critique (elle mettait en évidence les problèmes), progressivement, les chercheurs se sont efforcés de proposer des solutions permettant de limiter au maximum le caractère subjectif de la notation.

Q9.1 Comment améliorer la fiabilité des résultats de l'évaluation?

Une évaluation fiable donne un résultat similaire, même si elle est menée par des évaluateurs différents ou plusieurs fois par le même évaluateur.

Dans le cas des évaluations basées sur des standards, il est possible de répondre à ces exigences, car les standards sont des comportements visibles, des résultats décrits de manière objective. Cependant, les exigences de validité et de fiabilité peuvent être contradictoires. Par exemple, les tests à choix multiples sont les évaluations les plus fiables car ils ne laissent aucune place au jugement professionnel dans la notation, mais tout ne peut pas être facilement évalué par un test à choix multiples, comme la capacité à produire des preuves par le biais d'un protocole expérimental ou d'une observation, ou la capacité à évaluer l'information dans toute sa profondeur.

En outre, les évaluations plus qualitatives de compétences complexes (qui se réfèrent à de multiples connaissances et compétences combinées) perdent en fiabilité, dans le sens où différents évaluateurs peuvent donner un jugement différent. En outre, même les tests apparemment fiables peuvent être soumis à des biais lors de la correction des résultats.

Lors de la correction, plusieurs conditions peuvent biaiser (influencer dans un certain sens) le jugement porté sur une copie/un sujet spécifique :

- ▶ Le contexte, lorsqu'un correcteur ajuste son barème en fonction du niveau des élèves : à performances égales, et toutes choses égales par ailleurs, un élève est jugé par son professeur comme un « bon élève » dans une classe, alors qu'il peut être contraint de doubler son année scolaire dans une autre. Tout dépend non pas de la performance particulière de l'élève en termes absolus, mais de sa performance en termes de performance par rapport à celle de ses camarades. De ce fait, les résultats de fin d'année sont souvent distribués selon une courbe de Gauss (quelques élèves « faibles », quelques élèves « forts », et la grande majorité dans la « moyenne »). Ce phénomène s'appelle la loi de Posthumus et résulte d'une sorte de prototype communément admis : il y a peu d'élèves exceptionnels (très faibles ou très bons) et la distribution des notes est très uniforme. Les tests internes aboutissent souvent à ce type de distribution, même si les tests externes, appliqués aux mêmes élèves, indiquent des niveaux moyens très variables et des dispersions différentes.
- ▶ La place de la copie dans la pile/la qualité des autres copies : une même copie est notée différemment par l'enseignant selon son ordre de correction (selon qu'elle se trouve parmi les premières ou les dernières feuilles de la pile, selon qu'elle suit une copie jugée « très bonne » ou « très faible » par le professeur). Cet effet est appelé « effet de contraste ou de séquence » : la copie qui suit une copie brillante peut être désavantagée et vice versa.
- ▶ Les informations complémentaires dont dispose le correcteur sur l'élève : le jugement de l'évaluateur peut notamment être influencé par l'origine sociale, l'apparence physique, l'attitude ou le comportement des étudiants, ainsi que les réussites/échecs passés, sans oublier les divers stéréotypes associés aux groupes ethniques ou encore au sexe. Par exemple, certains correcteurs ont tendance à donner de meilleures notes aux apprenants issus des milieux les plus favorisés ou à ceux qui ont obtenu de bons résultats dans le passé ; les étudiants jugés « plus beaux », bien habillés ou qui ont un accent particulier peuvent être mieux notés ; c'est ce qu'on appelle l'« effet de halo ». Il semble que les filles sont désavantagées plus fréquemment que les garçons au niveau de leurs notes, notamment en raison des phénomènes de « menace du stéréotype » et de la reproduction des inégalités sociales, en termes de genre sexuel. En revanche, dans d'autres circonstances, les filles sont mieux évaluées parce qu'elles ont un « meilleur comportement » en classe, et qu'elles maîtrisent mieux les règles et les exigences du « métier d'élève » – en d'autres termes, elles sont plus proches que leurs homologues masculins du comportement idéalement attendu par l'enseignant.
- ▶ L'état du correcteur : son jugement peut être influencé par des facteurs personnels tels que la fatigue, la distraction, et le conduire à donner, à des moments différents, des notes différentes à une même copie.

Quelles précautions prendre pour contourner ces difficultés ? Cinq approches peuvent être utilisées pour révéler les biais de notation, et donc chercher à les réduire :

- (a) le même ensemble de copies est corrigé plusieurs fois par le même correcteur, à des moments différents, sans que ce dernier en soit conscient, ce qui permet de mesurer la stabilité au sein des correcteurs ;
- (b) le même jeu de copies est corrigé par plusieurs correcteurs différents, ce qui permet de mesurer l'accord entre correcteurs ;
- (c) la même copie est placée dans un ensemble de copies, dans des positions différentes (précédées par des copies meilleures ou plus faibles), ce qui permet de mesurer l'effet de contraste ou effet de séquence ;
- (d) la même copie est placée dans un ensemble de copies dont les valeurs sont plus ou moins dispersées (tantôt parmi des copies ayant toutes reçu la même note lors d'une évaluation précédente, tantôt parmi des copies de qualité très variable) ;
- (e) la même copie est corrigée par plusieurs groupes de correcteurs qui disposent de différentes informations complémentaires sur l'étudiant, ses notes antérieures...

Q9.2 Comment améliorer la validité d'une évaluation?

L'évaluation devrait permettre d'affirmer que la performance observée (la réponse à un QCM ou à une réponse ouverte, les actions que l'élève met en place dans la classe pendant une investigation...) reflète le niveau de connaissance ou de compétence développé par l'élève. Mais la performance est ponctuelle et ne correspond qu'à ce que l'apprenant produit à un moment donné, à un niveau d'engagement donné. Il suffit de penser au théâtre : une seule représentation peut être mauvaise pour de nombreuses raisons liées au contexte ! Ainsi, plusieurs facteurs peuvent réduire la cohérence entre la compétence ou la connaissance que l'on veut évaluer et la performance constatée. Ces facteurs peuvent se présenter à différents moments de l'évaluation :

- ▶ Au moment de la création des items, des questions ou des problèmes constituant l'évaluation, il arrive aux concepteurs :
 - D'ajouter des tâches parasites ne portant pas sur les connaissances ou les compétences que l'on désire évaluer. Elles influencent alors négativement les performances.
 - De proposer des items, des questions et des problèmes qui ne mesurent pas réellement ce que l'on souhaite évaluer (problème de validité). La condition de validité implique en effet celle de pertinence. Par exemple, si l'on veut évaluer la capacité à raisonner sur un phénomène naturel, il ne suffit pas de poser des questions sur le lexique du phénomène en question ou de demander des définitions, car cela ne permet pas d'extraire des informations pertinentes sur le raisonnement.
- ▶ Au moment de l'évaluation. Les conditions dans lesquelles les élèves travaillent peuvent « activer » la menace du stéréotype, créer du stress inutile ou provoquer une charge cognitive délétère.
- ▶ Au moment de l'analyse des réponses. L'évaluateur peut être influencé par différents biais cognitifs. Par exemple, le biais de confirmation est la tendance à interpréter des informations d'une manière qui confirme ou soutient ses propres croyances. Le biais de confirmation peut influencer la manière dont l'évaluateur juge la production d'un sujet sur la base de la première image qu'il s'est faite des productions de l'élève.

Quelles précautions prendre pour contourner ces difficultés ?

L'évaluation demande prudence et modestie. L'évaluation formative doit clairement être favorisée tout au long du processus d'apprentissage et être présentée aux élèves pour ce qu'elle est : une aide à apprendre, où l'erreur est un moteur sans aucun autre enjeu. On évite ainsi le stress dû aux enjeux de sélection pour centrer les élèves sur leurs apprentissages et les retours nécessaires à leur bon déroulement. Cette manière d'évaluer aura notamment un impact positif sur les élèves en situation socio-économique défavorisée et sur les filles, qui font souvent l'objet de stéréotypes sociaux, mais aussi sur la motivation intrinsèque et l'image de soi de tous les élèves. Pour toutes ces raisons, l'évaluation formative améliore la capacité de l'élève à autoréguler ses apprentissages et l'aide donc à gagner en autonomie.

Q9.3 Comment améliorer l'équité d'une évaluation?

Une évaluation équitable doit donner à tout le monde la possibilité d'exprimer tout son potentiel. Or, respecter les mêmes conditions pour tous n'est pas nécessairement ce qu'il y a de mieux pour garantir l'équité.

Par exemple, on peut considérer que donner le même temps de réponse à tout le monde ou ne pas aider ou assister certains élèves plus que d'autres répond à l'exigence d'une évaluation juste. Mais ceci ne prend pas en compte le fait que pour certains élèves, ces conditions sont handicapantes et qu'elles ne permettent donc pas d'exprimer pleinement et librement les connaissances et les capacités que l'on vise à évaluer.

Du moment où le but de l'évaluation n'est pas celui d'un concours de sélection, mais une évaluation dans le cadre d'un parcours d'apprentissage, il vaut mieux donner aux élèves toute l'aide dont ils ont besoin pour les mettre en condition de réussir leurs évaluations et de montrer tout leur potentiel.

Tout comme demander à tous les élèves de passer leurs tests sans lunettes ne serait pas du tout une condition de réussite équitable pour tous. Ainsi, le fait de demander à tout le monde d'écrire ou de prendre le même temps pour répondre ne conduit pas à des évaluations justes !

Dans toutes les tâches d'évaluation, la performance est en outre fortement dépendante de la familiarité des élèves avec le contexte et le contenu/sujet. Il faut en effet rappeler que lors des apprentissages, les compétences sont liées à un contexte et à un contenu particuliers. Le contexte peut faire la différence, pour le meilleur et pour le pire :

- ▶ au niveau de la compréhension ;
- ▶ au niveau de l'engagement à répondre.

Ceci pose notamment un problème lorsque l'on veut évaluer les compétences liées à la démarche scientifique car, inévitablement, les questions des tests, tout comme les observations en classe, portent ou sont effectuées à propos d'objets et d'événements particuliers, pour répondre à des questions sur des phénomènes particuliers. Or, un élève peut être capable de planifier une investigation appropriée (par exemple, planifier une expérience avec contrôle des variables) à une situation pour laquelle il a une certaine connaissance, mais ne pas y parvenir si le sujet ne lui est pas familier.

L'influence du contenu et du contexte soulève des questions d'équité, car tous les apprenants sont inégaux scolairement et socialement devant la variété des contenus et des contextes.

Certains élèves, par exemple ceux issus de milieux défavorisés, peuvent ne pas avoir acquis l'expérience nécessaire à la compréhension d'un élément du test. Des élèves peuvent éprouver des difficultés avec la langue du test, alors qu'il s'agit de leur deuxième ou troisième langue. Ces élèves, ainsi que ceux qui ont des besoins physiques particuliers ou des difficultés d'apprentissage peuvent par conséquent être empêchés de montrer ce qu'ils savent faire.

Quelles précautions prendre pour contourner ces difficultés ? Voici quelques suggestions :

- ▶ faire en sorte que les élèves se sentent impliqués dans l'utilisation des compétences si l'on veut évaluer ce qu'ils savent réellement faire : les tâches doivent être authentiques et attrayantes, mais aussi cohérentes avec ce qu'ils ont appris ;
- ▶ dans un premier temps, faire en sorte que le contexte de la tâche soit familier ou proposer plusieurs contextes différents pour une même tâche, en laissant les élèves libres de choisir celui qui leur convient ;
- ▶ dans un second temps, entraîner les élèves à rencontrer des tâches incluant un certain degré de nouveauté afin qu'ils apprennent et osent utiliser leurs connaissances ou leurs compétences dans le cadre d'un transfert, et favoriser ainsi leur généralisation le cas échéant ;
- ▶ veiller à ce que des tâches pensées pour évaluer les compétences liées à la démarche scientifique ne soient pas trop demandeuses en termes de connaissances et de compréhension et, vice versa, que les tâches pensées pour évaluer la compréhension n'impliquent pas une difficulté trop importante en termes de compétences liées à la démarche ;
- ▶ en raison des limites des tests écrits (QCM, réponses ouvertes) et des tests faisant appel à la pratique, il est utile d'ajouter d'autres types d'évaluations qui se déroulent sur une période plus longue ou utilisent d'autres modalités (observation en classe, évaluation pratique individuelle ou en groupe...).

Q9.4 Quels points d'attention particuliers concernant les évaluations en sciences?

Dans le cas des évaluations portant sur des contenus scientifiques, il faut veiller à ne pas inclure dans les évaluations des compétences parasites pouvant pénaliser les performances de certains élèves. En voici des exemples de différente nature à prendre en considération :

- ▶ La barrière linguistique, liée à une distorsion entre la formulation des questions et les compétences linguistiques de certains élèves. Nous avons déjà vu que, de manière plus subtile, le contexte de la question posée peut influencer l'expression de la compétence. Dans ce cas, ne pas hésiter, par moments, à prendre en considération les besoins spécifiques de ces élèves pour, par exemple, leur lire les consignes d'évaluation et s'assurer de leur bonne compréhension. Il est également possible d'aider à la rédaction des réponses le cas échéant.
- ▶ La menace du stéréotype est une condition pouvant affecter les performances des personnes qui s'identifient à un stéréotype négatif (par exemple, pour une fille, le stéréotype selon lequel « les filles ne sont pas bonnes en sciences » peut affecter ses performances dans cette discipline). Les sciences et les mathématiques font partie des domaines dans lesquels les effets de la menace du stéréotype peuvent être plus facilement mis en évidence, car il s'agit de tâches valorisées et considérées difficiles. La menace du stéréotype peut agir de façon implicite, sans que la personne se dise explicitement « je suis une fille, donc je ne suis pas bonne en sciences ou en maths ! ». Les stéréotypes implicites et explicites sont présents dès l'enfance et affectent les performances et les choix de carrière. Une explication possible de la menace du stéréotype est qu'elle conduit à une charge cognitive élevée, qui nuit aux performances. Mais cette menace a plusieurs facettes et différents mécanismes y sont probablement impliqués en raison du type de stéréotype.
- ▶ Une charge cognitive excessive (de par le contexte ou les contenus de l'évaluation) peut affecter les résultats de l'évaluation. La charge cognitive est la quantité de ressources cognitives investies par un individu lors de l'exécution d'une tâche. La charge cognitive est un concept qui remonte aux années 1960 et qui est utilisé dans diverses disciplines de la psychologie (notamment cognitive, sociale et développementale), la neuropsychologie et dans de nombreux

domaines d'application : la sécurité des transports, l'ergonomie des systèmes d'information ou encore les effets du stress sur les performances sportives, ainsi que, pour ce qui nous concerne ici, l'éducation. La charge cognitive peut affecter la performance : lorsque les ressources cognitives sont insuffisantes, parce que la tâche est trop complexe, parce que l'élève utilise ses ressources pour une autre tâche ou parce que la tâche est présentée de telle manière qu'elle ajoute des tâches invisibles à la tâche principale. Par exemple, la nécessité de contrôler le stress ou les menaces liés aux stéréotypes peut priver l'élève de ressources précieuses pour faire face à la tâche d'évaluation, limitant ainsi la validité de l'évaluation, puisque ce qui est évalué n'est pas seulement la capacité à résoudre les tâches, mais aussi la capacité à résoudre la tâche + la capacité à surmonter le stress et à faire taire la menace liée aux stéréotypes.

Q10. Où trouver des lectures pour justifier les propos de cette page?

Q10.1 Des approfondissements scientifiques

- ▶ Des fiches d'approfondissement sont disponibles en ligne sur le site de la Fondation *La main à la pâte*, à la page [Sciences cognitives et éducation](#).

Pour aller plus loin on pourra également consulter :

- ▶ Cnesco (2023). Conférence de consensus. « L'évaluation en classe, au service de l'apprentissage des élèves ».
- ▶ CSEN (2022). Le passeur : Les enjeux de la métacognition à l'école.
- ▶ MOOC La psychologie pour les enseignants, produit par l'École normale supérieure et le réseau Canopé – « Motivation et métacognition : Les évaluations propices aux motivations d'apprentissage ».

Synthèses de la littérature scientifique

- ▶ Black, P., & William, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- ▶ Bennett, R. E. (2011). Formative Assessment : A Critical Review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25.
- ▶ Briggs, D. C., Ruiz-Primo, M. A., Furtak, E., Shepard, L., & Yin, Y. (2012). Meta-Analytic Methodology and Inferences About the Efficacy of Formative Assessment. *Educational Measurement: Issues and Practice*, 31(4), 13-17.
- ▶ Florin, A., Tricot, A., Chesné, J.-F., Piedfer-Quêney, L., Simonin-Kunerth, M. (2023). Dossier de synthèse : L'évaluation en classe, au service de l'apprentissage des élèves. Cnesco, Cnam.
- ▶ Kingston, N., & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice*, 30(4), 28-37.
- ▶ Hattie, J. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge.
- ▶ Tricot, A., & Chesné, J.-F. (2020). *Numérique et apprentissages scolaires : rapport de synthèse*. Centre national d'étude des systèmes scolaires (Cnesco), Conservatoire national des arts et métiers (Cnam).

Livres

- ▶ Brookhart, S. M. (2013). *How to Create and Use Rubrics for Formative Assessment and Grading*. Alexandria, VA : ASCD.
- ▶ Brookhart, S. M., & Nitko, A. J. (2019). *Educational Assessment of Students*, 8th Edn. Boston, MA : Pearson.
- ▶ Moss, C. M., & Brookhart, S. M. (2019). *Advancing Formative Assessment in Every Classroom: A Guide for Instructional Leaders*. ASCD.
- ▶ Style, A., Andrade, H. L., & Cizek, G. J. *Handbook of Formative Assessment*.
- ▶ Tomlinson, C. A., & McTighe, J. (2006). *Integrating Differentiated Instruction & Understanding by Design: Connecting Content and Kids*. ASCD.

Articles

- ▶ Autin, F., Batruch, A., & Butera, F. (2019). The Function of Selection of Assessment Leads Evaluators to Artificially Create the Social Class Achievement Gap. *Journal of Educational Psychology*, 111(4), 717.
- ▶ Broadfoot, P., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). [Assessment for Learning: 10 Principles](#).
- ▶ Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Welsh, M. E. (2016). Century of Grading Research: Meaning and Value in the Most Common Educational Measure. *Review of Educational Research*, 86(4), 803-848.
- ▶ Cameron, J., & Pierce, W. D. (1994). Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis. *Review of Educational Research*, 64(3), 363-423.
- ▶ Crouzevialle, M., & Darnon, C. (2019). On the Academic Disadvantage of Low Social Class Individuals: Pursuing Performance Goals Fosters the Emergence of the Achievement Gap. *Journal of Educational Psychology*, 111(7), 1261.
- ▶ Hayek, A. S., Toma, C., Oberlé, D., & Butera, F. (2015). Grading Hampers Cooperative Information Sharing in Group Problem Solving. *Social Psychology*, 46(3), 121.
- ▶ Harlen W, Deakin Crick R (2002). [A systematic review of the impact of summative assessment and tests on students' motivation for learning](#) (EPPI-Centre Review, version 1.1*). In: Research Evidence in Education Library. Issue 1. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- ▶ Harlen, W. (2013). Assessment & Inquiry-Based Science Education: Issues in Policy and Practice. Global Network of Science Academies (IAP) Science Education Programme (SEP).
- ▶ Hattie, J. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge.
- ▶ Koenka, A. C., Linnenbrink-Garcia, L., Moshontz, H., Atkinson, K. M., Sanchez, C. E., & Cooper, H. (2021). A Meta-Analysis on the Impact of Grades and Comments on Academic Motivation and Achievement: A case for Written Feedback. *Educational Psychology*, 41(7), 922-947.
- ▶ Lepper, M. R., Henderlong, J., & Gingras, I. (1999). Understanding the Effects of Extrinsic Rewards on Intrinsic Motivation - Uses and Abuses of Meta-Analysis: Comment on Deci, Koestner, and Ryan (1999).
- ▶ McMillan, J. H., Venable, J. C., & Varier, D. (2013). Studies of the Effect of Formative Assessment on Student Achievement: So Much More Is Needed. *Practical Assessment, Research & Evaluation*, 18(2), n2.
- ▶ Nyberg, K., Koerber, S., & Osterhaus, C. (2020). How to Measure Scientific Reasoning in Primary School: A Comparison of Different Test Modalities. *European Journal of Science and Mathematics Education*, 8(3), 137-144.
- ▶ Opitz, A., Heene, M., & Fischer, F. (2017). Measuring Scientific Reasoning - A Review of Test Instruments. *Educational Research and Evaluation*, 23(3-4), 78-101.
- ▶ Pennington, C. R., Heim, D., Levy, A. R., & Larkin, D. T. (2016). Twenty Years of Stereotype Threat Research: A Review of Psychological Mediators. *PLOS One*, 11(1), e0146487.
- ▶ Piekny, J., & Maehler, C. (2013). Scientific Reasoning in Early and Middle Childhood: The Development of Domain-General Evidence Evaluation, Experimentation, and Hypothesis Generation Skills. *British Journal of Developmental Psychology*, 31(2), 153-179.
- ▶ Pulfrey, C., Darnon, C., & Butera, F. (2013). Autonomy and Task Performance: Explaining the Impact of Grades on Intrinsic Motivation. *Journal of Educational Psychology*, 105(1), 39.
- ▶ Pulfrey, C., Buchs, C., & Butera, F. (2011). Why Grades Engender Performance-Avoidance Goals: The Mediating Role of Autonomous Motivation. *Journal of Educational Psychology*, 103(3), 683.
- ▶ Ryan, R. M., & Deci, E. L. (1996). When Paradigms Clash: Comments on Cameron and Pierce's Claim that Rewards Do Not Undermine Intrinsic Motivation. *Review of Educational Research*, 66(1), 33-38.
- ▶ Siarova, H., Sternadel, D., & Mašidlauskaitė, R. (2017). Assessment Practices for 21st Century Learning: Review of Evidence. NESET II report. Luxembourg : Publications Office of the European Union.
- ▶ Sodian, B., Zaitchik, D., & Carey, S. (1991). Young Children's Differentiation of Hypothetical Beliefs from Evidence. *Child Development*, 62, 753-766.
- ▶ Souchal, C., Toczek, M. C., Darnon, C., Smeding, A., Butera, F., & Martinot, D. (2014). Assessing Does Not Mean Threatening: The Purpose of Assessment as a Key Determinant of Girls' and Boys' Performance in a Science Class. *British Journal of Educational Psychology*, 84(1), 125-136.
- ▶ Tomlinson, C. A., & McTighe, J. (2010). *Conversations : Integrating Differentiated Instruction and Understanding by Design*. ASCD.

- ▶ Wiliam, D., & University of London. Institute of Education. (2009). *Assessment for Learning: Why, What and How ?*
- ▶ Willingham, D. L'apprentissage devrait-il être sa propre récompense ? (trad. de l'anglais, article paru dans *American Educator*, 2007-2008)
- ▶ Zimmerman, C. (2007). The Development of Scientific Thinking Skills in Elementary and Middle School. *Developmental Review*, 27(2), 172-223.

Q10.2 Des outils pour s'inspirer dans la création de ses évaluations

Pour des évaluations concernant les **connaissances scientifiques**, on pourra consulter :

- ▶ Évaluations nationales [CEDRE-DEPP sciences](#)
- ▶ Évaluations Timss sciences, qui font explicitement référence à trois types de domaines cognitifs : « connaître », « appliquer » et « raisonner » (seulement une partie des items sont accessibles).
 - [Consulter un exemple d'items de sciences Timss 4 2019 \(CM1\)](#)
 - [Consulter un exemple d'items de sciences Timss 8 2019 \(4^e\)](#)

Il existe également des tests standardisés – développés dans le cadre de recherches en éducation, de recherches sur le développement cognitif de l'enfant ou encore de recherches sur la compréhension que le grand public a de la science – qui s'intéressent spécifiquement à la **mesure des capacités de raisonnement scientifique** (raisonner comme des scientifiques, en termes de méthode) et à la compréhension de la nature de la science. Ils sont souvent en anglais, difficiles d'accès.

À titre d'exemple :

- ▶ [The Science P Reasoning Inventory](#) (école primaire) (EN) ;
- ▶ [The Science K Reasoning Inventory](#) (école maternelle) (EN) ;
- ▶ [Views of the Nature of Science Questionnaire](#) (VNOS) (EN).

L'exemple le plus connu **d'évaluation de la culture ou de la littératie scientifique** est représenté par l'enquête PISA, dirigée par l'OCDE. Pour se faire une idée plus précise de son contenu, on pourra consulter les liens suivants :

- ▶ [PISA 2015 en culture scientifique : exemples d'exercices proposés aux élèves](#) ;
- ▶ [Analyse de quelques questions PISA 2015 en culture scientifique](#) ;
- ▶ [PISA 2006](#) (EN) ;
- ▶ [PISA Take the Test](#) (EN).

Parmi les **tests concernant les attitudes, motivations et valeurs**, nous pouvons citer :

- ▶ [Questionnaire de l'étude ROSE \(Relevance of Science Education\)](#), qui mesure l'intérêt des élèves pour les sciences et les technologies, le désir d'apprendre ou de s'y engager pour une future carrière dans ces disciplines.
- ▶ [Test of Science Related Attitudes \(TOSRA\)](#), qui évalue les attitudes liées à la science selon sept dimensions : implications sociales de la science, normalité des scientifiques, attitude à l'égard de la recherche scientifique, adoption d'attitudes scientifiques, plaisir des cours de sciences, intérêt des loisirs pour la science et intérêt pour les carrières scientifiques.
 - [Modified Attitude Towards Science Inventory \(mATSI\)](#), qui mesure les attitudes des élèves à l'égard des sciences, en fonction de facteurs tels que la perception qu'ont les élèves du professeur de sciences, l'anxiété à l'égard des sciences, la valeur des sciences dans la société...

Pour des exemples de la façon de construire et d'utiliser des grilles d'évaluation (certaines faisant référence à des compétences, d'autres à des standards) avec niveaux de progression et descripteurs de performance :

- ▶ [Cellule d'appui pédagogique, UTC : Construire une grille critériée d'évaluation](#) ;
- ▶ [The Victorian Curriculum and Assessment Authority : Guide to Formative Assessment Rubrics](#) (EN) ;
- ▶ [Australian Education Research Organization – Aero : Formative assessment rubrics](#) (EN).

Pour des exemples de tests portant sur les pratiques (niveau baccalauréat), on pourra consulter :

- ▶ Éduscol : [Banques des évaluations des compétences expérimentales \(ECE\)](#).

Pour des exemples de grilles d'évaluation notée, voir celles de notation des épreuves du bac :

- ▶ [Grille d'évaluation pour l'épreuve de l'enseignement de spécialité « SVT » de la classe de terminale.](#)
- ▶ [ISLE-Based Labs, accessibles via le site Scientific Abilities.](#)

Auteurs

Elena PASQUINELLI, Frédéric PEREZ, Nathalie GIORGI, Yara KANE, Evelyne TOUCHARD

Remerciements

Emilie CAROSIN

Cette ressource a été produite avec le soutien du Ministère de la Culture

Soutenu
par



**MINISTÈRE
DE LA CULTURE**

*Liberté
Égalité
Fraternité*

Date de publication

Mars 2024

Licence

Ce document a été publié par la Fondation La main à la pâte sous la licence Creative Commons suivante : Attribution + Pas d'Utilisation Commerciale + Partage dans les mêmes conditions.



Le titulaire des droits autorise l'exploitation de l'œuvre originale à des fins non commerciales, ainsi que la création d'œuvres dérivées, à condition qu'elles soient distribuées sous une licence identique à celle qui régit l'œuvre originale.

Fondation *La main à la pâte*

43 rue de Rennes

75 006 Paris

01 85 08 71 79

contact@fondation-lamap.org

Site : www.fondation-lamap.org

 FONDATION
La main à la pâte
POUR L'ÉDUCATION À LA SCIENCE