

**Guide de l'enseignant sur l'évaluation  
sommativ en classe des élèves au Maroc**



Produit en collaboration avec :

Royaume du Maroc



Ministère de l'Education Nationale,  
du Préscolaire & des Sports



MILLENNIUM CHALLENGE ACCOUNT MOROCCO  
وكالة حساب تمرد الألفية-المغرب



MILLENNIUM  
CHALLENGE CORPORATION  
UNITED STATES OF AMERICA

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

Ce document, ainsi que les données et cartes qu'il peut comprendre, sont sans préjudice du statut de tout territoire, de la souveraineté s'exerçant sur ce dernier, du tracé des frontières et limites internationales, et du nom de tout territoire, ville ou région.

©OCDE 2023

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

# Remerciements

Le guide de l'enseignant sur l'évaluation sommative en classe a été élaboré par l'Organisation de coopération et de développement économique (OCDE), dans le cadre du déploiement de la sous-composante « Évaluation des acquis des élèves et des performances scolaires » du projet « Éducation secondaire » relevant du programme de coopération « Compact II », conclu entre le Gouvernement du Royaume du Maroc et le Gouvernement des États-Unis, représenté par Millennium Challenge Corporation (MCC), et dont la mise en œuvre a été confiée à l'Agence Millennium Challenge Account-Morocco (MCA-Morocco).

L'équipe de l'OCDE souhaite exprimer sa gratitude à l'ensemble des institutions qui ont participé à l'élaboration de ce guide. Ainsi, nous avons pu bénéficier des précieux conseils et retours du Ministère de l'Éducation Nationale, du Préscolaire et des Sports (ci-après dénommé le Ministère de l'Éducation ou le Ministère), notamment le Centre National de l'Évaluation et des Examens (CNEE) au Maroc, sous la direction de Monsieur Mohamed Sassi et son équipe, notamment messieurs Said Boudergua, Mohammed Lamghari, et Abdelhakim Ech-chaouyqui qui étaient impliqués dans la coordination du projet 3.1 de la convention. Par ailleurs, nous tenons à exprimer notre gratitude aux inspecteurs, enseignants, et à tous les participants pour avoir pris part aux ateliers destinés à informer et à affiner le guide et fourni leurs retours constructifs.

Nous tenons aussi à remercier l'Agence MCA-Morocco et le responsable du projet en son sein, Messieurs Mohamed Mirisse et Rachid El Otmani, qui ont assuré, tout au long du processus du projet, le suivi et la liaison entre les parties prenantes du côté du Maroc et de l'OCDE, l'appui à la mise en œuvre des projets et la revue des livrables.

Le guide a été rédigé par Mesdames Dominique Lafontaine, Françoise Crepin, Isabelle Demonty, et Monsieur Antoine Bodin. Nous tenons à les remercier vivement pour leurs précieuses contributions intellectuelles, ainsi que leur soutien et leurs conseils réguliers tout au long du projet.

Finalement, nous remercions Eleonore Morena, pour son appui précieux en matière de révision du guide, et l'équipe de Lushomo, pour la conception graphique qui rend le guide plus accessible.

L'équipe du projet était dirigée par Elizabeth Fordham, la supervision des projets du programme de coopération était assurée par Soumaya Maghnouj (octobre 2020 – mars 2022) et Lynn Abi Raad (mars 2022 – décembre 2022) ; et le projet 3 était coordonné par Lynn Abi Raad et incluait Till Kadereit, Romane Viennet et Malek Abou-Jawdeh (Secrétariat de l'OCDE).

# Résumé exécutif

Ce résumé exécutif a pour but de présenter les points les plus importants et les principes centraux du guide. Il fournit ainsi un aide-mémoire concis pour passer rapidement en revue les principes clés pour le contrôle continu. Il ne remplace en aucun cas le guide lui-même, qui donne l'ensemble des informations nécessaires à la bonne compréhension et la mise en œuvre de ces principes.

On distingue généralement deux grandes catégories d'évaluation des apprentissages : l'évaluation sommative et l'évaluation formative. La première est destinée à contrôler les acquis des élèves à la fin d'une période d'apprentissage et prend la forme du contrôle continu au Maroc.

Ce guide est consacré à l'évaluation sommative, même si les techniques qui y sont proposées peuvent être utilisées dans le cadre de l'évaluation formative. Il vise à fournir des pistes et des techniques pour améliorer les pratiques d'évaluation en cours au Maroc. Plus précisément, l'objectif principal de ce guide méthodologique est de proposer des suggestions claires pour renforcer la fiabilité (qui comprend la validité et la fidélité) et l'équité des contrôles continus.

Les grandes étapes à suivre pour garantir un contrôle continu fiable et équitable sont ainsi présentées :



## L'élaboration des épreuves

- *Définir l'objet de l'évaluation* : bien définir ce que l'on cherche à évaluer est le premier pas vers une épreuve de qualité. Si l'objet n'a pas été défini au préalable, il est important qu'un cadre de référence vienne guider les concepteurs des épreuves avant même l'élaboration des questions, afin de préciser les connaissances ou compétences à évaluer.
  - La *table de spécification* permet d'opérationnaliser le cadre de référence choisi en croisant les connaissances, les compétences ou capacités évaluées, avec le nombre et les formats de questions. Celle-ci guide la conception des questions et l'analyse des résultats.
- *Choisir les supports et élaborer les questions* : pour garantir un diagnostic valide, il est incontournable que chacune des compétences ou capacités sélectionnées soit évaluée par plusieurs questions, de différents formats et de différents niveaux de complexité. Les questions peuvent être:
  - Fermées à réponse choisie (alternative vrai/faux ou questions à choix multiples [QCM]).
  - Ouvertes à réponse rédigée (réponse brève ou réponse construite).

Chaque format de question présente des avantages et des inconvénients. Les questions fermées permettent une correction rapide et fidèle, alors que les questions ouvertes fournissent davantage d'informations sur les démarches mises en œuvre par l'élève.

- En général, en rédigeant des questions, il faut s'assurer qu'elles évaluent bien la connaissance ou compétence que l'on souhaite évaluer, qu'elles sont formulées de manière claire et précise, et qu'elles sont indépendantes les unes des autres.
- Les questions fermées étant plus complexes à élaborer, le guide fournit des règles générales pour développer des QCM de qualité, notamment : réduire la part du hasard, concevoir des distracteurs pertinents, choisir des contextes authentiques, choisir soigneusement les supports, formuler des questions et des réponses claires et précises, et assurer le caractère exclusif de la réponse correcte, entre autres.



### La passation des épreuves

- La passation est la période pendant laquelle les élèves sont soumis à l'épreuve. Elle devrait être normalisée autant que possible afin de renforcer l'égalité entre les élèves.
- L'organisation de la passation se fait sous la responsabilité du chef d'établissement et comprend les aspects suivants :
  - a. Mettre en place les conditions logistiques pour la passation.
  - b. Informer les élèves de la date suffisamment à l'avance (par exemple, au moins une semaine).
  - c. Assurer une supervision adéquate pendant l'épreuve.
  - d. Définir des mesures à prendre en cas de retard, d'absence ou d'achèvement anticipé.
  - e. Mettre en place des dispositifs pour les élèves à besoins éducatifs particuliers.
- Il est recommandé d'élaborer une feuille de route type pour les personnes surveillant l'épreuve, qui précise notamment le moment et la durée de la passation, les ressources et les pauses autorisées pendant l'épreuve, entre autres aspects.



### Correction des épreuves

- Il est important de prévoir au départ des grilles de correction précises afin de limiter l'interprétation et la subjectivité des corrections, particulièrement pour les questions ouvertes à réponse rédigée.
- Afin d'informer les élèves sur leurs acquis et leur marge de progrès, il est recommandé de « faire parler la note » en annotant les copies.
- Plusieurs phénomènes mis en évidence par les recherches en docimologie<sup>1</sup> menacent l'objectivité et la validité de l'évaluation pendant la phase de correction. L'enseignant peut, par exemple, réduire un éventuel biais par :
  - a. L'élaboration de critères de correction.
  - b. Des corrections collectives (dans ce cas, il est préférable qu'un enseignant ne corrige pas les copies de ses propres élèves).
  - c. Des corrections par question ou bloc de questions.

<sup>1</sup> La docimologie désigne l'étude scientifique des examens et des procédés de notation en général



## Communication et utilisation des résultats des épreuves

- Dès que les épreuves seront corrigées, les résultats seront communiqués aux différents acteurs, y compris l'enseignant lui-même et ses élèves, les parents, l'équipe éducative et l'administration.
- Une communication constructive et adaptée à chaque acteur est fortement conseillée, mettant l'accent sur la contextualisation du résultat et l'interprétation de ses conséquences pour l'apprentissage.
- Ce guide propose un modèle de document de suivi pour faciliter la communication. Entièrement ou partiellement informatisé et adapté au contexte, celui-ci pourrait inclure :
  - a. La liste des connaissances et des compétences contrôlées (rapportée aux attentes curriculaires).
  - b. L'épreuve elle-même et la liste des items prévus pour la prise d'information sur la performance (rapportée aux attentes).
  - c. Le barème de notation (points attribués à chaque item<sup>2</sup>).
  - d. Le détail des notes et appréciations de chacun des élèves ayant passé l'épreuve.
- 1. La docimologie désigne l'étude scientifique des examens et des procédés de notation en général
- 2. Un item est un élément de recueil d'information dans un test. Une question est une unité d'évaluation qui peut comporter plusieurs items.



## Biais dans l'évaluation

Pendant le processus d'évaluation, il appartient à l'enseignant d'être conscient de l'existence de potentiels biais et de chercher à les éviter. Ces biais peuvent être :

- Liés aux caractéristiques des élèves, notamment au genre, mais aussi à l'origine socioculturelle ou ethnique ou à la langue parlée par les élèves. Ces biais peuvent intervenir pendant la conception de l'épreuve et la correction des copies. Pour les éviter, il est notamment recommandé d'éviter les thèmes fortement genrés lors de la conception d'épreuves.
- Liés aux caractéristiques de la classe (l'effet *Posthumus*) : des recherches ont montré que les enseignants adaptent leurs exigences en fonction du niveau moyen des élèves de leur classe. Cela peut, au moment d'une évaluation, conduire à un traitement inéquitable des élèves selon la classe qu'ils fréquentent puisqu'à compétence égale, un élève pourrait réussir dans une classe faible ou moyenne et échouer dans une classe forte.

<sup>2</sup> Un item est un élément de recueil d'information dans un test. Une question est une unité d'évaluation qui peut comporter plusieurs items.



2.3.1	Évaluations en langue française .....	53
2.3.2	Évaluations en mathématiques .....	58
2.4	Notes .....	64
3	La passation des épreuves .....	65
3.1	Développer les conditions de passation d'une évaluation sommative .....	66
3.2	Développer une feuille de route type .....	69
3.3	Notes .....	73
4	La correction des épreuves .....	74
4.1	Assurer une correction fidèle : grilles de correction et de codage .....	75
4.2	Corriger de manière informative : l'annotation des copies .....	79
4.3	Assurer une notation cohérente .....	81
4.4	Principaux phénomènes menaçant l'objectivité et la validité de l'évaluation .....	82
4.5	Notes .....	84
5	La communication et l'utilisation des résultats .....	85
5.1	Communication des résultats aux différents acteurs .....	87
5.1.1	Communication aux élèves et avec les élèves .....	87
5.1.2	Communication avec l'équipe éducative .....	89
5.1.3	Communication aux parents et avec les parents .....	90
5.1.4	Communication avec l'administration .....	90
5.2	Communication des résultats à travers le document de suivi du contrôle continu .....	91
5.3	Notes .....	92
6	Les biais liés au genre et à d'autres caractéristiques des élèves : une menace pour l'équité .....	93
6.1	Les biais liés aux caractéristiques des élèves .....	94
6.2	Biais liés aux caractéristiques de la classe : l'effet Posthumus .....	97
6.3	Notes .....	99
	Annex A. Glossaire .....	100
	Annex B. Éduométrie : pour aller plus loin .....	113
	Références .....	119

# Graphiques

Graphique 1.1 Composants de qualité d'une évaluation des élèves .....	14
Graphique 1.2 Les grandes étapes du processus du contrôle continu .....	15
Graphique 1.3 Types d'évaluation des élèves .....	17
Graphique 1.4 Les éléments d'une évaluation .....	20
Graphique 2.1 Définition de l'objet de l'évaluation .....	23
Graphique 2.2 Formats possibles des questions .....	27
Graphique 2.3 Catégories principales de la taxonomie de Bloom .....	29
Graphique 3.1 Les différents aspects de l'organisation de la passation .....	68
Graphique 3.2 Éléments d'une feuille de route type .....	70
Graphique 4.1 Principaux phénomènes menaçant l'objectivité et la validité de l'évaluation .....	83
Graphique 4.2 Solutions pour limiter les biais dans la correction .....	84
Graphique 5.1 Communication des résultats du CC aux différents acteurs concernés .....	86
Graphique 5.2 Les principales étapes de la communication avec les élèves .....	88
Graphique 5.3 La communication entre les enseignants sur les résultats du contrôle continu .....	89
Graphique 6.1 Les niveaux des biais liés aux caractéristiques des élèves .....	95
Graphique 6.2 Le mythe de la courbe de Gauss .....	98
Graphique B-1 Exemple d'une courbe de réponse de l'IRT .....	118

# Tableaux

Tableau 2.1 Exemple de table de spécification (4e secondaire, mathématiques, 2017) .....	26
Tableau 2.2 Avantages et inconvénients des différents formats de questions .....	30

# Encadrés

Encadré 1.1 Sciences de l'éducation, sciences cognitives, didactique .....	13
Encadré 1.2 Vocabulaire et concepts .....	16
Encadré 1.3 Évaluation normative et évaluation critériée .....	18
Encadré 1.4 Attention à la complexité de l'évaluation .....	19
Encadré 2.1 Bonne pratique — liste de contrôle .....	23
Encadré 2.2 La taxonomie de Bloom .....	29
Encadré 4.1 Bonne pratique – le prétest .....	76
Encadré 6.1 L'équité : une question de justice .....	94
Encadré 6.2 La menace du stéréotype .....	96

# Introduction

**1.1** Objectif et structure du guide

**1.2** Usages et fonctions de l'évaluation

**1.2.1** *L'évaluation sommative (contrôle continu)*

**1.2.2** *L'évaluation formative*

**1.2.3** *Les épreuves et les évaluations externes*

**1.3** Notes

2

**Les grandes étapes de l'élaboration des épreuves**

3

**La passation des épreuves**

4

**La correction des épreuves**

5

**La communication et l'utilisation des résultats**

6

**Les biais liés au genre et à d'autres caractéristiques des élèves ou de la classe**

L'évaluation des acquis des élèves a toujours été une préoccupation centrale des enseignants. En effet, comment leur serait-il possible d'enseigner sans garder à l'esprit les objectifs de leur enseignement et sans se soucier des effets de cet enseignement ?

De ce fait, l'évaluation, qui consiste avant tout à accorder autant que possible les objectifs de l'enseignement et les effets obtenus, occupe une place centrale dans l'action pédagogique.

Alimentées depuis plus d'un siècle par de nombreuses recherches menées dans tous les pays ainsi que par l'expérience partagée des enseignants, les conceptions de l'apprentissage et de l'évaluation ont considérablement évolué. Le cerveau de l'élève n'est plus considéré comme une boîte plus ou moins vide qu'il s'agirait de remplir ; les sciences cognitives, les neurosciences et les recherches en didactique des disciplines apportent chaque jour de nouvelles connaissances\* qui obligent à repenser les processus éducatifs et évaluatifs (encadré 1.1).



### Encadré 1.1

#### Sciences de l'éducation, sciences cognitives, didactique

Les recherches menées depuis au moins un siècle par les sciences de l'éducation, les sciences cognitives et la didactique des disciplines, ainsi que les besoins nouveaux des sociétés, modifient de façon continue les conceptions sur l'enseignement et les objectifs qui lui sont assignés.

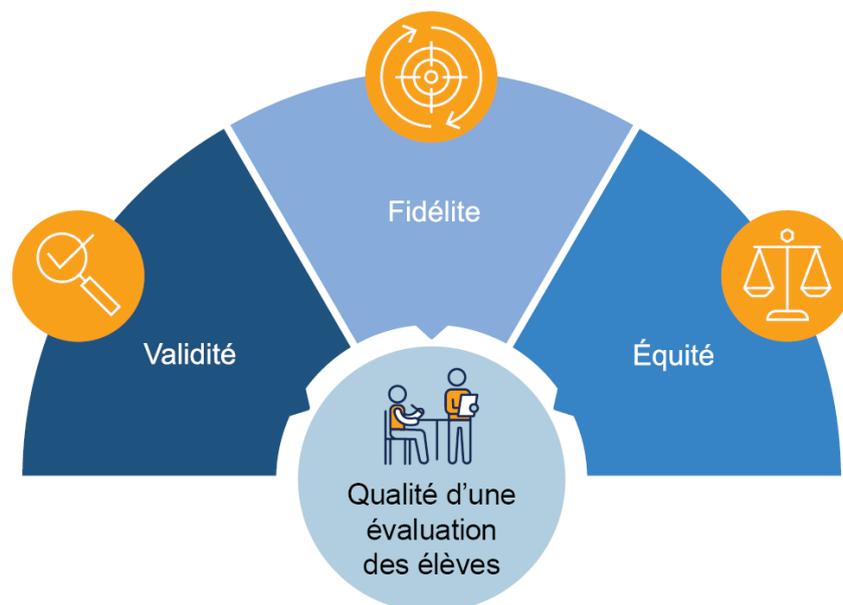
La didactique a mis en évidence un certain nombre de phénomènes qu'il n'est plus possible d'ignorer ; elle a notamment mis en lumière le rôle des erreurs dans la construction du savoir. À un certain moment de l'apprentissage, l'erreur est normale, doit être reconnue, analysée et, à terme, remplacée par un savoir plus conforme à la situation.

Les sciences cognitives, renforcées aujourd'hui par les neurosciences, valident ce que, bien souvent, les enseignants savent déjà sur la base de leur expérience, par exemple que chaque apprentissage demande du temps et nécessite des périodes de renforcement ; elles montrent aussi que le stress et l'anxiété nuisent aux apprentissages. Elles soulignent encore que les retours positifs sont indispensables aux apprentissages. Certaines études montrent qu'un retour négatif devrait être compensé par cinq retours positifs. De ce fait, la bienveillance et l'empathie sont des attitudes souhaitables de la part des enseignants, ce qui n'exclut ni la rigueur ni l'objectivité nécessaires à l'évaluation.

Source : Développé par les auteurs du guide

En outre, pour des raisons aussi bien économiques que sociales, les sociétés modernes portent un intérêt grandissant à la qualité de l'enseignement et à ses effets. Aujourd'hui, au-delà du rapport personnel entre l'enseignant et ses élèves, l'évaluation intéresse toute une chaîne d'acteurs, de l'élève aux plus hautes sphères de la société.

Graphique 1.1 Composants de qualité d'une évaluation des élèves



Source : Développé par les auteurs du guide

Généralement, les responsables et les acteurs des systèmes éducatifs s'attendent à ce que les résultats des évaluations menées en classe soient fiables. Cette fiabilité, recouvre deux notions : la **validité** et la **fidélité** (Rémond, 1980<sup>[11]</sup>), (Hivon, 1989<sup>[21]</sup>) (Voir Annexe A pour une définition des concepts).

De plus, il est indispensable que les évaluations soient **équitables**. Au Maroc, cela constitue une priorité du Ministère de l'Éducation nationale, du Préscolaire et des Sports<sup>1</sup>, guidé par la loi-cadre qui exige l'efficacité, la crédibilité et la fiabilité des évaluations diagnostiques, formatives et certificatives (Ministère de l'Éducation nationale du Maroc, 2020<sup>[31]</sup>), en cohérence avec les objectifs généraux définis dans la Vision stratégique de la réforme 2015-2030.

En conséquence, la mission confiée aux enseignants est devenue plus exigeante et plus complexe que jamais. Elle est aussi devenue plus riche et plus valorisante.

Au Maroc, comme dans la plupart des pays, l'évaluation des élèves est cadrée par un ensemble de normes et d'instructions officielles, notamment par les programmes d'études officiels et les notes ministérielles. Cependant, dans le respect du cadre, les enseignants disposent d'une marge d'autonomie importante : ils peuvent en particulier interroger leurs pratiques et innover.

Sans disqualifier les pratiques en cours, ce guide cherche à fournir des pistes et des techniques pour améliorer ces pratiques d'évaluation, tout en proposant une ouverture sur l'innovation.

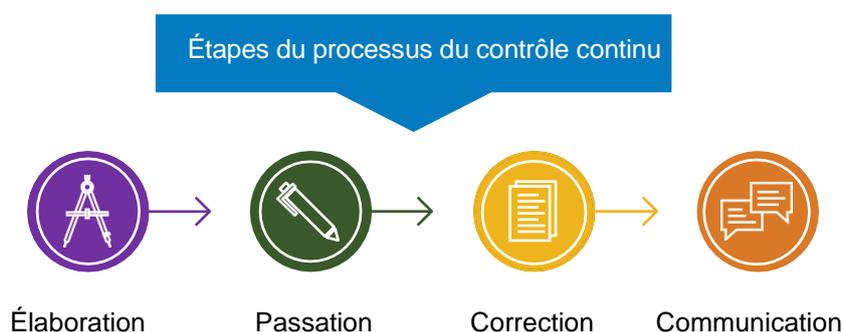
« Sans disqualifier les pratiques en cours, ce guide cherche à fournir des pistes et des techniques pour améliorer ces pratiques d'évaluation, tout en proposant une ouverture sur l'innovation. »

## 1.1 Objectif et structure du guide

L'objectif de ce guide méthodologique est de proposer des suggestions claires pour renforcer la fiabilité\*<sup>2</sup> (qui comprend la validité\* et la fidélité\*) et l'équité des contrôles continus, tout en tenant compte des spécificités et des contraintes qui pèsent sur ceux-ci dans le contexte éducatif du Maroc.

Dans cette perspective, les différentes étapes qui structurent l'évaluation seront successivement abordées et développées de l'élaboration, à la passation, à la correction, et à la communication. La conception de l'épreuve consiste en une réflexion préalable à l'élaboration proprement dite. Dans la phase de conception, il ne s'agit pas encore de rédiger des questions, mais seulement de bien préparer cette élaboration en repérant précisément ce qu'il s'agit d'évaluer, en faisant le choix des types de questions et d'épreuves les mieux appropriées au projet d'évaluation, et en construisant une table de spécification\*<sup>3</sup> qui servira de guide à la construction de l'épreuve (Voir le chapitre 2, La définition de l'objet de l'évaluation : que veut-on évaluer ?). La figure suivante présente les grandes étapes de l'évaluation.

**Graphique 1.2 Les grandes étapes du processus du contrôle continu**



Source : Développé par les auteurs du guide

Ce premier chapitre du guide consiste en une mise en contexte du contrôle continu : le but de ces évaluations et leurs objets. Ce chapitre rappelle, de façon générale, les fonctions et les usages de l'évaluation.

Les autres chapitres du guide développent les étapes du processus du contrôle continu. Il s'agit de son élaboration (chapitre 2, qui revient sur la conception), de sa passation (chapitre 3), de sa correction (chapitre 4) et, finalement, de la communication et de l'utilisation des résultats (chapitre 5).

L'évaluation est sujette à de nombreux biais que l'enseignant doit connaître et chercher à éviter. Ces biais menaçant l'objectivité et la validité de la correction d'une évaluation sont présentés dans le chapitre 4. Il semble cependant important de souligner d'autres biais ; à savoir le biais de genre (concernant directement la moitié des élèves) et le biais d'ajustement des résultats d'une classe à une conception, en général inconsciente, de ce que doit être la distribution des notes (effet Posthumus). Ce dernier est généralisé dans la plupart des systèmes éducatifs et souvent ignoré.

Ces deux biais et les précautions qu'il convient de prendre pour les éviter sont développés dans le chapitre 6.

Ce guide est destiné à l'ensemble des enseignants du collège. Il développe donc des principes généraux. Les exemples d'application ont été tirés des domaines du français, des mathématiques ainsi que, d'une façon plus restreinte, de la biologie pour illustrer de manière plus pratique les principes du guide. Toutefois, il est important de souligner que le contenu de ce guide reste transférable aux différentes disciplines et au différents niveaux scolaires avec des adaptations nécessaires.



### Encadré 1.2 Vocabulaire et concepts

Le vocabulaire des sciences de l'éducation est varié et évolue constamment. Il dépend notamment des courants pédagogiques et de leurs évolutions. Ainsi, certains termes sont des synonymes, d'autres termes peuvent avoir un sens différent selon les auteurs ou les acteurs qui l'emploient.

Une définition précise des termes utilisés est nécessaire afin de permettre une communication et compréhension mutuelles. Pour cette raison, les termes importants à la compréhension d'un chapitre sont définis dans le guide. Par ailleurs, lors de sa première occurrence dans le guide, un terme nécessitant une définition sera muni d'un astérisque renvoyant à un glossaire en annexe (\*).

En ce qui concerne ce guide, on peut penser aux termes suivants (dans l'ordre alphabétique) : Aptitude(s), Attitude(s), Capacité(s), Compétence(s), Connaissance(s), Faculté(s), Habilité(s), Norme(s), Objectif(s), Processus, Savoir-Faire, Savoir(s). Mais la liste ne s'arrête pas là.

Nous avons pris soin de limiter la liste des termes techniques utilisés dans ce guide au strict nécessaire. Notons que l'important réside moins dans les mots que dans les concepts et les conceptions qu'ils cherchent à traduire, et l'on sait que la correspondance mot-concept est rarement univoque.

Par ailleurs, ce guide, dans un souci de lisibilité, emploie le masculin générique pour renvoyer à des ensembles mixtes composés d'hommes et de femmes.

Source : Développé par les auteurs du guide

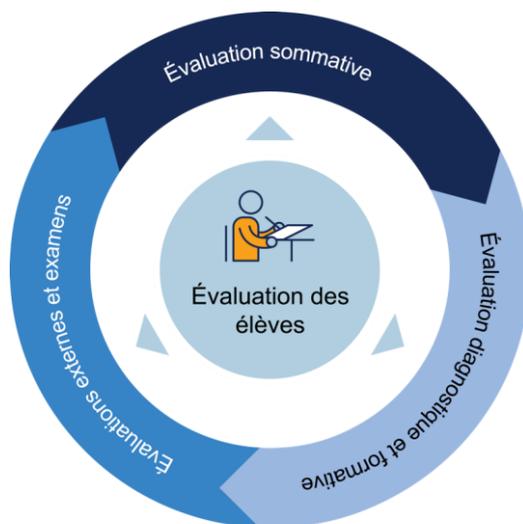
## 1.2 Usages et fonctions de l'évaluation

La tâche essentielle de l'enseignant consiste à prendre en charge les élèves en général et chaque élève en particulier avec ses points forts et ses points faibles, et sa motivation, qu'il lui faut souvent maintenir. De ce point de vue, sa tâche est déjà complexe et l'évaluation peut apparaître comme une charge supplémentaire lourde à porter.

Pourtant, malgré un enseignement soutenu, l'élève n'apprend pas toujours avec la régularité souhaitée, d'où les raisons de s'intéresser à la façon dont il apprend, à la qualité de ses

apprentissages et à la réalité de ses acquis. Ces raisons conduisent à plusieurs types d'évaluation qui n'ont pas toutes les mêmes fonctions ni les mêmes usages. On distingue habituellement deux grandes catégories d'évaluation : l'évaluation sommative\* et l'évaluation formative\*.

**Graphique 1.3 Types d'évaluation des élèves**



Source : Développé par les auteurs du guide

### 1.2.1 L'évaluation sommative (contrôle continu)

L'évaluation sommative est destinée à contrôler les acquis des élèves **à la fin d'une période d'apprentissage**. Au Maroc, l'évaluation sommative dans la classe prend la forme du contrôle continu.

Le contrôle continu est strictement cadré par les instructions officielles et conditionne largement la progression d'une année à l'autre au collège. En outre, il compte pour 30%<sup>4</sup> dans l'examen de fin de cycle du collège.

Ce guide se focalise sur le contrôle continu, dont il s'agit d'améliorer la fiabilité et l'équité. Toutefois, les réflexions et techniques proposées dans les pages qui suivent peuvent aussi être utilisées avec profit dans le cadre de l'évaluation formative.

De plus, le contrôle continu peut et se doit même d'avoir des retombées formatives. Il fournit en effet à l'enseignant des informations qui l'aideront à réguler son action, et aux élèves des informations sur leurs forces, leurs progrès et les points à améliorer.

### 1.2.2 L'évaluation formative

L'évaluation formative est l'évaluation **pratiquée au quotidien par l'enseignant pour accompagner son enseignement**. Son objectif n'est pas de contrôler les acquis mais d'accompagner l'élève dans son apprentissage en fournissant des indications (à l'élève lui-même ainsi qu'à l'enseignant) pour l'aider à surmonter les obstacles rencontrés.

Ce type d'évaluation fait intégralement partie du processus d'enseignement. En observant ses élèves en cours d'apprentissage – lorsqu'ils font des exercices, interviennent, posent des questions, vont au tableau, réussissent une tâche, se trompent, entre autres – l'enseignant évalue. Ainsi, les exercices d'application du cours, les devoirs à la maison, les projets individuels ou collectifs participent de l'évaluation formative.

Selon le besoin ou la nécessité, cette évaluation peut prendre différentes formes, dont les traditionnelles épreuves « papier-crayon », mais aussi d'autres modalités d'évaluation dite « informelles » ou « non instrumentées » comme l'observation, le questionnement et le dialogue entre élèves et avec les élèves.

L'évaluation diagnostique\* est une forme d'évaluation formative spécifiquement destinée à identifier les difficultés des élèves avant une séquence d'apprentissage.

Dans la plupart des systèmes éducatifs, l'évaluation formative a pris une place prépondérante. Elle contribue d'autant plus aux apprentissages que l'élève a intégré l'idée que cette évaluation sert à ses apprentissages et non à le sanctionner ou à le noter.

À l'exception de cette section, qui vise à définir et à présenter le concept, nous n'approfondirons pas davantage ce type d'évaluation, un guide distinct lui étant consacré.



### Encadré 1.3

#### Évaluation normative et évaluation critériée

Une évaluation peut être normative ou critériée. Elle est dite « critériée » lorsqu'elle s'appuie sur des critères précis, définis en cohérence avec les objectifs d'apprentissage. Les productions de chaque élève sont donc évaluées en fonction de ces critères, sans comparaison avec les productions des autres élèves. Cette forme d'évaluation permet aux enseignants de recueillir des informations sur l'apprentissage des élèves, et de fournir un feedback constructif aux élèves pour qu'ils puissent s'améliorer.

Une évaluation est qualifiée de « normative » quand elle cherche à établir un classement entre les élèves en comparant leurs productions respectives (on mettra par exemple une note de 9/10 à la meilleure production, une note de 8 à une production un peu moins bonne... et une note de 3/10 à la plus mauvaise production).

D'une manière générale, il est recommandé d'évaluer les apprentissages des élèves avec une visée critériée et non normative, le diagnostic porté sur les apprentissages de l'élève étant posé en référence à des critères précis<sup>5</sup>.

Source : Développé par les auteurs du guide

## 1.2.3 Les épreuves et les évaluations externes

Les examens\* sont aussi un mode d'évaluation sommative. Il n'en sera question dans ce guide que dans la mesure où le contrôle continu intervient pour 30 % dans l'examen de fin du cycle collégial.

En plus des examens, qui, en grande partie, dépendent d'instances extérieures à l'établissement, les enseignants sont confrontés aux résultats d'évaluations externes nationales ou internationales telles que le Programme international pour le suivi des acquis des élèves [PISA], l'étude Progress in International in Reading Literacy Study [PIRLS] ou l'étude Trends in International Mathematics and Science Study (TIMSS).

Ces dernières évaluations visent à évaluer le système et n'ont qu'un impact indirect sur les enseignants et sur les élèves. Elles interviennent dans la régulation du système éducatif et, donc, dans la régulation des pratiques enseignantes. Il est souhaitable que les enseignants s'y intéressent, mais ce n'est pas l'objet de ce guide.

Un système d'évaluation des élèves bien aligné implique que les différents types d'évaluation (y compris l'évaluation en classe, les examens, et l'évaluation à grande échelle) s'y déroulent de façon coordonnée et portent sur les mêmes objectifs globaux définis par le curriculum, les programmes d'études et les diverses instructions officielles. Les enseignants ont un rôle à jouer dans cet alignement en s'assurant de bien prendre en compte les objectifs officiels dans leur enseignement comme dans le contrôle continu.



#### Encadré 1.4

##### Attention à la complexité de l'évaluation

Avant d'aborder chacune des étapes du processus du contrôle continu, il convient de rappeler que, même en réalisant les différentes étapes du processus du contrôle continu de la manière la plus efficace possible, l'image des acquis et des faiblesses des élèves, renvoyée au travers d'une épreuve, est nécessairement incomplète.

L'épreuve devrait prendre en compte l'ensemble des objectifs (connaissances, savoir-faire\*, compétences\*) définis par les instructions officielles. Cela n'est pas toujours possible.

L'épreuve ne permet de prendre en compte qu'une partie du savoir qu'a l'élève par rapport au domaine qu'il s'agit d'évaluer. Là encore, il s'agit de construire l'épreuve pour que l'élève puisse au mieux y investir son savoir.

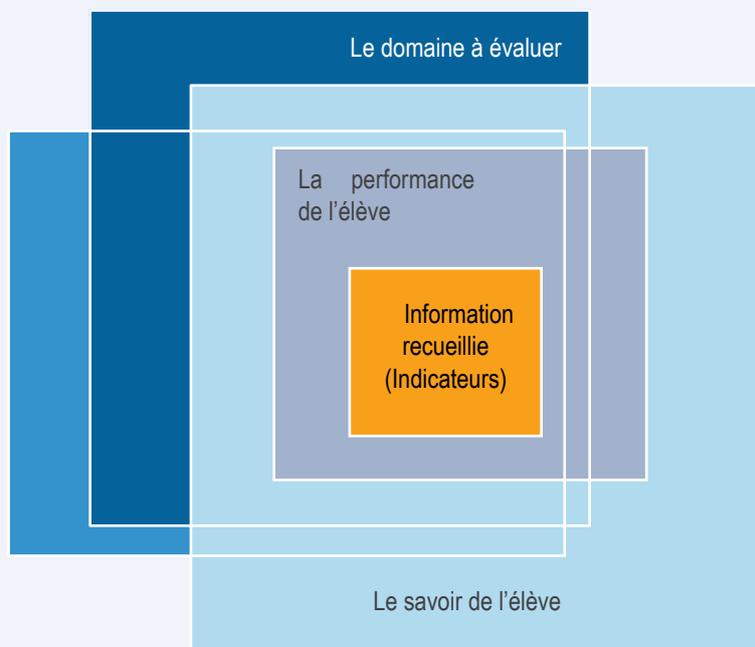
Elle peut aussi déborder en partie de ce qu'il s'agit vraiment d'évaluer. C'est par exemple ce qui se passe lorsque l'on prend en compte le soin et l'orthographe dans un contrôle de mathématiques. Certains débordements sont légitimes mais il convient de les limiter.

La production de l'élève constitue sa performance. En général, cette performance ne couvre qu'une partie des éléments contrôlés par l'épreuve et ne rend compte que d'une partie du savoir que l'élève aurait pu mobiliser.

L'information que l'enseignant va recueillir pour établir son jugement, et pour attribuer une note à l'élève, va être une interprétation de sa performance et ne pourra que difficilement en saisir tous les éléments. Les éléments saisis sont les indicateurs du savoir de l'élève par rapport à l'objet de l'évaluation.

Le graphique 1.4 illustre les éléments qui viennent d'être décrits et qui ne sont jamais tout à fait superposables.

**Graphique 1.4 Les éléments d'une évaluation**



Source : Développé par les auteurs du guide

Tout cela est normal et ne fait qu'illustrer le fait que le savoir n'est pas mesurable au sens de la physique. **Le problème de l'enseignant (ou de l'évaluateur) est donc d'optimiser le processus de passage du domaine à évaluer à l'information recueillie.** Ce guide cherche à faciliter cette optimisation.

Source : Développé par les auteurs du guide

## 1.3 Notes

---

<sup>1</sup> Jusqu'à la fin du guide, nous nous y référerons comme au « Ministère de l'éducation »

<sup>2</sup> Lors de sa première occurrence dans le guide, un mot nécessitant une définition sera muni d'un astérisque renvoyant à un glossaire en annexe (\*).

<sup>3</sup> Voir glossaire dans l'annexe sous le terme « Taxonomie »

<sup>4</sup> Au moment de la rédaction du guide.

<sup>5</sup> Voir « Guide de l'Enseignant sur l'Évaluation Formative et la Différenciation » (OCDE, non publié<sup>[18]</sup>).

1

Section

2

Introduction

# Les grandes étapes de l'élaboration des épreuves

**2.1** La définition de l'objet de l'évaluation : que veut-on évaluer ?

**2.1.1** *La table de spécification*

**2.2** Le choix des supports (textes) et la conception proprement dite des questions

**2.2.1** *Les formats des questions et leurs avantages et inconvénients*

**2.2.2** *L'importance de varier les questions, leurs formats et leur complexité*

**2.2.3** *Quelques règles à respecter*

**2.2.4** *Focus sur la construction de QCM*

**2.2.5** *Focus sur la construction des questions ouvertes*

**2.3** Illustration par des exemples concrets

**2.3.1** *Évaluations en langue française*

**2.3.2** *Évaluations en mathématique*

**2.4** Notes

3

La passation des épreuves

4

La correction des épreuves

5

La communication et l'utilisation des résultats

6

Les biais liés au genre et à d'autres caractéristiques des élèves ou de la classe

## 2.1 La définition de l'objet de l'évaluation : que veut-on évaluer ?

En ce qui concerne la précision du domaine à évaluer, deux cas de figure peuvent se présenter :

1. **Soit l'objet de l'évaluation n'a pas été défini au préalable** : Dans ce cas, il revient aux concepteurs des épreuves (les enseignants en ce qui concerne les évaluations en classe) de se donner un cadre de référence, c'est-à-dire décider et préciser, avant toute élaboration de questions, quelles connaissances, capacités ou compétences vont être évaluées.
2. **Soit cet objet est préalablement fixé**, de manière plus ou moins précise, par les curricula ou les programmes ou, comme dans le cas des contrôles continus, par les notes ministérielles qui décrivent la fréquence et le contenu couvert par chaque épreuve.

Dans les deux cas, il est possible et souvent nécessaire de s'appuyer sur des taxonomies\* et typologies existantes, en fonction des disciplines.



### Encadré 2.1

#### Bonne pratique — liste de contrôle

Même dans le second cas de figure, une liste de contrôle peut être établie. Il est important qu'une telle liste détaille le domaine ciblé, en précisant les connaissances, les savoir-faire et les compétences visées. Bien définir ce que l'on cherche à évaluer est un premier pas pour une épreuve de qualité. Cette liste sera utilisée en fin de phase d'élaboration pour contrôler que l'épreuve produite remplit bien les critères\* prévus.

Source : Développé par les auteurs du guide

### Graphique 2.1 Définition de l'objet de l'évaluation



Source : Développé par les auteurs du guide

Par **exemple**, pour l'examen normalisé régional du niveau collégial, le cadre de référence<sup>1</sup> précise que pour l'épreuve de **lecture**, trois capacités<sup>2</sup> sont à évaluer : Identifier, Comprendre et Réagir (Centre national de l'évaluation et des examens, Non publié<sup>[4]</sup>). Ces trois capacités générales seront ensuite définies de manière plus précise dans ce même document.



Par exemple, la capacité Réagir « amène l'élève à prendre position, à justifier ou à défendre son point de vue en se référant à des éléments contenus dans le texte ». Tous les mots de ces définitions sont importants et toute déviation par rapport à cette définition fait en sorte que l'évaluation ne sera pas valide, c'est-à-dire qu'elle ne mesurera pas ce qu'elle prétend ou doit mesurer.

Enfin, une déclinaison encore plus précise (on parle à ce propos d'opérationnalisation de la capacité) reprend les principaux processus que peut recouvrir la capacité Réagir : « il s'agit donc essentiellement de juger, d'évaluer, de se situer par rapport au texte, d'exprimer son appréciation du texte », toujours en se référant au texte comme spécifié dans la définition.

Dans ce cadre de référence, il est en outre stipulé que l'épreuve de lecture doit porter sur un texte littéraire. Ceci signifie donc que l'épreuve évalue dans les faits la lecture de textes littéraires et non la lecture de manière plus large. Ce choix ne peut se comprendre que si le texte littéraire faisait l'objet spécifique des unités d'enseignement pendant la période sur laquelle porte le contrôle. Si on veut évaluer un domaine de manière valide, il faut évidemment identifier les connaissances, les savoir-faire et les compétences qui structurent le domaine.

En **mathématiques**, le cadre de référence qui fixe les balises du contrôle continu pour les trois années du cycle secondaire préparatoire (précisé dans une note ministérielle) (Ministère de l'Education nationale du Maroc, 2010<sup>[5]</sup>) définit ces éléments.

Ce cadre s'articule autour d'une série de notions organisées en domaines principaux et sous-domaines (comme le théorème de Thalès), dont chacune est travaillée en classe durant 5 semaines environ.

L'évaluation en mathématiques ne porte pas seulement sur la connaissance stricte de ces thématiques et se décline en trois niveaux : le premier concerne la mobilisation directe et explicite de connaissances (définition, propriété, théorème, technique, règle, algorithme, expression, ...) ; le deuxième cherche à voir si l'élève parvient à utiliser ses connaissances dans des situations habituelles (comme utiliser la technique algébrique dans un calcul) ; et le troisième renvoie à des situations inhabituelles envisagées à l'intérieur ou à l'extérieur des mathématiques (il s'agit par exemple de résoudre un



problème en utilisant la technique algébrique : la question posée à l'élève n'indique pas explicitement la nécessité de recourir à cette technique).

Ces éléments orienteront la construction des questions. En outre, l'analyse des résultats des élèves à la lumière de ces éléments autorisera à établir un premier diagnostic de leurs acquis et faiblesses : les problèmes se posent-ils au niveau de la connaissance de la ressource ciblée, au niveau de sa mobilisation en situations habituelles ou inhabituelles ?

### 2.1.1 La table de spécification

Une table de spécification est un outil qui permet d'opérationnaliser le cadre de référence choisi ou imposé et de servir de guide lors de la conception des questions et lors de l'analyse des résultats. Une table de spécification croise différentes dimensions :



1

Les connaissances



2

Les compétences ou capacités<sup>2</sup> évaluées



3

Les formats de questions

Au Maroc, dans le cadre du contrôle continu, des tables de spécification sont déjà en partie définies par les orientations pédagogiques. L'important serait de compléter ces tables qui, en l'état, ne tiennent pas compte des formats de questions. Si on prend l'exemple du français (lecture) pour l'examen au collège, la table prévoit 6 items\*, des pourcentages et un nombre de points par capacité (Identifier, Comprendre, Réagir) mais ne croise pas avec les types d'items. De plus, on peut craindre que 6 items au total pour évaluer la lecture ne suffisent pas pour assurer la validité de l'épreuve.

Le croisement des différentes dimensions est important, pour éviter, par exemple, d'évaluer systématiquement une compétence ou un domaine de savoir à l'aide d'un seul format de question, d'un seul type de texte en français ou d'un seul type de problème en mathématiques.

« Il est important de prévoir dès le départ des questions de difficultés variées pour chaque dimension, afin de cibler les acquis de tous, y compris des élèves en difficulté. »

Il est important de prévoir dès le départ des questions de difficultés variées pour chaque dimension, afin de cibler les acquis de tous, y compris des élèves en difficulté (Voir exemple de table de spécification ci-dessous).

Si cela n'est pas possible, en raison du faible nombre d'items, il faut, a minima et en planifiant les contrôles continus successifs, prévoir une progression dans la difficulté. Il serait par exemple absurde que les questions et items portant sur la capacité à faire des inférences soient plus

complexes en début d'année ou que le texte support de fin d'année soit plus simple que celui de début d'année.



## Exemple de table de spécification

Tableau 2.1 Exemple de table de spécification (4e secondaire, mathématiques, 2017)

Contenus	Processus		
	 Connaître (20 items)	 Appliquer (29 items)	 Transférer (13 items)
Approche graphique d'une fonction (31 items)	6 items	21 items	4 items
Les fonctions du premier degré (31 items)	14 items	8 items	9 items

**⚠ Il y avait davantage d'items au prétest**

Note : Dans certains curricula, on parle de processus plutôt que de compétences.

Source : Développé par les auteurs du guide basé sur (Fédération Wallonie-Bruxelles, s.d.<sup>[6]</sup>)

Dans l'exemple ci-dessus, chaque processus et chaque contenu est évalué par plusieurs items de format varié et il sera possible, une fois le test soumis aux élèves et corrigé, de calculer des scores pour chaque contenu et chaque processus. Cela nécessite de disposer d'une dizaine d'items pour chaque sous-score que l'on souhaite calculer. L'analyse de ces résultats pourra également conduire à des activités d'approfondissement ciblées sur les besoins d'apprentissage des élèves.

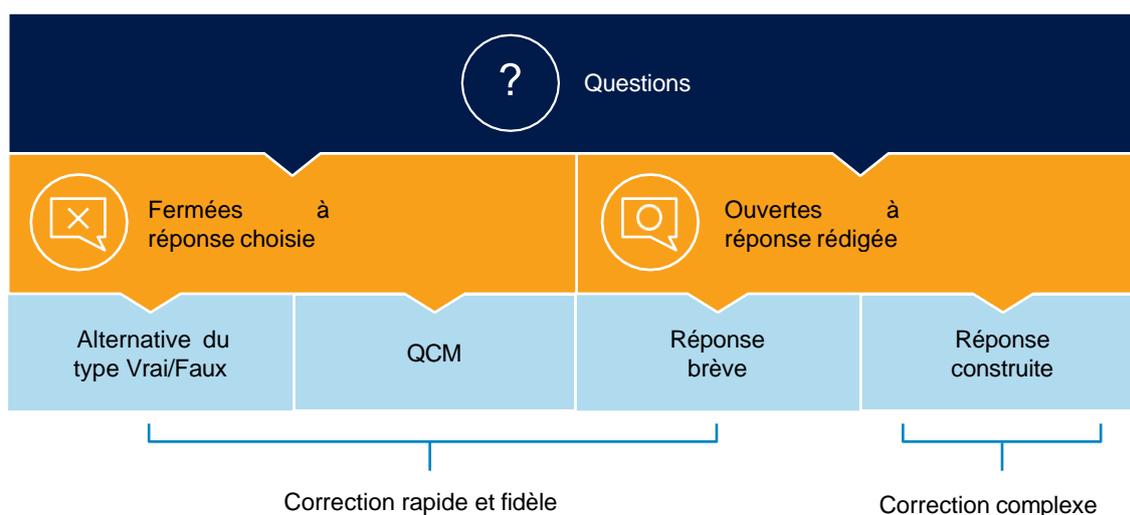
## 2.2 Le choix des supports (textes) et la conception proprement dite des questions

### 2.2.1 Les formats des questions et leurs avantages et inconvénients

#### 2.2.1.1 Les différents formats de questions

Les questions utilisées lors d'une épreuve peuvent généralement prendre les formes illustrées par le schéma suivant, en distinguant les questions ouvertes et fermées.

Graphique 2.2 Formats possibles des questions



Note : QCM : Question à choix multiple.

Source : Développé par les auteurs du guide

L'idée largement répandue selon laquelle une question fermée est simple et une question ouverte est complexe n'est pas exacte. Une question à choix multiples (QCM) bien conçue peut exiger de la part de l'élève des démarches d'un haut niveau de complexité (voir encadré 2. 2). Une telle question implique cependant un travail soigneux de conception.

Voici un exemple de QCM pour les sciences et son tableau de résultats.



## Exemple d'une question fermée complexe

Cette question figure parmi les plus complexes de l'épreuve, elle se situe d'ailleurs au quatrième niveau de complexité des apprentissages cognitifs (sur six) dans la taxonomie\* de Bloom (voir encadré 2.2) : Analyser. Comme le montre le tableau de résultats, elle n'est réussie que par 29 % des élèves. Les distracteurs<sup>3</sup> sont par ailleurs assez attractifs : ils ont tous été choisis par au moins 10 % des élèves.



Voici 5 classements de ces animaux. Un seul est scientifiquement correct. Lequel ?  
Fais une croix à côté du seul classement qui est scientifiquement correct.

Classement 1 Animaux à classer

```

graph TD
    A[Animaux à classer] --> B[Insecte]
    A --> C[Insecte]
    B --> D[Guêpe]
    C --> E[Mouche]
    B --> F[Araignées]
    C --> G[Crustacés]
    F --> H[Épeire, Tégénaire]
    G --> I[Crabe, Écrevisse]
        
```

Classement 2 Animaux à classer

```

graph TD
    A[Animaux à classer] --> B[qui volent]
    A --> C[qui ont des pattes]
    B --> D[Guêpe, Mouche]
    C --> E[Tégénaire, Épeire, Crabe, Écrevisse]
    D --> F[qui ont 8 pattes]
    D --> G[qui ont des pinces]
    F --> H[Épeire, Tégénaire]
    G --> I[Crabe, Écrevisse]
        
```

Classement 3 Animaux à classer

```

graph TD
    A[Animaux à classer] --> B[Araignées]
    A --> C[Crustacés]
    B --> D[Épeire, Tégénaire]
    C --> E[Crabe, Écrevisse]
    B --> F[Insecte qui pique]
    C --> G[Insecte qui ne pique pas]
    F --> H[Guêpe]
    G --> I[Mouche]
        
```

Classement 4 Animaux à classer

```

graph TD
    A[Animaux à classer] --> B[qui volent]
    A --> C[qui ne volent pas]
    B --> D[Guêpe, Mouche]
    C --> E[Tégénaire, Épeire, Crabe, Écrevisse]
    D --> F[qui ont des pinces]
    D --> G[qui n'ont pas de pinces]
    F --> H[Crabe, Écrevisse]
    G --> I[Épeire, Tégénaire]
        
```

Classement 5 Animaux à classer

```

graph TD
    A[Animaux à classer] --> B[Faux Insectes]
    A --> C[Insectes]
    B --> D[Épeire, Tégénaire]
    C --> E[Guêpe, Mouche]
    D --> F[qui pique]
    C --> G[qui ne pique pas]
    F --> H[Guêpe]
    G --> I[Mouche]
        
```

Item 58	Moyenne de l'échantillon
<b>Réponse correcte : le quatrième classement</b>	29%
Le 1 <sup>er</sup> classement	15%
Le 2 <sup>o</sup> classement	21%
Le 3 <sup>o</sup> classement	17%
Le 5 <sup>o</sup> classement	10%
Plusieurs choix	3%
Omission	5%

Source : (Fédération Wallonie-Bruxelles, s.d.[7])

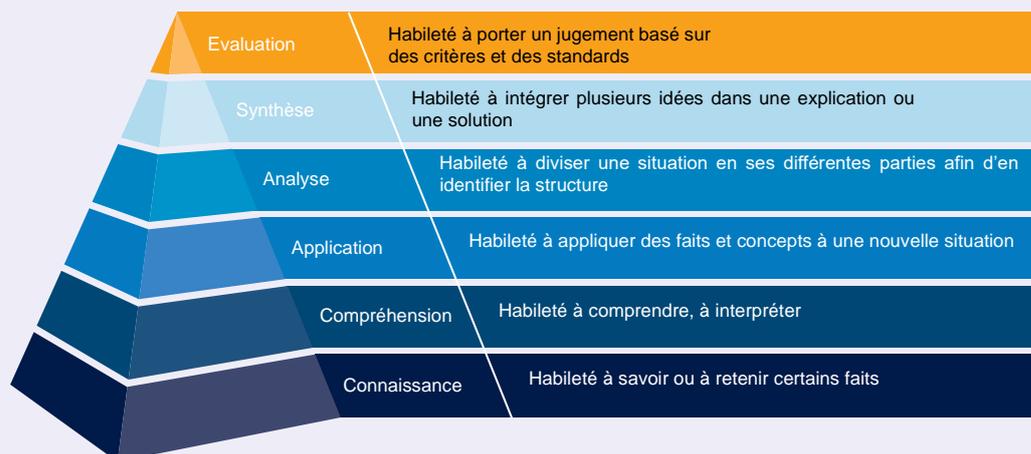


## Encadré 2.2

### La taxonomie de Bloom

En 1956, Benjamin Bloom et ses collaborateurs ont publié un cadre pour catégoriser les objectifs éducatifs, la Taxonomie des objectifs éducatifs, connu comme la taxonomie de Bloom. Le cadre se compose de six catégories principales : connaissance, compréhension, application, analyse, synthèse et évaluation. L'idée centrale de la taxonomie de Bloom est de classer ce que les éducateurs veulent que les élèves connaissent ou sachent faire, dans une hiérarchie allant du plus simple au plus complexe. Les différents niveaux sont considérés comme successifs, un niveau devant être maîtrisé avant que le niveau suivant puisse être atteint. En 2001, la taxonomie a été révisée par (Anderson et al., 2001<sup>[8]</sup>) dans le but de l'adapter aux objectifs de l'enseignement moderne, notamment en inversant l'ordre des deux niveaux les plus élevés.

#### Graphique 2.3 Catégories principales de la taxonomie de Bloom



Source : Université de Poitiers (2014<sup>[9]</sup>) <sup>[6]</sup>, Taxonomie de Bloom et roue pédagogique, <https://blogs.univ-poitiers.fr/troy/2014/05/09/taxonomie-de-bloom-et-roue-pedagogique/> (consulté le 13 juillet 2022).

Le plus souvent utilisée pour évaluer une variété de niveaux cognitifs, la taxonomie de Bloom a plusieurs fonctions et applications. Ainsi, disposer d'un ensemble organisé d'objectifs aide les enseignants à planifier et dispenser un enseignement approprié, concevoir des tâches et des stratégies d'évaluation valables, et s'assurer que l'enseignement et l'évaluation sont alignés sur les objectifs. En plus, la taxonomie peut aider les enseignants à mieux comprendre le processus éducatif et le développement cognitif complexe, ce qui permet de fixer des priorités pour les objectifs d'apprentissage et peut orienter l'organisation des leçons. Par exemple, les compétences de niveau inférieur devraient être développées avant l'introduction de compétences de niveau supérieur.

La taxonomie de Bloom peut également être utilisée comme une liste de contrôle pour s'assurer que tous les niveaux d'un domaine ont été évalués.

Source :

1. Armstrong, P. (2010<sup>[10]</sup>), Bloom's Taxonomy, <https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/> (consulté le 13 juillet 2022) ;
2. Huit, W. (2011<sup>[11]</sup>), « Bloom et al.'s taxonomy of the cognitive domain », <http://www.edpsycinteractive.org/topics/cognition/bloom.html> (consulté le 13 juillet 2022) ;
3. University of Central Florida (s.d.<sup>[12]</sup>), Bloom's Taxonomy, <https://ctl.ucf.edu/teaching-resources/course-design/blooms-taxonomy/> (consulté le 13 juillet 2022).

### 2.2.1.2 Les avantages et inconvénients de chaque format de question

D'une façon générale, chaque **format de question** présente des avantages et des inconvénients. En voici une brève synthèse.

**Tableau 2.2 Avantages et inconvénients des différents formats de questions**

	 <b>Questions fermées</b>		 <b>Questions ouvertes</b>	
	<b>Alternatives</b> Deux solutions mutuellement exclusives (vrai-faux, oui-non)	<b>QCM</b> Choix entre plusieurs propositions dont une est correcte <sup>1</sup>	<b>Réponse brève</b> L'élève fournit une réponse dont la forme est imposée (une date, une somme, un nom...)	<b>Réponse construite</b> L'élève construit sa réponse
 <b>Avantages</b>	Ces formes de question permettent : <ul style="list-style-type: none"> <li>• Une large couverture de la matière (les élèves y répondant plus rapidement, les items peuvent être plus nombreux)</li> <li>• Une correction rapide et plus objective</li> <li>• Une possibilité moindre de contourner la difficulté que dans les questions ouvertes à réponse construite.</li> </ul>			Cette forme de question : <ul style="list-style-type: none"> <li>• Est rapidement rédigée</li> <li>• Permet d'évaluer la structuration de la pensée.</li> </ul>
 <b>Inconvénients</b>	L'élève peut deviner la bonne réponse	Difficulté à rédiger des distracteurs qui informent sur les démarches des élèves	Le caractère unique et personnel de la réponse à une question ouverte (surtout à réponse construite) implique : <ul style="list-style-type: none"> <li>• un temps de correction assez long</li> <li>• des désaccords entre correcteurs</li> <li>• des biais tels que l'effet de halo, de contagion et d'ordre de correction*</li> <li>• une difficulté à réaliser un examen couvrant toute la matière</li> <li>• Une possibilité de contourner la difficulté</li> </ul>	
	Ces formes de questions peuvent entraîner : <ul style="list-style-type: none"> <li>• Une tentation de poser des questions de détail et une réduction de la vision du monde</li> <li>• Une mémorisation des distracteurs (dans certaines conditions seulement)</li> <li>• Une difficulté à évaluer des performances relevant de l'expression spontanée</li> <li>• Une facilité à copier sur d'autres élèves</li> </ul>			

Note:

1. Il existe également des questions à choix multiples dont plusieurs solutions sont correctes.
2. Dans une question ouverte à réponse construite, un élève peut espérer « grappiller » quelques points en faisant la preuve de connaissances qu'il possède même si ce n'est pas précisément ce qui est ciblé par la question.
3. Les biais tels que l'effet de halo, de contagion et d'ordre de correction seront exposés dans le chapitre 4 dédié à la correction.

Source : Développé par les auteurs du guide

Dans le cas d'une évaluation que l'on souhaite valide, fidèle et équitable, il convient de rechercher la plus grande objectivité possible dans les corrections. Dans ce contexte, les questions fermées de type binaire (vrai/faux), les QCM et les questions ouvertes à réponse brève présentent l'avantage d'être corrigées rapidement et d'une façon fidèle. Par exemple, les tests standardisés sont appréciés pour leur fidélité, particulièrement si le codage des réponses est automatisé.

Les questions ouvertes à réponse construite peuvent fournir davantage d'informations sur les démarches mises en œuvre par l'élève ; elles sont préférables pour l'évaluation de certains aspects (par exemple, la réflexion critique). En revanche, dans le contexte qui nous occupe, elles présentent



l'inconvénient majeur d'une correction longue, complexe et n'offrant que peu de garantie de fidélité (désaccord entre correcteurs) et d'objectivité (effet de halo, de contagion, d'ordre de correction)<sup>4</sup>.

## 2.2.2 L'importance de varier les questions, leurs formats et leur complexité

Chacune des compétences ou capacités sélectionnées sera évaluée par plusieurs questions, de différents formats et de différents niveaux de complexité (voir encadré 2.2). Cette condition est incontournable pour aboutir à un diagnostic valide, qui permette d'identifier le plus précisément possible ce que les élèves sont capables de faire à un moment donné et ce qui leur pose problème.

Par exemple, si une capacité est évaluée exclusivement via des QCM et une autre capacité évaluée exclusivement via des questions ouvertes, il existe une confusion potentielle entre l'objet évalué (la capacité) et la modalité d'évaluation : comment interpréter d'éventuelles différences de résultats ?

Imaginons une épreuve de lecture dans laquelle la capacité « Identifier » est évaluée uniquement par des questions fermées (vrai-faux ou à choix multiple) et la capacité « Réagir » uniquement au travers de questions ouvertes construites. Si on veut ensuite comparer les résultats des élèves pour ces deux capacités, on se trouve en difficulté vu la confusion entre l'objet évalué (la capacité) et la modalité d'évaluation (questions fermées ou questions ouvertes). Les différences de résultats correspondent-elles à une moins bonne maîtrise de la capacité de lecture (« identifier » ou « Réagir ») ou à une différence dans la capacité à répondre à des questions fermées ou à élaborer des réponses ouvertes ? Les deux étant totalement confondus, il est impossible de le savoir et le risque est de mal interpréter les différences de résultats.

Pour aboutir à un diagnostic le plus précis possible, l'importance d'évaluer un domaine de savoir ou une capacité à l'aide de questions de niveaux de difficulté variés a été soulignée. Au moment de sélectionner les compétences à évaluer, il

« Pour aboutir à un diagnostic le plus précis possible, l'importance d'évaluer un domaine de savoir ou une capacité à l'aide de questions de niveaux de difficulté variés a été soulignée. »

convient donc de bien examiner la possibilité de les décliner en différents niveaux de difficulté et de les aborder à travers différents formats de questions.



## Exemple de question mesurant une capacité à des niveaux variés de difficulté

Pour illustrer la possibilité de mesurer une capacité à des niveaux variés de difficulté, voici, comment, pour une épreuve d'évaluation en mathématiques en 1<sup>e</sup> année du cycle secondaire collégial, il est possible de décliner la compétence « Construire une figure simple » en proposant un tracé en trois situations :

- L'utilisation d'instruments, déjà bien travaillée en primaire (équerre et règle graduée), pour la première situation :

### Situation 1

Construire un losange dont les diagonales mesurent 6 cm et 8 cm.

Niveau visé : 1<sup>e</sup> Année Collégiale

- L'utilisation d'instruments, déjà bien travaillée en primaire (rapporteur), pour la deuxième situation :

### Situation 2

Construire un triangle ABC sachant que l'angle de sommet B mesure  $70^\circ$  et l'angle de sommet C mesure  $40^\circ$ .

Niveau visé : 1<sup>e</sup> Année Collégiale

- La mobilisation d'un raisonnement déductif dans la troisième situation :

### Situation 3

Construire le point A pour que le triangle ABC soit isocèle de base [BC] et que le triangle ADC soit rectangle en D.

Réalise les constructions en crayon et ne les efface pas.

Trace les triangles en bleu.

Niveau visé : 1<sup>e</sup> Année Collégiale

Source : Développé par les auteurs du guide

En somme, la recommandation générale est d'inclure dans une même épreuve plusieurs questions de différents formats et de différents niveaux de complexité pour mesurer une connaissance, une compétence ou une capacité de manière précise. La table de spécification permettra aux enseignants d'assurer que les connaissances et les compétences ou capacités évaluées sont couvertes par un nombre de questions adéquat et une variété de formats et de niveaux de difficulté

et de complexité. C'est à ces conditions que l'on pourra approcher au mieux la validité souhaitée, et donc, la fiabilité de l'épreuve.

## 2.2.3 Quelques règles à respecter:

### 2.2.3.1 S'assurer que l'item évalue bien ce qui est visé

En lecture, par exemple, il faut à tout prix éviter les questions de compréhension de texte auxquelles il est possible de répondre sans avoir lu le texte. Dans ce cas, les questions sont des questions de connaissance, et alors, elles n'évaluent pas ce qui est visé – la compréhension du texte (voir l'exemple ci-dessous).



### Exemple de question portant sur un texte qui décrit quelques grandes villes du Maroc

**Quelle est la capitale du Maroc ?**

- Casablanca
- Tanger
- Rabat
- Marrakech

Source : Développé par les auteurs du guide

### 2.2.3.2 La formulation des questions : clarté, précision et concision

Quel que soit le format de question, la formulation de la consigne et de la question doit être aussi brève que possible, précise et claire pour tous les élèves. Il faut éviter les termes vagues et aussi éviter d'inclure dans la question des informations sans lien direct avec ce que l'on évalue.

Il n'est pas rare qu'une réponse fautive soit due à une mauvaise interprétation d'une consigne plutôt qu'à la non-maîtrise de la capacité ciblée. De façon générale, comment interpréter les résultats des élèves en termes de niveau de maîtrise de la capacité ciblée si on n'a pas la certitude que la consigne est comprise de la même façon par tous ?

Une bonne façon de réfléchir à la clarté de la consigne et de la question est d'identifier les compétences minimales à mobiliser pour y répondre. On peut également se poser les deux questions suivantes :

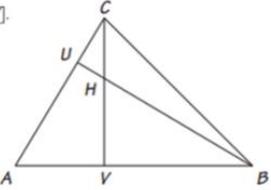
- Ce que constitue la réponse attendue à chaque question est-il absolument clair ?
- Les élèves savent-ils clairement ce qu'on attend d'eux (sans nécessairement savoir comment s'y prendre) ?



## Exemple de clarification de la consigne d'une question en mathématique

Voici, par exemple, comment une question de mathématiques a été clarifiée durant la conception. Initialement, la question était la suivante :

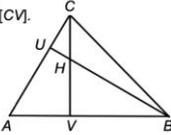
Dans le triangle  $ABC$  ci-contre, on a tracé les hauteurs  $[BU]$  et  $[CV]$ . Celles-ci se coupent en  $H$ .



Démontrez que les triangles  $BHV$  et  $CHU$  sont semblables.

Lors de l'élaboration des critères de correction de la question, le terme « démontre » a suscité des débats : en effet, les exigences des enseignants à ce niveau peuvent être très contrastées selon les établissements scolaires. Afin que les élèves puissent cibler précisément ce qui était attendu, la question a donc été reformulée comme suit :

Dans le triangle  $ABC$  ci-contre, on a tracé les hauteurs  $[BU]$  et  $[CV]$ . Celles-ci se coupent en  $H$ .



Démontrez que les triangles  $BHV$  et  $CHU$  sont semblables.

Hypothèses (on sait que...)

Conclusion (on doit démontrer que...)

Démonstration
ATTENTION : N'oubliez pas d'énoncer les propriétés utilisées

L'idée ici est de s'assurer que les élèves comprennent bien que, derrière le terme "démontrez", on attendait la formulation des hypothèses, de la conclusion, et de la démonstration proprement dite. Ce n'est donc pas une aide au sens de décomposition de la tâche en sous-tâches, mais plutôt une clarification de la consigne "démontrez", comprise différemment selon la manière dont l'enseignant a travaillé en classe, avec les élèves.

Source : (Fédération Wallonie-Bruxelles, s.d.[6]) (adapté par les auteurs du guide)

### 2.2.3.3 L'indépendance des items

L'indépendance des items entre eux constitue une autre règle générale qui peut être résumée de la façon suivante : il faut être attentif au fait que la réponse à un item ne dépende pas de la réponse à un item précédent.



#### Exemple d'items dépendants dans une épreuve de mathématiques

La question ci-dessous illustre ce problème de dépendance entre items :

Dans chaque cas, entoure le nombre le plus grand.

a.)  $2\sqrt{5}$  et  $5\sqrt{2}$

b.)  $\frac{1}{3+2\sqrt{5}}$  et  $\frac{1}{3+5\sqrt{2}}$

Niveau visé : 3<sup>e</sup> Année Collégiale Source : Développé par les auteurs du guide

A priori, ces items a et b dans l'exemple ci-dessus sont intéressants car, bien qu'évaluant le même processus (comparer des expressions contenant des racines carrées), ils présentent un niveau de difficulté contrasté : l'item b implique en effet d'analyser une expression fractionnaire, ce qui est sans doute plus difficile pour les élèves que d'analyser les deux expressions de l'item a. La personne qui a conçu cet item a sans doute souhaité réduire le temps nécessaire pour répondre à l'item b, en permettant à l'élève de s'appuyer sur l'analyse réalisée à l'item a. Effectivement, un élève qui répond correctement à l'item a pourra simplement analyser l'impact d'un dénominateur plus grand sur la taille d'une fraction. Toutefois, si l'élève se trompe à l'item a, il risque d'être pénalisé à l'item b, même s'il a bien raisonné à l'item b, simplement parce qu'il s'appuie sur une réponse erronée à l'item a. En ce sens, ces deux items sont dépendants.



#### Solution pour l'exemple d'items dépendants

Soit proposer d'autres expressions pour l'item b, soit préciser dans la correction que l'on tiendra compte de la réponse proposée à la première partie de la question (item a) pour juger la qualité de la réponse à la deuxième (item b) : si l'élève entoure l'item a et l'item b, il aura un point sur deux, car son analyse à l'item b est tout à fait pertinente s'il pense que b est plus grand que a.

En plus, de tels items dichotomiques ne sont acceptables que dans un questionnaire en QCM avec redondance de la capacité évaluée. Sinon, la réponse au hasard donne une chance sur deux de répondre juste.

Un autre exemple d'items dépendants est présenté ci-dessous en matière de compréhension de textes.



### Exemple d'items dépendants dans une épreuve en compréhension de textes en 2e secondaire

**a) Quel est le métier de la maman de Naïa ?**

COCHE la réponse correcte

- Infirmière
- Gardienne du camp
- Professeur
- Assistante sociale

**b) Sur quel(s) indice(s) du texte te bases-tu pour répondre ?**

Source : Développé par les auteurs du guide

Ces deux items sont dépendants. L'élève qui a coché une réponse incorrecte à l'item a ne pourra fournir un indice correct tiré du texte. La réussite à l'item b est conditionnée par la réussite à l'item a.



### SOLUTION

Il y a deux possibilités pour éviter le problème. Soit, on ne pose qu'une question :

« *Quels indices du texte montrent que le métier de la maman de Naïa est assistante sociale ?* ». Soit, pour attribuer les points, on exige que l'élève coche « assistante sociale » **et** qu'il fournisse un indice correct du texte. Ceci rend toutefois la question plus exigeante.

### 2.2.3.4 Dans certains cas, éviter de confronter l'élève aux erreurs

Dans les apprentissages où la mémoire visuelle intervient, il faut éviter l'usage de distracteurs que l'élève peut retenir. C'est typiquement le cas pour l'orthographe : l'élève peut fixer des erreurs en pensant que c'est la réponse correcte. Ceci concerne surtout les jeunes élèves.



## Exemples de confrontation des élèves aux erreurs

Coche le mot qui correspond à l'image.



- Mento
- Manteau
- Mantôt
- Mentau

Source : Développé par les auteurs du guide

## 2.2.4 Focus sur la construction de QCM

Un QCM est composé :

- D'une **consigne** (on explique à l'élève ce qu'il doit faire (par exemple, Coche une case...))
- D'une **amorce** (la définition du problème, la question)
- De **propositions de solution** (la solution correcte à découvrir parmi plusieurs distracteurs).

La consigne doit être aussi courte que possible et commencer par un verbe qui indique clairement le comportement attendu de l'élève (Lisez attentivement, Cochez, Entourez...).

Le choix des distracteurs (choix de réponses dans un QCM autres que la réponse correcte) est crucial. Ceux-ci doivent se rapprocher de la réponse correcte par certains aspects, sans en être trop proches. En d'autres termes, ils doivent être attractifs, mais pas trop<sup>5</sup>.

Une QCM dont les distracteurs sont simplement des réponses inventées un peu au hasard, voire des réponses fantaisistes, n'a que peu d'intérêt. Un élève un peu malin aura vite fait d'éliminer les réponses peu vraisemblables. Dans ce cas, on n'évalue pas vraiment la compétence ou le savoir visé, mais son intelligence.

« Les QCM seront plus intéressants d'un point de vue diagnostic si certains des distracteurs correspondent à des erreurs courantes des élèves. »

Les QCM seront plus intéressants d'un point de vue diagnostic si certains des distracteurs correspondent à des erreurs courantes des élèves. Tout d'abord, ceci donne une certaine vraisemblance aux propositions (aux yeux des élèves). En outre, cette façon

de procéder fournit des indications plus précises sur les démarches de l'élève et peut clarifier les pistes de remédiation pour l'élève. Il est rarement possible de trouver trois distracteurs correspondant à des erreurs typiques, mais si au moins un des distracteurs y correspond, c'est déjà intéressant.

Dans les sections qui suivent, nous présentons des suggestions de bonnes pratiques pour améliorer la qualité des questions à choix multiples (QCM) :

#### 2.2.4.1 Réduire la part du hasard

Pour les QCM, il est conseillé de prévoir quatre propositions de réponses au moins (la réponse correcte et trois distracteurs) pour réduire la part du hasard à 25 % maximum.

Dans les questions de type vrai-faux, l'élève a en effet une chance sur deux de répondre correctement en choisissant au hasard et sa compétence peut vite être surestimée si on n'utilise que ce format. Une compétence ne peut donc pas être évaluée au travers d'un ou quelques items d'un tel format. Il convient de multiplier les items et/ou de varier les formats de questions.



### Exemples de QCM qui ne réduisent pas la part du hasard

#### Coche la réponse correcte

Fatima écrit à ses amis pour...

- les inviter.
- les remercier.

#### Coche la réponse correcte

La lettre a un caractère...

- amical.
- officiel.

Source : Développé par les auteurs du guide



## SOLUTION

On peut éviter le problème en regroupant plusieurs items fermés de type vrai-faux en une seule question et n'attribuer les points (selon la pondération prévue) que si l'élève a répondu correctement à tous les items regroupés. On peut aussi ajouter une troisième possibilité comme « Le texte ne permet pas de le dire », ce qui réduit la part du hasard à un tiers plutôt qu'à 50 %.



## Exemple de question qui réduit la part du hasard

Vrai, faux ou le texte ne permet pas de le dire ?

**TRACE une croix** au bon endroit.

	Vrai	Faux	Le texte ne permet pas de le dire.
Les plantes carnivores sont des espèces végétales.			
Les plantes carnivores peuvent vivre vingt ans.			
Les plantes carnivores se nourrissent uniquement d'insectes, d'araignées et d'escargots.			

Source : (Fédération Wallonie-Bruxelles, s.d.[13])

#### 2.2.4.2 Concevoir des distracteurs qui renseignent sur les démarches erronées des élèves



### Exemple 1 de distracteurs qui renseignent sur les démarches erronées des élèves

**Sept personnes ont déjeuné dans un restaurant. Le plat est facturé à 70 DH. Elles ont bu chacune une bouteille d'eau minérale à 9 DH la bouteille.**

Le montant de l'addition s'élève à :

- 513 DH
- 553 DH
- 560 DH
- 520 DH

Coche la bonne réponse.

Source : Développé par les auteurs du guide



### SOLUTION

Les distracteurs traduisent des erreurs de calcul (mais il est difficile d'identifier lesquelles). Il serait intéressant qu'ils traduisent des erreurs de raisonnement.

Par exemple,

- oubli de tenir compte des bouteilles d'eau 4900  $\rightarrow (7 \times 70)$
- oubli de multiplier les bouteilles d'eau 499  $\rightarrow (7 \times 70) + 9$
- oubli de multiplier les repas 133  $\rightarrow 70 + (7 \times 9)$
- oubli de multiplier par le nombre de convives 79  $\rightarrow 70 + 9$
- oubli de la priorité des opérations 4473  $7 \times 70 + 7 \times 9$ .



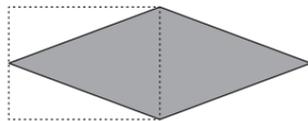
## Exemple 2 des distracteurs qui renseignent sur les démarches erronées des élèves

On te donne chaque fois l'aire du rectangle en pointillé. Recherche l'aire des figures grisées, sans utiliser ton matériel.



Aire du rectangle en pointillé :  $8 \text{ cm}^2$

Aire du triangle isocèle grisé : .....  $\text{cm}^2$



Aire du rectangle en pointillé :  $12 \text{ cm}^2$

Aire du losange grisé : .....  $\text{cm}^2$

Niveau visé : 1<sup>re</sup> Année Collégiale

Source : (Fédération Wallonie-Bruxelles, s.d.<sup>[6]</sup>)

Dans cette question, l'idée est que l'élève repère visuellement que l'aire du losange est la moitié de celle du rectangle, ce qui peut être intéressant, dans le cadre d'une question de raisonnement. Toutefois, les distracteurs informent peu sur des démarches erronées comme, par exemple, les élèves qui oublient de diviser l'aire du rectangle par 2 ( $630 \text{ m}^2$ ), ceux qui divisent l'aire du rectangle par 4 ( $157,5 \text{ m}^2$ ) en étant attirés par les 4 rectangles déterminés par les diagonales ou encore des élèves qui additionnent les côtés comme s'il s'agissait du périmètre et non de l'aire ( $106 \text{ m}^2$ ).



### Exemple 3 des distracteurs qui renseignent sur les démarches erronées des élèves

Deux caissiers ont vendu 10 640 billets d'entrée au stade ; le premier a vendu 250 billets de plus que le deuxième. Calcule le nombre de billets vendus par le premier.

5 245 billets

5 345 billets

5 445 billets

5 545 billets

Niveau visé : 1<sup>e</sup> Année Collégiale

Source : Développé par les auteurs du guide

La présence de grands nombres complique inutilement le problème. Si l'idée est de vérifier dans quelle mesure les élèves parviennent à résoudre un tel problème de partages inégaux, ce serait plus intéressant de limiter la complexité des calculs à effectuer.

#### 2.2.4.3 Choisir des contextes authentiques



### Exemple de contextes authentiques

À la rentrée des classes, l'enseignant d'arts plastiques a acheté 45 pochettes de 12 feutres pour ses élèves.

Combien de feutres met-il à la disposition de ses élèves ?

540 feutres

57 feutres

33 feutres

135 feutres

Niveau visé : 1<sup>e</sup> Année Collégiale

Source : Développé par les auteurs du guide

Le contexte choisi n'a en fait aucun sens. Quel est l'intérêt de compter le nombre total de feutres ?



## SOLUTION

La question pourrait être : « L'enseignant veut acheter des pochettes de feutres pour ses 45 élèves. Chaque pochette coûte 12 F. Combien payera-t-il ? ».

Contrairement aux questions précédentes, deux des distracteurs fournissent quelques informations :

- 57 : l'élève a additionné les données au lieu de les multiplier.
- 33 : l'élève a soustrait les données au lieu de les multiplier.

En revanche, pour le troisième distracteur, on ne voit pas bien pourquoi l'élève multiplierait le nombre de pochettes par 3.

### 2.2.4.4 Choisir soigneusement les supports

Les questions ci-dessous avaient été proposées pour une évaluation à large échelle. Aucune n'a été retenue en raison des défauts de conception qu'elles contiennent.



## Exemple de fautes de conception des questions

Lis le texte et réponds aux questions qui suivent.

Dans ton village, chaque année, les paysans font des feux de brousse pour préparer les champs. Ils se plaignent que les forêts soient remplacées par des savanes et que les récoltes ne soient plus bonnes.

### 1. Que font les paysans chaque année dans ton village ?

- A. Cultivent l'arachide
- B. Se plaignent des savanes
- C. Font des feux de brousse pour préparer les champs
- D. Vendent leurs récoltes

### 2. Pourquoi, les paysans font-ils des feux de brousse ?

- A. Pour détruire les forêts
- B. Pour préparer les champs
- C. Pour planter d'autres arbres
- D. Pour libérer de l'espace

**3. À quel personnage correspond le mot « ils » dans le texte ?**

- A. L'homme
- B. Enfants
- C. Femmes
- D. Paysans

Source : Développé par les auteurs du guide

Concernant l'item 1:

- Le fait d'indiquer « dans ton village » plutôt que d'évoquer un village fictif peut amener les élèves à répondre, à juste titre, que dans leur village, chaque année, les paysans vendent leurs récoltes. Ceci rend les trois distracteurs plausibles, même s'ils ne se rapportent pas directement au texte.
- La proposition C est une partie de phrase parfaitement identique à celle du texte. Ceci constitue un indice qui peut conduire l'élève à la solution, même s'il n'a pas compris ce qu'il a lu. Par ailleurs, cette proposition est plus longue que les autres, ce qui la rend plus attractive.
- En outre, cette proposition C fournit la réponse à l'item 2. D'une façon générale, il convient de toujours vérifier que la réponse à un item n'est pas fournie par un autre item.

Concernant l'item 3:

- « À quel personnage » devrait être au pluriel puisqu'il s'agit de mettre en correspondance avec le pronom « ils ».
- La reconnaissance des marques du pluriel intervient puisqu'une seule proposition est au singulier (ce qui l'exclut dès le départ) et les autres au pluriel.
- Les distracteurs et le texte ne conviennent pas pour évaluer la compréhension du pronom « ils » puisque seuls les paysans sont évoqués dans le texte.
- Enfin, vu sa nature et les informations qu'il contient, le texte proposé ne permet d'évaluer que la reconnaissance d'informations explicites voire de correspondances mot à mot.

#### 2.2.4.5 Formuler des phrases syntaxiquement correctes

Même si cela paraît évident, il convient d'être attentif à la cohérence syntaxique entre l'amorce et les distracteurs.



#### Exemple 1 de QCM avec des phrases syntaxiquement incorrectes

**L'organe qui sécrète la bile est le**

1. foie
2. rate
3. estomac
4. hypophyse

Source : Développé par les auteurs du guide

La formulation de l'amorce (particulièrement, l'utilisation de « le ») élimine directement les solutions 2, 3 et 4.



#### Exemple 2 de QCM avec des phrases syntaxiquement incorrectes

Fès est à	l'Ouest	de Rabat
	le Nord	
	l'Est	
	le Sud	

Source : Développé par les auteurs du guide

La formulation de l'amorce élimine directement les solutions 2 et 4 (Fès est « à le » Nord ne « fonctionne » pas).

Dans l'exemple suivant, des défauts de formulation rendent la recherche de la solution soit plus complexe, soit plus simple.



### Exemple 3 de QCM avec des phrases syntaxiquement incorrectes

Dans TOUS les déserts, il tombe :

1. au moins 30 cm d'eau.
2. moins de 30 cm d'eau.
3. moins de 10 cm d'eau.
4. il ne pleut jamais.

Source : Développé par les auteurs du guide

Plusieurs faiblesses caractérisent ce QCM :

- Une erreur syntaxique dans la quatrième proposition : elle est incohérente par rapport à l'amorce de la question (« il tombe... il ne pleut jamais »).
- Les propositions 2 et 3 sont emboîtées : si la solution 3 est correcte, la solution 2 l'est automatiquement.
- Enfin, la présence du mot « jamais » renforce le caractère peu attractif du distracteur 4. Des termes comme « tous », « toujours » ou « jamais » sont à éviter dans la formulation des distracteurs.

#### 2.2.4.6 Éviter les formes négatives

Les formes négatives dans la question compliquent inutilement la tâche des élèves. Il ne suffit pas d'éviter les "ne... pas". Des mots comme absence, jamais, interdiction, nier, supprimer, éviter, impossible, etc. comportent une idée de négation. Il n'est pas rare que les erreurs des élèves soient le reflet d'une erreur de compréhension de la question plutôt que de la non-maîtrise du savoir-faire visé. Les erreurs à ce genre de questions sont très difficiles à interpréter :



### Exemple de formes négatives dans une question fermée

Vrai ou faux ?

Un quadrilatère dont les diagonales ne se coupent pas en leur milieu n'est pas un parallélogramme.

Source : Développé par les auteurs du guide

Formellement, cette question vise les propriétés des diagonales d'un quadrilatère. La présence d'une double négation complique fortement l'analyse.



## SOLUTION

Il serait plus simple de supprimer cette double négation. La formulation proposée est : « Un quadrilatère dont les diagonales se coupent en leur milieu est un parallélogramme ».

### 2.2.4.7 *Formuler des propositions d'égale longueur, de même vraisemblance et d'un degré de complexité équivalent*

Les propositions correctes ont tendance à être souvent plus longues que les propositions fausses, et comme elles ont l'air plus précises, elles sont plus attractives.



## Exemple 1 de propositions de différentes longueurs/ vraisemblance/ degré de complexité

**Coche la proposition correcte.**

Watson et Crick sont célèbres parce qu'ils...

- sont physiciens
- ont mis au point le code génétique
- ont découvert l'origine de la trisomie 21
- ont élaboré le modèle d'ADN représenté par une double hélice torsadée constituée de quatre bases azotées, de désoxyribose et d'acide phosphorique

Source : Développé par les auteurs du guide



## Exemple 2 de propositions de différentes longueurs/ vraisemblance/ degré de complexité

**Dans le domaine des statistiques, la moyenne est...**

- Un nombre, positif, négatif, ou nul, qui résume l'information fournie par l'ensemble des données d'une série statistique
- Le nombre de données de la population
- La valeur apparaissant le plus fréquemment dans la série de données
- Le milieu d'un ensemble de données classées de la plus petite à la plus grande

Niveau visé : 3<sup>e</sup> Année Collégiale

Source : Développé par les auteurs du guide

La dernière proposition dans l'exemple 1 et la première proposition dans l'exemple 2 sont plus longues que les autres, ce qui les rend plus attractives pour les élèves, qui les choisiront même s'ils n'ont aucune idée de la réponse correcte.

### 2.2.4.8 Éviter les indices qui conduisent à la réponse

Par exemple, l'utilisation d'un terme identique dans l'amorce et dans la proposition correcte aura tendance à attirer les élèves (voir l'exemple sous la section Choisir soigneusement les supports).

### 2.2.4.9 Formuler une seule idée par proposition

Quand une proposition contient plus d'une information, les erreurs des élèves sont très difficiles à interpréter. Par rapport à quelle information l'élève a-t-il commis une erreur ?



## Exemple de plusieurs idées par proposition

Vrai ou faux ?

Dans une série de données, la moyenne arithmétique consiste à calculer la somme des données et à la diviser par le nombre de données : c'est un nombre qui est toujours différent de 0.

Entoure ton choix      VRAI    FAUX

Niveau visé : 2<sup>e</sup> Année Collégiale

Source : Développé par les auteurs du guide

Dans cette question vrai-faux, la proposition contient deux informations. Seule la deuxième partie de la proposition est fautive (la moyenne peut être nulle). Comment interpréter les acquis des élèves qui répondent « faux » ? Considèrent-ils que la définition de la moyenne est incorrecte ou ont-ils bien compris que la moyenne n'est pas un indicateur de dispersion ?

Cet exemple illustre aussi l'importance d'avoir cerné précisément ce que l'on souhaite évaluer au travers de chaque question. Effectivement, chacune des deux informations contenues dans la proposition porte sur une connaissance différente.

#### 2.2.4.10 Vérifier le caractère exclusif de la réponse correcte

Cette vérification peut s'effectuer en répondant à la question : les conditions fournies dans la proposition sont-elles suffisantes pour que la proposition soit absolument vraie ou absolument fautive ?



### Exemple de questions sans caractère exclusif

Fable de Jean de la Fontaine, *Le loup et l'agneau*

#### Pourquoi le loup s'en prend-il à l'agneau ?

- Pour venger son espèce
- Pour se nourrir
- Pour le punir de l'avoir insulté
- Pour le punir de polluer son eau

Source : Développé par les auteurs du guide dans le cadre du prétest 2016 d'une évaluation externe en Belgique francophone. Item non diffusé.

Plusieurs réponses (voire toutes) sont acceptables.

#### 2.2.4.11 Prévoir plus de propositions que d'éléments à appairer

Une question à appariement est une variante de QCM dans laquelle, pour chaque élément à relier, l'élève a le choix entre plusieurs possibilités. Toutefois, pour qu'il en soit ainsi, il convient de prévoir un plus grand nombre de propositions que d'éléments sur lesquels porte la question.

Si les éléments à relier entre eux sont en nombre identique, l'élève qui a commis une erreur en commettra automatiquement une deuxième. L'élève qui n'a pas commis d'erreur dans les trois premiers items obtient automatiquement la dernière réponse correcte. Les items sont donc dépendants.



## Exemple de QCM avec plus de propositions que d'éléments à appairer

Relie au moyen d'un trait chaque oeuvre d'opéra à son compositeur.

- |                       |                       |                                 |
|-----------------------|-----------------------|---------------------------------|
| Carmen                | <input type="radio"/> | <input type="radio"/> Rossini   |
| Les Noces de Figaro   | <input type="radio"/> | <input type="radio"/> Bizet     |
| Manon                 | <input type="radio"/> | <input type="radio"/> Massenet  |
| Le Barbier de Séville | <input type="radio"/> | <input type="radio"/> Mozart    |
|                       |                       | <input type="radio"/> Berlioz   |
|                       |                       | <input type="radio"/> Gounod    |
|                       |                       | <input type="radio"/> Offenbach |

Source : Développé par les auteurs du guide

### 2.2.5 Focus sur la construction des questions ouvertes

Bien que les questions ouvertes (à réponse brève ou construite) soient plus faciles et rapides à rédiger que les QCM, les critères de correction sont plus complexes à concevoir et à utiliser, particulièrement pour les questions à réponse construite.

Une bonne question ouverte à réponse construite doit être élaborée en étroite interaction avec les critères de correction que l'on envisage. En effet, il n'est pas rare, lorsqu'on réfléchit aux critères de correction d'une telle question que l'on se rende compte qu'elle sera en réalité difficile à corriger de manière fidèle et équitable. Des modifications dans l'énoncé de la question pourront alors être envisagées pour éviter ce problème.

Pour illustrer ce propos, voici une question pour laquelle l'élaboration des critères de correction a permis d'améliorer la formulation de la question. Cette question portait sur l'interprétation des données présentées dans le tableau suivant :



## Exemple de question pour laquelle l'élaboration des critères de correction permet d'améliorer la formulation

Les données du tableau ci-dessous présentent l'âge obligatoire d'entrée à l'école pour les 28 pays membres de l'Union européenne (UE) pour l'année scolaire 2019-2020.

Âge obligatoire d'entrée à l'école	Nombre de pays de l'UE (effectifs)
3	2
4	3
5	7
6	15
7	1

Source : (Fédération Wallonie-Bruxelles, s.d.<sup>[14]</sup>) d'après (Rakocevic et Fournier, 2020<sup>[15]</sup>)

L'idée de la question était d'amener les élèves à critiquer une affirmation d'un ministre qui considérait que son pays était le seul à fixer l'âge d'entrée à l'école le plus tôt dans la vie.

Étant donné l'intention de l'item qui consistait à interpréter les données d'un tableau, il était nécessaire que les élèves fondent leur argumentation sur une analyse des données du tableau. Les critères de correction envisageaient donc deux types de réponses acceptables :

- L'affirmation du ministre est fausse car deux pays imposent aux enfants d'entrer à l'école à 3 ans.
- L'affirmation du ministre est fausse car la seule case du tableau où l'effectif est 1 concerne l'âge le plus élevé d'entrée à l'école.

Vu la nécessité d'utiliser explicitement les données du tableau pour répondre à la question, cette dernière a été formulée comme suit :

*Un ministre d'un pays membre de l'UE affirme que son pays est le seul à fixer l'âge obligatoire d'entrée à l'école le plus tôt dans la vie. Précise si cette affirmation est correcte et explique ta réponse en te basant sur les données du tableau.*

Lorsqu'on établit des **critères de correction**, il est indispensable de cibler très précisément la (ou les) capacité(s) que l'on souhaite évaluer. Par la suite, il s'agit de définir le critère et l'indicateur de réussite correspondant. Les **critères de réussite** décrivent les qualités de la production attendue. Les indicateurs de réussite quant à eux décrivent les signes mesurables qui permettent de statuer sur le respect des critères définis. Les **indicateurs de réussite** gagnent également à être accompagnés d'exemples de réponses prototypiques d'élèves, acceptables et non acceptables.

Pour illustrer ces propos, voici la manière dont la capacité évaluée, le critère et l'indicateur de réussite ont été définis dans le cadre de la question présentée ci-dessus.



## Exemple de critère de correction pour une question ouverte

<b>Question</b>
Un ministre d'un pays membre de l'UE affirme que son pays est le seul à fixer l'âge obligatoire d'entrée à l'école le plus tôt dans la vie. Précise si cette affirmation est correcte et explique ta réponse en te basant sur les données du tableau.
<b>Capacité évaluée</b>
Critiquer une information, en référence à un ensemble de données statistiques.
<b>Critère de réussite</b>
Analyse correcte de l'affirmation du ministre en référence aux données du tableau
<b>Indicateur de réussite</b>
Faux, avec une explication pertinente basée sur les données du tableau.
<b>Exemples de réponses</b>
Exemples de réponse acceptable <ul style="list-style-type: none"> <li>- C'est faux car il y a deux pays qui fixent l'entrée à l'école à 3 ans</li> <li>- C'est faux, car 7 ans n'est pas l'âge où les élèves entrent à l'école le plus tôt dans la vie.</li> </ul> Exemples de réponses non acceptables : <ul style="list-style-type: none"> <li>- C'est faux.</li> <li>- C'est vrai, dans ce pays, les élèves entrent à l'école à 3 ans.</li> </ul>

Source : Développé par les auteurs du guide

## 2.3 Illustration par des exemples concrets

Dans cette section, nous proposons de revenir sur les thématiques développées dans le point précédent, en ancrant la réflexion dans deux domaines de contenu particuliers : la langue française et les mathématiques.

### 2.3.1 Évaluations en langue française

Lors d'un séminaire portant sur la conception et la validation d'épreuves d'évaluation sommative des acquis des élèves, les participants ont construit deux épreuves d'évaluation de la compréhension de l'écrit. Chacune de ces épreuves est accompagnée d'un cadre de référence, d'une table de spécification et d'un guide de correction.

 <b>Exemple 1 d'une évaluation en langue française</b> Gustave Flaubert, <i>Un cœur simple</i> .	
Cadre de référence et table de spécification	
Composante à évaluer	La compréhension de l'écrit
Niveau cible	3 <sup>ème</sup> année du collège
Période	Sixième
Contrôle	2
Objectifs visés	<ul style="list-style-type: none"><li>• Reconnaître des éléments qui caractérisent le texte</li><li>• Utiliser des informations en vue d'inférer</li><li>• Se positionner par rapport au texte</li></ul>
Les contenus à évaluer	<ul style="list-style-type: none"><li>• Eléments paratextuels</li><li>• Les personnages</li><li>• La narration</li><li>• Les sentiments</li></ul>
Support choisi	Texte littéraire (extrait d'un conte de Gustave Flaubert, <i>Un cœur simple</i> )
Les capacités évaluées : nombre et pourcentage d'items	Identifier      Nombre d'items : 2      (25%) Comprendre      Nombre d'items : 5      (62,5%) Réagir      Nombre d'items : 1      (12,5%)
Format des items et barème de notation	<ul style="list-style-type: none"><li>• Item 1 : relever les éléments paratextuels (QRC) 2 x 0,5 pt</li><li>• Item 2 : identifier le type de texte (QCM) 1 pt</li><li>• Item 3 : inférer (vrai/faux) 0,5 pt</li><li>• Item 4 : inférer (QRC) 0,5 pt</li><li>• Item 5 : analyser des contenus (QCM) 0,5 pt</li><li>• Item 6 : inférer (QCR) 0,5 pt</li><li>• Item 7 : analyser des contenus (QRC) 1 pt</li><li>• Item 8 : évaluer des contenus et argumenter (QCR) 1 pt</li></ul>

**Lis le texte et réponds aux questions qui suivent**

*Félicité, la servante\**, a reçu en cadeau un perroquet, et s'est prise d'une grande affection pour lui.

Elle l'avait posé sur l'herbe pour le rafraîchir, s'absenta une minute; et, quand elle revint, plus de perroquet! D'abord elle le chercha dans les buissons, au bord de l'eau et sur les toits, sans écouter sa maîtresse qui lui criait :

— Prenez donc garde ! vous êtes folle !

Ensuite elle inspecta tous les jardins de Pont l'Evêque ; et elle arrêtait les passants.

— Vous n'auriez pas vu (...) par hasard, mon perroquet ?

A ceux qui ne connaissaient pas le perroquet, elle en faisait la description. Tout à coup, elle crut distinguer derrière les moulins, au bas de la côte, une chose verte qui voltigeait. Mais au haut de la côte, rien ! Un porte-balle\* lui affirma qu'il l'avait rencontré tout à l'heure, à Sainte-Melaine, dans la boutique de la mère Simon. Elle y courut. On ne savait pas ce qu'elle voulait dire. Enfin elle rentra, épuisée, les savates en lambeaux, la mort dans l'âme ; et, assise au milieu du banc, près de Madame, elle racontait toutes ses démarches quand un poids léger lui tomba sur l'épaule : Loulou !

Gustave Flaubert, *Un cœur simple*, 1877

\*La servante : la bonne, la femme de ménage

\*Un porte-balle : un vendeur ambulancier

\*La mort dans l'âme : très triste

1. Complète le tableau suivant : (2 x 0,5 pt)

Titre de l'œuvre	Nom de l'auteur
.....	.....

2. À quel type appartient ce texte ? (1pt)

Coche la bonne réponse.

- Informatif
- Narratif
- Argumentatif
- Injonctif

3. Lis l'énoncé suivant et coche vrai ou faux. (0,5pt)

	Vrai	Faux
- Félicité a acheté un perroquet		

4. Pourquoi la maitresse a-t-elle dit à Félicité : « Prenez garde ! Vous êtes folle » ?

.....  
..... (0,5pt)

5. Qu'est ce qui a poussé Félicité à chercher son perroquet partout ?

Coche la bonne réponse. (0,5pt)

- Elle voulait l'offrir à sa maitresse.
- Elle voulait le soigner.
- Elle voulait le retrouver
- Elle voulait le vendre.

6. Comment les efforts de Félicité ont-ils été récompensés ?

.....  
.....(0,5 pt)

7. Le dernier mot du texte est Loulou. Qui est Loulou ?

.....  
.....(1 pt)

8. Si tu étais à la place de Félicité, aurais-tu fait la même chose ? Justifie ta réponse.

.....  
.....(1pt)

#### Guide de correction

Item 1. Un cœur simple (0, 5pt) / Gustave Flaubert (0,5pt)

Item 2. Narratif (1pt)

Item 3. Faux (0,5pt)

Item 4. Accepter toute réponse qui montre l'imprudence de Félicité. (chercher dans les buissons, au bord de l'eau, sur les toits...../ Félicité risque sa vie.....) (0,5pt)

Item 5. Elle voulait le retrouver (0,5 pt)

Item 6. Félicité a retrouvé son perroquet ou son perroquet est revenu / accepter toute réponse qui va dans ce sens. (0,5pt)

Item 7. Loulou est le perroquet de Félicité. (1 pt)

Item 8. Accepter toute réponse justifiée. (exemple : si j'étais à la place de Félicité, j'aurais fait la même chose parce que moi aussi j'aime les perroquets ou je serais triste si mon animal avait disparu....) (1pt)

Source : Élaboré lors d'un séminaire portant sur la conception et la validation d'épreuves d'évaluation sommative des acquis des élèves au Maroc



## Exemple 2 d'une évaluation en langue française

D'après Marcel Pagnol, Le château de ma mère.

### Cadre de référence et table de spécification

<b>Composante à évaluer</b>	Compréhension de l'écrit		
<b>Niveau cible</b>	3 <sup>ème</sup> année du cycle secondaire collégial		
<b>Période</b>	6		
<b>Contrôle</b>	1		
<b>Objectifs visés</b>	Vérifier la capacité de l'élève à mobiliser ses acquis des trois séquences du projet pour : <ul style="list-style-type: none"><li>• reconnaître des éléments caractéristiques du texte</li><li>• utiliser des informations en vue d'inférer</li><li>• se situer par rapport au texte</li></ul>		
<b>Les contenus</b>	<ul style="list-style-type: none"><li>• Paratexte</li><li>• Histoire/narration</li><li>• Narrateur/auteur/personnage</li><li>• Les sentiments</li></ul>		
<b>Support</b>	Texte de 120 mots (extrait d'un genre romanesque)		
<b>Capacités évaluées : nombre et pourcentage d'items</b>	Identifier	Nombre d'items : 2	(33%)
	Comprendre	Nombre d'items : 3	(50%)
	Réagir	Nombre d'items : 1	(17%)
<b>Format des items et barème de notation</b>	<ul style="list-style-type: none"><li>• Item 1 : relever les éléments paratextuels (QRC) 2 x 0,5 pt</li><li>• Item 2 : identifier le type de texte (QCM) 1 pt</li><li>• Item 3 : inférer (QCM) 1 pt</li><li>• Item 4 : inférer (vrai/faux) 2 x 0,5 pt</li><li>• Item 5 : repérer une information explicite (QRC) 2 x 0,5 pt</li><li>• Item 6 : évaluer le contenu et argumenter (QCR) 1 pt</li></ul>		

### Lis le texte et réponds aux questions qui suivent

Quand j'étais petit, parfois ma mère me déposait dans la classe de mon père l'instituteur. Je restais assis, bien sage au premier rang.

Un beau matin, ma mère me laissa à ma place et sortit sans rien dire pendant que mon père écrivait : « La maman a puni son petit garçon qui n'était pas sage ». Je criai : « Non, ce n'est pas vrai ! »

Mon père se retourna soudain, me regarda, surpris... Et pour vérifier si j'étais vraiment capable de lire, il me donna un livre de lecture. Je lus facilement plusieurs pages.

Je crois qu'il eut ce jour-là la plus grande joie, la plus grande fierté de sa vie.

D'après Marcel PAGNOL, Le château de ma mère.

## Questions

### Lecture : (6 points)

1. Réponds aux questions suivantes (0,5 point x2)

A.	Qui est l'auteur du texte ? .....
B.	De quelle œuvre ce texte est-il extrait ? .....

2. Coche la bonne réponse. (1 point)

Ce texte....

A.	décrit un lieu	
B.	raconte un évènement	
C.	fournit des informations	
D.	donne des conseils	

3. Coche la bonne réponse. (1 point)  
Où le père du narrateur exerce-t-il son travail ?

A.	à l'hôpital	
B.	à l'usine	
C.	à la ferme	
D.	à l'école	

4. Réponds par vrai ou faux à chacune des propositions suivantes :  
(0,5 point x2)

A.	Le père donne à son enfant un livre pour l'encourager.	.....
B.	Le père est étonné par la capacité de son fils à lire.	.....

5. Quels sentiments le père éprouve-t-il à la fin du texte ? (0,5 point x2)

.....  
.....

6. L'enfant n'a pas aimé la phrase écrite par son père au tableau. (1 point)

A. Êtes-vous d'accord avec la réaction de l'enfant ?

Oui	<input type="checkbox"/>
Non	<input type="checkbox"/>

B. Justifie ta réponse en une phrase.

.....

#### Guide de correction

Item	Réponses attendues	Barème de notation
1	a. Marcel PAGNOL b. Le château de ma mère.	0,5 point x2
2	Raconte un évènement	1 point
3	À l'école	1point
4	A. Le père donne à son enfant un livre pour l'encourager. (faux) B. Le père est étonné par la capacité de son fils à lire. (vrai)	0,5 point x2
5	Joie et fierté	0,5 point x2
6	<i>Peu importe que l'élève ait choisi oui ou non, c'est la cohérence entre son choix et sa justification qui compte. Accepter toute formulation équivalente. Ex.</i> <ul style="list-style-type: none"><li>- Oui, car l'enfant n'a pas été puni/car l'enfant était bien sage</li><li>- Non, car il n'est pas correct de crier en classe/ car ce n'est pas de lui qu'il s'agissait dans la phrase du père</li></ul>	1 point

Source : Élaboré lors d'un séminaire portant sur la conception et la validation d'épreuves d'évaluation sommative des acquis des élèves au Maroc

### 2.3.2 Évaluations en mathématiques

Les exemples suivants sont des productions de deux ateliers réunissant chacun 4 ou 5 inspecteurs de mathématiques (octobre 2022). La consigne était de se placer dans les conditions conformes aux directives pédagogiques tout en s'efforçant de prendre en compte les recommandations du guide.

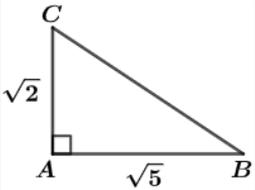
Si les évaluations suivantes, élaborées au cours de l'atelier, peuvent servir d'exemples authentiques adapté au contexte marocain, il convient de mentionner qu'elles ne sont pas totalement abouties aux yeux même des inspecteurs vu les contraintes de temps.



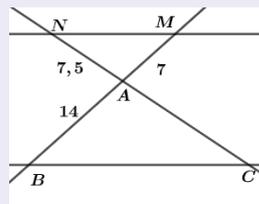
## Exemple 1 d'une évaluation en mathématiques

L'épreuve devait couvrir les points suivants du programme :

- Ordre et comparaison des nombres
- Théorème de Thalès

	Collège	Contrôle 2 Niveau : 3 <sup>e</sup>	Novembre 2022	
	Barème	L'usage de la calculatrice n'est pas autorisé		NIVEAU TAXONOMIQUE
Exercice 1 (2 points)	1 1	Soient : $a = \frac{6}{7}$ et $b = \frac{4}{5}$  1) Vérifier que : $a - b = \frac{2}{35}$  2) En déduire la comparaison de $a$ et $b$		1
Exercice 2 (4 points)	2 2	Soient : $a = \frac{1}{1+3\sqrt{2}}$ et $b = \frac{1}{1+2\sqrt{5}}$  1) Comparer $3\sqrt{2}$ et $2\sqrt{5}$  2) En déduire que $a > b$		2
Exercice 3 (2 points)	2	Montrer géométriquement que : $\sqrt{7} < \sqrt{5} + \sqrt{2}$  		3

Exercice 4 (2 points)	1	<p>On donne la figure ci-contre telle que :</p> <ul style="list-style-type: none"> <li>• Les droites (BC) et (NM) sont parallèles,</li> <li>• <math>AB=14</math> ; <math>AM=7</math> et <math>AN=7,5</math></li> </ul> <p>1) AC est égal à :</p> <p>A. 13 B. 14 C. 15 D. 16</p> <p>2) Justifier votre choix</p>	1
	1		



### Grille de correction

Exercice	Questions	Éléments de réponses	Barème
Exercice 1 (2 points)	1)	Toute méthode correcte sans faute de calcul	1
		Toute méthode correcte avec une faute de calcul	0,5
		Toute méthode correcte avec deux fautes de calcul	0,25
		Autre	0
	2)	Toute méthode correcte qui utilise la question 1	1
		Toute méthode correcte qui utilise la question 1 avec une faute.	0,5
Autre		0	
Exercice 2 (4 points)	1)	Toute méthode correcte sans faute	2
		Toute méthode correcte avec une faute de calcul	1,5
		Toute méthode correcte avec deux fautes de calcul	0,75
		Autre	0
	2)	Toute méthode correcte qui utilise la question 1	2
		Toute méthode correcte qui utilise la question 1 avec une étape fautive	1,5
		Toute méthode cohérente avec le résultat obtenu pour la question 1	1
		Autre	

Exercice 3 (2 points)		Utilisation de Pythagore pour démontrer que $BC = \sqrt{7}$ sans faute de calcul	1
		Utilisation de Pythagore pour démontrer que $BC = \sqrt{7}$ avec une faute de calcul	0,5
		Utilisation de l'inégalité triangulaire	1
Exercice 4 (3 points)	1)	C. $AC=15$	1
	2)	Toute méthode correcte sans faute de calcul	2
		Toute méthode correcte avec une faute de calcul	1,5
		Toute méthode correcte avec deux fautes de calcul	0,75
	Autre	0	

### Remarques

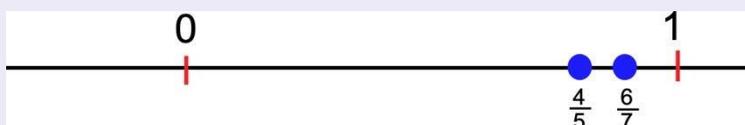
Les pratiques habituelles amènent à décomposer les tâches d'évaluation. On le remarque ici, tant dans l'exercice 1 que dans l'exercice 2. Cela conduit à abaisser le niveau taxonomique et rend plus difficile le respect de la table de spécification, laquelle suppose que 20% des items soient au niveau 3 de complexité.

Cela conduit aussi à limiter la validité de l'évaluation.

Ici, l'exercice 1 cherche à vérifier que les élèves savent comparer deux fractions. Le 1) leur impose une démarche et, simultanément, leur facilite la tâche. Mais, les élèves qui réussissent l'ensemble de l'exercice (1) et 2)), auraient-ils réussi le 2) tout seul ?

D'autre part, pourquoi imposer une démarche ?

Ici, l'élève peut très bien, en référence à la droite numérique, constater que  $a$  et  $b$  sont inférieurs à 1 et que  $a$  est plus proche de 1 que  $b$ , ce qui suffit pour conclure.



Les mêmes remarques peuvent être faites pour l'exercice 2.

La troisième question est de niveau de complexité 3. L'élève est un peu aidé par la figure qui doit l'orienter sur le théorème de Pythagore, mais la difficulté demeure. Il faut ensuite penser à l'inégalité triangulaire.

Cette question mériterait d'être testée avant d'être utilisée dans un contrôle.

Si la question est bien ou moyennement bien réussie lors du test, tout va bien. Si elle est mal ou très mal réussie, cela ne retire rien à son intérêt mais il faudra se demander ce qu'il faudrait faire lors des séquences d'enseignement pour qu'une telle question ait davantage de chance d'être réussie. Bien sur il ne s'agit pas de préparer les élèves en dévoilant la question avant le contrôle.

Source : Élaboré lors d'un séminaire portant sur la conception et la validation d'épreuves d'évaluation sommative des acquis des élèves au Maroc



## Exemple 2 d'une évaluation en mathématiques

### Exercice 1

- Calculer
- 1)  $2+3 \times 4$
  - 2)  $5 \times 3 + 1$

**Exercice 2** Dans les questions suivantes, une seule réponse est correcte.

Cocher la réponse correcte :

- 1)  $7 ( 1,3 + 3,7 )$  est égal à :
  - 28,7
  - 35
  - 12,8
- 2)  $3,1 \times 2,6 + 3,1 \times 7,4$ . est égal à :
  - 31
  - 15,46
  - 82,584
- 3) J'ai 20 DH. J'achète trois stylos à 2 DH chaque et un cahier de 48 pages à 8,5 DH. Il me reste :
  - 6,5 DH
  - 5,5 DH
  - 13,5 DH

**Exercice 3 Cocher les bonnes réponses.**

- 4) Le nombre  $\frac{6}{4}$  peut s'écrire :
  - 1,5
  - $6 - 4$
  - $\frac{9}{6}$
  - $\frac{(6+3)}{(4+3)}$
- 5) Soient les nombres  $a = \frac{1}{2}$  ;  $b = \frac{3}{4}$  et  $c = 5$ .  
 $a \times b + c$ . est égal à :
  - 1
  - $\frac{43}{8}$
  - $\frac{8}{9}$
  - $\frac{86}{16}$

### Exercice 4

La capacité d'un réservoir est 60 l.

On remplit les  $\frac{3}{4}$  de ce réservoir.

Calculer la capacité de la partie vide du réservoir.

### Exercice 5

Relie chaque opération à son résultat.

$$\frac{1}{2} + \frac{1}{3}$$

$$\frac{3}{5} \times \frac{2}{3}$$

$$\frac{12}{7} - \frac{5}{7}$$

•  $1$

•  $\frac{1}{5}$

•  $\frac{5}{6}$

•  $\frac{2}{5}$

### Barème

Exercice 1	1)	1,5 pts
	2)	1,5 pts
Exercice 2	1)	1,5 pts
	2)	2 pts
	3)	2 pts
Exercice 3	1)	2 pts
	2)	3,5 pts
Exercice 4		2 pts
Exercice 5		3,5 pts

### Remarques

Les exercices 1 et 2 sont composés de questions à choix multiples.

Il est fortement recommandé de proposer au moins 4 réponses dans chaque cas.

D'autre part, le fait de placer dans une même épreuve des QCM pour lesquelles une seule réponse exacte est proposée et d'autres pour lesquelles il peut y voir plusieurs réponses exactes peut être source de confusion pour les élèves.

Source : Élaboré lors d'un séminaire portant sur la conception et la validation d'épreuves d'évaluation sommative des acquis des élèves au Maroc

## 2.4 Notes

---

<sup>1</sup> Examen normalisé régional pour l'obtention du certificat du cycle collégial – Cadre de référence de la langue Française. Document transmis par le Centre national de l'évaluation et des examens (CNEE), non daté, non numéroté.

<sup>2</sup> Les notes ministérielles consultées par l'équipe éditoriale en langue française utilisent le terme de capacités là où dans d'autres curriculums on parlerait de compétences. Comme le dit l'étymologie du terme, les capacités comme les compétences incluent les savoir-faire.

<sup>3</sup> Les distracteurs dans les QCM sont les propositions de réponses plausibles autres que la réponse correcte.

<sup>4</sup> Pour plus de détails, voir section 5, La correction des épreuves.

<sup>5</sup> Dans les épreuves standardisées, on utilise des techniques statistiques pour évaluer la qualité des distracteurs (r.bis). Voir l'Annexe B pour aller plus loin.

1

Introduction

2

Les grandes étapes de l'élaboration des épreuves

Section

3

# La passation des épreuves

**3.1** Développer les conditions de passation d'une évaluation sommative

**3.2** Développer une feuille de route type

**3.3** Note

4

La correction des épreuves

5

La communication et l'utilisation des résultats

6

La communication et l'utilisation des résultats



La passation, c'est la période pendant laquelle les élèves sont soumis à l'épreuve. La passation suppose une préparation avant et un accompagnement pendant, qui peut consister en une simple surveillance ou en des activités plus complexes. La fiabilité autant que l'équité de l'évaluation dépendent en partie du sérieux avec lequel cette opération est réalisée.

« Si l'on veut mettre les élèves sur un pied d'égalité par rapport aux contrôles continus, il convient d'unifier autant que possible la passation, au moins au sein d'un établissement scolaire, voire dans des entités plus larges. »

Si l'on veut mettre les élèves sur un pied d'égalité par rapport aux contrôles continus, il convient d'unifier autant que possible la passation, au moins au sein d'un établissement scolaire, voire dans des entités plus larges.

L'organisation concrète de la passation des contrôles continus pose divers problèmes qui concernent en particulier le secret gardé sur les épreuves avant la passation, l'information préalable des élèves, le moment et la durée de la passation, les éventuels regroupements de classes, les mesures prises pour éviter le copiage ou la fraude.

L'organisation concrète de la passation des contrôles continus pose divers

Alors que l'évaluation formative ne pose pas de problème en ce qui concerne la sécurité et le copiage (les enjeux pour l'élève ne sont en effet pas de même nature que dans le cas de l'évaluation sommative), ces questions sont problématiques pour l'évaluation sommative.

## 3.1 Développer les conditions de passation d'une évaluation sommative

L'organisation de la passation dépend des possibilités de l'établissement et se fait sous la responsabilité du chef d'établissement, en collaboration avec l'inspecteur pédagogique et/ou les conseils au niveau des établissements. L'organisation de la passation comprend plusieurs aspects :

- **Mettre en place les conditions logistiques pour la passation du contrôle continu**

Les conditions de passation seront conçues de manière à permettre aux élèves de faire de leur mieux et à garantir l'intégrité de la passation des tests (c'est-à-dire minimiser le risque de fraude ou copiage). Dans le contexte des contrôles continus, les élèves passeront les épreuves dans un endroit sécurisé (par exemple, dans leurs classes, chacun depuis son propre bureau, avec un écart suffisant entre les bureaux), avec, dans la mesure du possible, un élève par table. Pour cela, la passation peut se dérouler dans un autre lieu que dans la salle de classe si elle n'est pas suffisamment équipée.

Dans le cas d'une épreuve commune à plusieurs classes, ces dernières devront passer les épreuves en même temps. Pour les contrôles continus élaborés de manière collective (avec d'autres enseignants d'un même établissement ou bien des enseignants de différents établissements), plusieurs classes peuvent être regroupées en un même lieu. Les établissements publics de grande taille peuvent collaborer avec les établissements privés et

publics de petite taille pour organiser et accueillir la tenue des épreuves, au cas où ces derniers ne disposeraient pas des ressources nécessaires

Dans ces cas, et selon la taille de l'établissement, du nombre de classes et d'élèves, les épreuves peuvent même être administrées par groupes tout en veillant à ne pas déranger et retarder la tenue des enseignements.

- **Informar les élèves de la date de passation suffisamment à l'avance**

Il est préférable de prévenir les élèves au moins une semaine à l'avance de la date et des horaires de passation des épreuves de contrôle continu (ou par un calendrier trimestriel ou annuel) et les informer du matériel dont ils devront disposer : dictionnaires, documents, instruments de géométrie, calculatrices. Il appartient aux enseignants de veiller à ce que le matériel admis au test ne procure pas un avantage injuste aux élèves les mieux équipés<sup>1</sup>.

- **Assurer une supervision adéquate pendant le déroulement du test**

Pendant la passation, le ou les enseignants assurent la surveillance sans se laisser distraire par d'autres activités. Ils répondent aux questions avec empathie, rassurent les élèves qui seraient anxieux, mais n'aident pas les élèves à traiter les questions.

Ils peuvent profiter de cette surveillance pour observer le comportement des élèves et prendre des informations qu'ils pourront utiliser au moment du retour des résultats. La façon dont l'élève s'implique dans une tâche est une information importante pour organiser les séances de soutien.

Les enseignants avertissent les élèves du temps qu'il leur reste, par exemple 15 minutes, puis 5 minutes avant la fin du temps écoulé, puis, lorsque le temps est écoulé, les enseignants ramassent les copies en vérifiant qu'elles sont toutes remises.

- **Définir les mesures prises en cas de retard, d'absence ou d'achèvement anticipé du test**

Il conviendra de définir au niveau de l'établissement :

- Ce qui se passe pour un élève qui arrive en retard : autoriser un retard justifié de 10 minutes maximum ? avec ou sans prolongation du temps ?
- Ce que l'on fait pour un élève absent à l'épreuve.
  - Si l'absence est justifiée : lui permettre de passer ultérieurement, sous surveillance, une forme modifiée de l'épreuve ?
  - Si l'absence n'est pas justifiée : sanctionner son absence et le contraindre à passer ultérieurement, sous surveillance, une forme modifiée de l'épreuve ?
- Ce qui se passe lorsqu'un élève a terminé l'épreuve avant la fin du temps alloué.
  - En premier lieu, lui demander de revoir son travail, puis, éventuellement :

- Permettre à l'élève d'effectuer une autre activité scolaire sans déranger les autres ?
- Autoriser l'élève à sortir ?

- **Mettre en place des dispositifs pour les élèves à besoins éducatifs particuliers**

Les élèves à besoins éducatifs particuliers peuvent aussi bénéficier de dispositions particulières. Ils peuvent, selon les cas, disposer de plus de temps ou être assistés par un personnel spécialisé (cas d'élèves malvoyants ou malentendants).

L'objectif des dispositions particulières est d'assurer l'égalité des chances pour tous les élèves et pour qu'aucun élève ne soit désavantagé par une caractéristique des exigences de l'épreuve qui ne soit pas liée aux concepts évalués par l'épreuve (les compétences, connaissances ou savoir-faire évalués). Toutefois, il faut veiller à ne pas compromettre l'évaluation en introduisant des dispositions d'accès inappropriées.

**Graphique 3.1 Les différents aspects de l'organisation de la passation**



Source : Développé par les auteurs du guide

## 3.2 Développer une feuille de route type

Comme cela se fait pour les évaluations externes standardisées dans d'autres pays, il est fortement recommandé d'élaborer une feuille de route type pour les personnes (enseignants dans ce cas) qui font passer

le test pour assurer que les conditions de passation sont unifiées autant que possible.

« ...il est fortement recommandé d'élaborer une feuille de route type pour les personnes (enseignants dans ce cas) qui font passer le test pour assurer que les conditions de passation sont unifiées autant que possible. »

Notamment, la feuille de route précisera les aspects suivants :

- La date, l'heure et l'emplacement de la passation du test.
- La durée de la passation : cela devra être réfléchi de manière à ce que le test soit aisément réalisable par les élèves dans le temps imparti. En mathématiques par exemple, un test d'une heure comportera environ 12 questions déclinées en une trentaine d'items environ.
- Les ressources dont les élèves peuvent se servir (dictionnaire, calculatrice).
- Les instructions que l'enseignant doit/peut fournir aux élèves.
- Les pauses autorisées.
- Les mesures prises en cas de retard, d'absence ou d'autres situations.

Graphique 3.2 Éléments d'une feuille de route type



Source : Développé par les auteurs du guide

Nous proposons ci-dessous deux exemples de feuille de route en français et en mathématiques pour des évaluations externes. Il convient de mentionner que tous les points inclus dans ces exemples ne s'appliqueront pas au contrôle continu au Maroc, mais ils visent à servir d'exemple dont les enseignants peuvent s'inspirer pour développer leur propre feuille de route adaptée.



### Exemple 1 d'une feuille de route pour la passation

L'épreuve de français se déroule dans tous les établissements scolaires durant la matinée du vendredi 14 juin 2019. L'épreuve doit commencer entre 8 h 15 et 8 h 45. Sa durée est de 200 minutes effectives.

Dès 7 h du matin, le jour de la passation, l'épreuve peut être répartie entre les enseignants des classes concernées.

Première partie : 2 x 50 minutes (livret 1) Pause

Deuxième partie : 2 x 50 minutes (livret 2)

### **Pour l'ensemble de l'épreuve**

Les référentiels habituellement utilisés par l'élève (manuels, fiches, dictionnaires...) sont laissés à sa disposition.

#### **Première partie (livret 1)**

##### *Compréhension du dossier informatif*

Le professeur distribue le portefeuille de documents (dossier informatif et récit de fiction) et le livret 1.

L'élève répond aux questions concernant la lecture du dossier informatif.

##### *Compréhension du récit de fiction*

L'élève répond aux questions concernant la lecture du récit de fiction.

Le professeur reprend le livret 1, le portefeuille de documents reste à la disposition de l'élève durant la deuxième partie de l'épreuve.

#### **Deuxième partie (livret 2)**

##### *Écoute d'une interview*

Le professeur explique oralement à l'élève les modalités de passation de l'écoute et s'assure que celui-ci a pris connaissance des questions.

Il invite l'élève à écouter une première fois l'interview (5 min 35 sec).

Il invite l'élève à écouter une deuxième fois l'interview (5 min 35 sec) et à répondre aux questions.

Remarque : l'élève utilise toutes les stratégies qui lui permettent de répondre aux questions (prise de notes, mémorisation...) au moment où il le souhaite.

##### *Outils liés à la maîtrise de la langue*

L'élève répond aux questions concernant les outils liés à la maîtrise de la langue.

##### *Tâche d'écriture*

L'élève réalise la tâche d'écriture.

Le professeur reprend le portefeuille de documents, le livret 2 et le brouillon éventuel.

Source : (Fédération Wallonie-Bruxelles, s.d.[15])



## Exemple 2 d'une feuille de route pour la passation

Cette année l'évaluation externe non certificative se déroulera entre le **04 et le 08 octobre 2021**.

En 5e année primaire, il s'agit de programmer trois séances de test de 50 minutes. Elles ne peuvent être consécutives : elles seront séparées au minimum par une récréation ou, idéalement, ne se dérouleront pas le même jour. Les différentes parties seront présentées aux élèves dans l'ordre du carnet.

### **PARTIE 1**

50 minutes maximum

Test de fluidité + Items 1 à 26

### **PARTIE 2**

50 minutes maximum

Items 27 à 53

### **PARTIE 3**

50 minutes maximum

Tâches d'écriture : E1 à E24

Il est légitime que chaque enseignant ait le souci de respecter les différences entre ses élèves. Toutefois, pour que la validité de l'épreuve soit assurée, il est impératif que la passation se déroule dans les mêmes conditions pour toutes les classes. Les consignes de passation permettent de garantir la pertinence des informations recueillies par ces évaluations.

Même si les élèves passent l'épreuve en compagnie de leur enseignant attiré, la présentation du carnet ne leur est peut-être pas familière ; il semble opportun de les informer et de les rassurer quant à cette passation.

Quelques jours avant le début de l'évaluation, décrivez-leur ce qu'est une évaluation externe non certificative en leur expliquant, par exemple, que :

- Tous les élèves de leur année, de tous les établissements scolaires de la Fédération Wallonie-Bruxelles (environ 50 000 enfants) participent à ce test durant la même semaine.
- Cette épreuve n'aura aucune conséquence sur les résultats indiqués dans le bulletin, mais il est toutefois important d'effectuer le travail le plus sérieusement possible, car elle a pour objectif de cerner les difficultés rencontrées par les élèves afin de chercher différents moyens d'y remédier.

Aucun document ne sera remis aux élèves avant la passation. Le carnet et le portfolio de chaque élève seront repris en fin de séance et redistribués lors de la suivante pour poursuivre l'épreuve.

Le premier jour de la passation, il est demandé aux élèves de compléter le cadre de la couverture du carnet : nom, prénom, classe, numéro d'ordre et établissement scolaire. Il s'avère utile de vérifier ces informations lors de la reprise des documents.

Avant chaque séance, l'enseignant repère, avec les élèves, les pages relatives à la partie considérée. Lors de cette découverte, il est utile de leur expliquer que les séries de cases carrées à droite des questions seront utilisées pour la correction et qu'ils ne doivent donc pas s'en préoccuper.

Le carnet a été conçu de façon à ce que chaque élève puisse travailler seul. Les élèves n'ont donc, en principe, pas besoin d'explications supplémentaires.

### **Consignes spécifiques à l'épreuve de 5e année de l'enseignement primaire**

Le test de fluidité de lecture (début de partie 71) doit être rigoureusement chronométré (3 minutes). Assurez-vous de disposer d'un chronomètre ou de cette fonction sur smartphone.

Pour ce test de fluidité, découvrez les consignes avec les élèves avant de donner le signal de départ.

Précisez aux élèves que les textes à consulter sont renseignés au début de chaque partie de l'épreuve. Insistez sur l'importance d'en prendre connaissance consciencieusement pour répondre aux questions.

Pour les questions à choix multiples, rappelez qu'une seule réponse est à cocher.

Le vocabulaire du carnet a été soigneusement choisi et les concepteurs de l'épreuve considèrent que ce vocabulaire appartient au lexique de la discipline. Quant au vocabulaire des textes, il est inhérent aux différents genres littéraires considérés. Dans un souci de standardisation des conditions de passation, il est demandé de ne pas clarifier des mots incompris.

L'utilisation de dictionnaires ou d'autres référentiels n'est pas autorisée pendant la réalisation de l'épreuve, ceci à nouveau afin de garantir des conditions de passation identiques pour tous les élèves. C'est la raison pour laquelle l'orthographe ne sera pas prise en compte dans les tâches d'écriture puisque les Socles de compétences prescrivent des niveaux de maîtrise de l'orthographe « en ayant recours à des référentiels d'orthographe d'usage et grammaticale » (Rondal et Broonen, 2021<sup>[16]</sup>).

Source : (Fédération Wallonie-Bruxelles, s.d.<sup>[13]</sup>)

## **3.3 Notes**

---

<sup>1</sup> Par exemple, les calculatrices modernes peuvent être de véritables encyclopédies et certaines permettent la communication à distance. La question des calculatrices ne concerne donc pas que les disciplines scientifiques. Dans certains cas, en mathématiques en particulier, on peut souhaiter l'utilisation des calculatrices (la compétence à tirer profit de cet instrument et d'autres ressources informatiques est l'un des objectifs de l'enseignement des mathématiques). Dans ce cas, les épreuves devront être élaborées en conséquence et, en particulier, ne pas donner un avantage aux élèves les mieux équipés. De plus, le matériel autorisé pourra être limité à certains modèles. Dans le cas d'une passation sur ordinateur, l'épreuve est en général accompagnée d'une calculatrice virtuelle dont les fonctions sont limitées à l'usage qu'il convient d'en faire.

1

Introduction

2

Les grandes étapes de l'élaboration des épreuves

3

La passation des épreuves

Section

4

## La correction des épreuves

**4.1** Assurer une correction fidèle : grilles de correction et de codage

**4.2** Corriger de manière informative : l'annotation des copies

**4.3** Assurer une notation cohérente

**4.4** Principaux phénomènes menaçant l'objectivité et la validité de l'évaluation

**4.5** Note

5

La communication et l'utilisation des résultats

6

Les biais liés au genre et à d'autres caractéristiques des élèves ou de la classe

La correction consiste à apprécier, annoter et noter les productions réalisées par les élèves au cours de la passation de l'épreuve. Les annotations sont qualitatives ; la notation quantitative est analogue à la mesure chiffrée de la performance de l'élève. La correction de l'épreuve est également l'action consistant à interpréter l'activité de l'élève pour lui attribuer une valeur et en tirer des informations utiles aux différents acteurs (élèves, enseignants, administration, parents, chacun selon ses besoins).

## 4.1 Assurer une correction fidèle : grilles de correction et de codage

Pour une correction fidèle, il faut prévoir dès le départ des grilles de correction précises, particulièrement pour les réponses ouvertes construites, qui limitent l'interprétation et la subjectivité des corrections et spécifient quel code ou quel point est attribué à quel type de réponse.

« Pour une correction fidèle, il faut prévoir dès le départ des grilles de correction précises, particulièrement pour les réponses ouvertes construites, qui limitent l'interprétation et la subjectivité des corrections. »

Sans cela chaque correcteur y va de son interprétation et une même réponse d'élève peut recevoir des appréciations ou des points très différents selon le correcteur. Même pour des matières réputées plus objectives, comme les mathématiques et les sciences, la fidélité entre correcteurs est un réel souci et pose en conséquence des problèmes en matière d'équité.

Si une même réponse reçoit des notes plus ou moins sévères selon le correcteur, les élèves ne sont clairement pas sur un pied d'égalité, leurs chances de réussite sont inégales et c'est injuste dans la mesure où celles-ci ne dépendent pas que de leurs acquis. Cet aspect est davantage crucial pour les années d'études où le contrôle continu intervient dans la certification (par exemple, en troisième année de collège).

En plus, comme mentionné au-dessus (voir l'encadré 1.3 « Évaluation normative et évaluation critériée »), une évaluation critériée permet de situer chaque élève de la classe par rapport aux objectifs visés et enseignés, et permet ainsi aux enseignants de recueillir des informations sur l'apprentissage des élèves.

Comme brièvement décrit précédemment, lorsqu'on élabore des critères de correction, il est utile de cerner très précisément les capacités que l'on souhaite évaluer. Il convient :

- D'identifier la ou les réponses prototypes correspondant au niveau d'exigence attendu ainsi que les limites tolérées par rapport à ces réponses prototypes.
- De rendre les critères explicites, envisager tous les cas de figure et illustrer les réponses à accepter ou à rejeter.

Il faut préciser que faire :

- En cas de partie de réponse non pertinente ou contradictoire
- En présence d'erreurs orthographiques ou grammaticales
- Si les élèves n'indiquent pas les unités (en mathématiques par exemple)
- S'ils ne formulent pas leur réponse au sein d'une phrase correcte
- En présence de réponses partiellement correctes.

Toutes ces questions doivent trouver réponse tout en gardant à l'esprit que la multiplication des critères et des codes de correction rend le travail des correcteurs plus ardu, plus long et potentiellement moins fiable.



#### Encadré 4.1

##### Bonne pratique – le prétest

Pour en revenir au guide de correction, il est toujours utile de tester les questions sur un échantillon même limité d'élèves non concernés par le contrôle (élèves de l'année supérieure ou inférieure par exemple), en particulier pour les questions ouvertes (à travers une étude pilote). Cela permet ainsi de recueillir des réponses d'élèves authentiques qui peuvent nourrir les grilles de correction. Une élaboration a priori par un adulte ne permet pas d'y arriver de manière efficace. Ce n'est sans doute pas possible de le faire pour tous les contrôles continus, mais il pourrait être intéressant d'effectuer cette opération de prétest dans l'année d'études où le contrôle continu intervient pour la certification (en 3<sup>e</sup> année de collège).

Source : Développé par les auteurs du guide

Nous proposons ci-dessous un exemple de grille de correction tiré de l'évaluation internationale PISA (Programme international pour le suivi des acquis des élèves).



## Exemple 1 de grille de correction

Question PISA libérée

### Question

#### Question 4 : ÉNERGIE ÉOLIENNE

PM922Q04 – 0 1 2 9

Quelle est la vitesse maximale à laquelle les extrémités des pales de l'éolienne peuvent tourner ? Décrivez votre raisonnement et donnez le résultat en **kilomètre par heure (km/h)**. Référez-vous aux informations fournies au sujet du modèle E-82.



### ÉNERGIE ÉOLIENNE

Zedville envisage de construire plusieurs éoliennes pour produire de l'électricité.

La mairie de Zedville a recueilli des informations sur le modèle suivant.

Modèle :	E-82
Hauteur du mât :	138 mètres
Nombres de pales :	3
Longueur d'une pale :	40 mètres
Vitesse maximale de rotation :	20 rotations par minute
Coût de construction :	3 200 000 zeds
Production :	0,10 zed par kWh généré
Frais d'entretien :	0,01 zed par kWh généré
Efficacité :	Opérationnel 97% de l'année

Remarque : Le kilowatt/heure (kWh) est une unité de mesure de l'énergie électrique.

## Critères de correction

### ÉNERGIE ÉOLIENNE : CONSIGNES DE CORRECTION Q 4

#### OBJECTIF DE LA QUESTION :

Description : Utiliser une modélisation en plusieurs étapes pour résoudre un problème dans un contexte cinétique

Domaine mathématique : Variations et relations

Contexte : Scientifique

Processus : Employer

#### ***Crédit complet***

Code 2 : Résultat exact déduit d'un raisonnement correct, complet et compréhensible. Le résultat doit être donné en km/h. Un schéma n'est pas obligatoire, tout comme une phrase séparée contenant la réponse.

- La vitesse maximale de rotation est de 20 rotations par minute ; la distance par rotation est  $2 \times \pi \times 40 \text{ m} \approx 250 \text{ m}$ , soit  
 $20 \times 250 \text{ m/min} \approx 5\,000 \text{ m/min} \approx 83 \text{ m/s} \approx 300 \text{ km/h}$ .

#### ***Crédit partiel***

Code 1 : Résultat exact déduit d'un raisonnement correct, complet et compréhensible, mais sans être exprimé en km/h. Ici encore, un schéma n'est pas obligatoire, tout comme une phrase séparée contenant la réponse.

- La vitesse maximale de rotation est de 20 rotations par minute ; la distance par rotation est  $2 \times \pi \times 40 \text{ m} \approx 250 \text{ m}$ , donc  
 $20 \times 250 \text{ m/min} \approx 5\,000 \text{ m/min} \approx 83 \text{ m/s}$ .

#### ***Pas de crédit***

Code 0 : Autres réponses.

Code 9 : Omission.

Source : (OCDE, 2013<sub>(17)</sub>)

## 4.2 Corriger de manière informative : l'annotation des copies

Si on veut que les contrôles continus servent a minima à informer les élèves sur leurs acquis et leur marge de progrès, il convient de ne pas se limiter aux notes, mais de « faire parler la note » en annotant les copies, ce qui constitue un premier pas vers un usage formatif/informatif et pas uniquement sommatif/certificatif des contrôles continus.

Les annotations présentes sur les épreuves lors de la correction permettront aux élèves de **comprendre la notation, d'avoir un aperçu de leurs erreurs et de leurs réussites, et donc de connaître les points à améliorer ou à valoriser** <sup>1</sup>.

Les annotations, pour l'essentiel, seront en rapport avec les connaissances et les compétences que l'épreuve est censée contrôler et ne seront pas limitées aux classiques « juste », « faux », « mal compris », ou autres, peu porteuses d'information. Elles doivent repérer dans les productions des élèves les éléments significatifs par rapport aux attentes de l'épreuve, par exemple, les éléments significatifs de réussite, des difficultés rencontrées par l'élève, d'incompréhensions ou d'échecs, les fausses pistes suivies par l'élève, ses connaissances apparemment manquantes, etc.

Les annotations sont portées directement sur la copie. Des codages simples (partagés avec les élèves à l'avance) peuvent être utilisés pour les annotations les plus fréquentes (par exemple, en français, FO = Faute d'Orthographe).

Deux illustrations commentées figurent dans les pages suivantes.



### Exemple 1 d'annotations sur une épreuve

Première illustration

2- Le corps d'un homme de 70 kg, contient environ 6 litres de sang. 2/4

Déterminer, en notation scientifique, le nombre de plaquettes contenues dans le corps d'un homme de 70 kg sachant qu'un millimètre cube de sang compte environ 250 000 plaquettes.

Indiquer toutes les étapes du raisonnement.

Pour info : 1l occupe un volume de 1dm<sup>3</sup>.

70kg → 6L de sang | 1dm<sup>3</sup> = 1l

250 000 · 6 = 1 500 000 plaquettes pour 6mm<sup>3</sup>

6mm<sup>3</sup> donc 6dm<sup>3</sup> = 6L donc 1 500 000 plaquettes est égal à 1,5 · 10<sup>6</sup> plaquettes pour 6L.

1,5 · 10<sup>6</sup> = nombre de plaquette pour 6L de sang

1dm<sup>3</sup> = 1 000 000 mm<sup>3</sup>.

Niveau visé : 2<sup>e</sup> Année Collégiale

Source : Exemple tiré d'une production étudiante

Dans son commentaire, l'enseignant montre assez clairement à l'élève où se situe le problème : bien que tout à fait pertinente, la démarche de l'élève n'a pas tenu compte du fait que 250 000 était la quantité de plaquettes présente dans 1 mm<sup>3</sup> et non 1 dm<sup>3</sup>.

L'élève peut donc comprendre assez rapidement son erreur.



## Exemple 2 d'annotations sur une épreuve

Deuxième illustration

Résoudre l'inéquation suivante et représenter l'ensemble des solutions sur une droite graduée :

$$\frac{3x}{2} - \frac{2x-1}{3} > \frac{2x-6}{4}$$

*mettre la fraction négative*

*Que fais-tu ? Il faut respecter la réduction au même dénominateur.*

$$\frac{18x - 4(8x-4)}{12} > \frac{6x-18}{12}$$

$$18x - 32x + 16 > 6x - 18$$

$$18x - 32x = 6x - 16 - 18$$

$$-20x > -34$$

$$x > \frac{34}{-20}$$

$$x < \frac{34}{20}$$

$$x < \frac{17}{10}$$

$$1/8$$

Niveau visé : 2<sup>e</sup> Année Collégiale

Source : Exemple tiré d'une production étudiante

Dans cette copie, l'enseignant identifie une première erreur, due à la réduction au même dénominateur : l'élève semble avoir multiplié le numérateur de la fraction par 4, mais également par 2 ce qui n'a pas de sens dans ce cas.

Par la suite, bien que tout à fait correcte, la démarche n'a pas été prise en compte par l'enseignant. Un commentaire supplémentaire aurait peut-être été utile afin d'expliquer à l'élève le sens des traits sur la suite de la copie.

Quoi qu'il en soit, c'est bien la réduction au même dénominateur que l'élève doit revoir et ceci est clairement mis en évidence par l'enseignant.

Si, dans ces deux exemples, l'enseignant stipule clairement les éléments qu'il appartient à l'élève d'approfondir, aucune annotation ne concerne les aspects que l'élève maîtrise visiblement dans ces deux situations. Il pourrait être intéressant de compléter les commentaires, en soulignant ce que l'élève parvient à réaliser également, pour éviter de nuire à la motivation de l'élève.

## 4.3 Assurer une notation cohérente

Pour assurer la fidélité des CC, il est nécessaire que la note chiffrée ait le même sens pour tous les enseignants et pour tous les utilisateurs. Pour cela, des descripteurs de performance de l'apprentissage attendu des élèves sur l'échelle de notation existante sur 20 pourraient être fournis (voir l'exemple ci-dessous). Ceux-ci doivent être complétés par des normes d'évaluation, qui précisent les critères permettant de déterminer le niveau de maîtrise des acquis de l'apprentissage, ainsi que par d'autres ressources pour les enseignants, telles que des guides et des exemples de copies corrigées et annotées des élèves<sup>2</sup>.

Au Maroc, les enseignants sont responsables de la correction des épreuves du CC. Le barème établi au moment de l'élaboration de l'épreuve est appliqué. Les productions des élèves pour le CC sont notées sur une échelle de 0 à 20. Cette note ne sera une information fiable que si l'échelle de notation a un sens univoque pour tous les enseignants et tous les utilisateurs. De ce fait, les enseignants d'un établissement ou d'une région peuvent se mettre d'accord sur une échelle de descripteurs (voir l'exemple ci-dessous).

La note des CC pourrait permettre de replacer les acquis de l'élève par rapport à l'ensemble des objectifs contrôlés. En associant à chaque épreuve la liste des objectifs ou des compétences contrôlées, l'enseignant pourrait identifier non seulement les objectifs atteints mais aussi ceux qui posent un problème.



### Exemple de descripteurs de performance

Note chiffrée	Niveau d'acquisition des connaissances, savoir-faire et compétences
18 à 20	L'élève a montré une excellente maîtrise de l'ensemble des compétences et des connaissances contrôlées par ce CC.
15 à 18	L'élève a montré une très bonne maîtrise des compétences et des connaissances contrôlées par ce CC.
13 à 15	L'élève a montré une assez bonne maîtrise des compétences et des connaissances contrôlées par ce CC. Plusieurs points doivent être retravaillés.
11 à 13	L'ensemble est satisfaisant. Cependant, une partie des compétences et des connaissances contrôlées par ce CC ne semblent pas acquises. L'élève doit revenir sur leur apprentissage.

9 à 11	L'ensemble est à peine passable. L'élève ne maîtrise partiellement qu'une partie des compétences et des connaissances contrôlées par ce CC. Un travail d'approfondissement est nécessaire.
7 à 9	L'ensemble est insuffisant. L'élève ne maîtrise partiellement qu'une très faible partie des compétences et des connaissances contrôlées par ce CC. Beaucoup d'apprentissages doivent être repris.
5 à 7	L'ensemble est très insuffisant. Une très faible partie des compétences et des connaissances contrôlées par ce CC semblent partiellement acquises. Beaucoup d'apprentissages doivent être repris. Des activités de soutien sont nécessaires.
< 5	Les lacunes de l'élève sont importantes. Une prise en charge particulière de l'élève s'impose (entretiens et soutien).

Source : Développé par les auteurs du guide

La grille proposée n'est qu'un exemple. Elle signifie seulement la nécessité d'un codage aussi univoque que possible, nécessaire à la fiabilité recherchée.

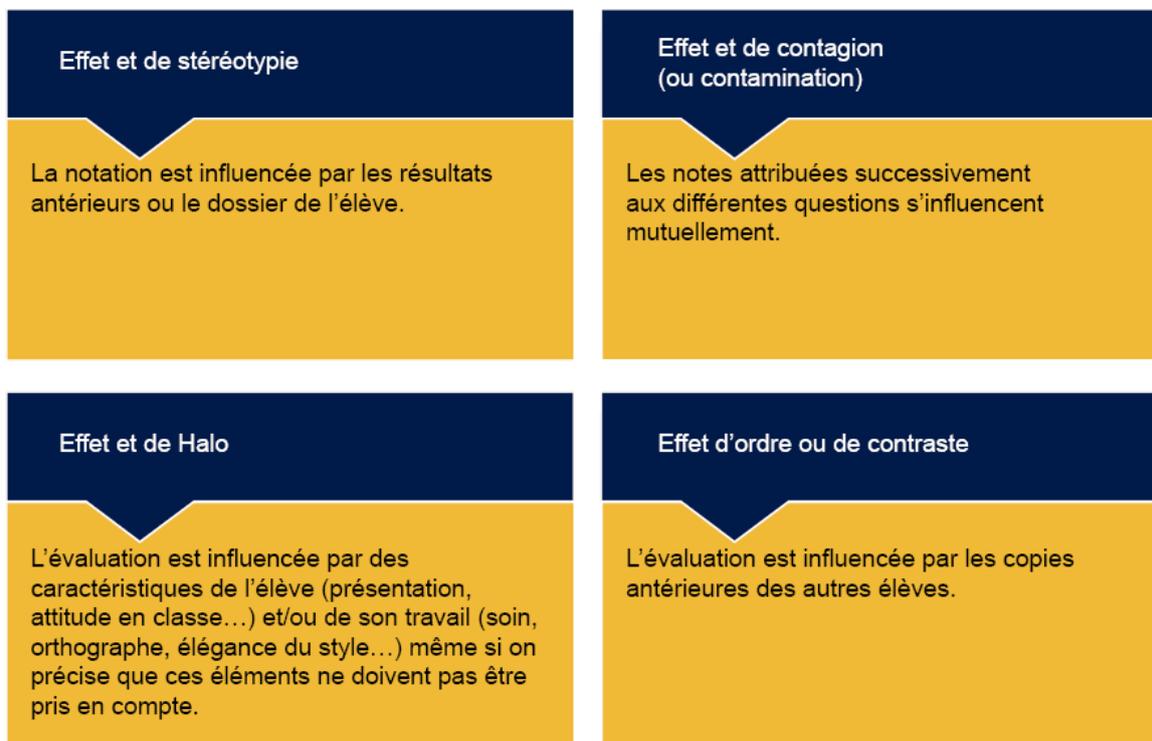
## 4.4 Principaux phénomènes menaçant l'objectivité et la validité de l'évaluation

Même avec une grille de correction présentant des critères et des indicateurs, certains biais peuvent affecter l'objectivité de la correction.

La recherche en docimologie a mis en évidence les principaux phénomènes menaçant l'objectivité et la validité de l'évaluation ; ceux-ci sont des menaces pour l'équité de l'évaluation. Voici les principaux biais :

- **Effet de stéréotypie** : la notation est influencée par les résultats antérieurs ou le dossier de l'élève.
- **Effet de halo** : l'évaluation est influencée par des caractéristiques de l'élève (présentation, attitude en classe...) et/ou de son travail (soin, orthographe, élégance du style...) même si on précise que ces éléments ne doivent pas être pris en compte.
- **Effet de contagion (ou contamination)** : les notes attribuées successivement aux différentes questions s'influencent mutuellement. Un élève qui aura très bien répondu aux quelques premières questions d'un test verra son résultat surestimé pour une question suivante qui n'est que moyennement (ou pas du tout) réussie, contrairement à l'élève qui aurait mal commencé le test.
- **Effet d'ordre ou de contraste** : l'évaluation est influencée par les copies antérieures des autres élèves. Une copie d'élève qui vient dans l'ordre après une copie brillante peut être moins bien jugée et vice-versa.

Graphique 4.1 Principaux phénomènes menaçant l'objectivité et la validité de l'évaluation



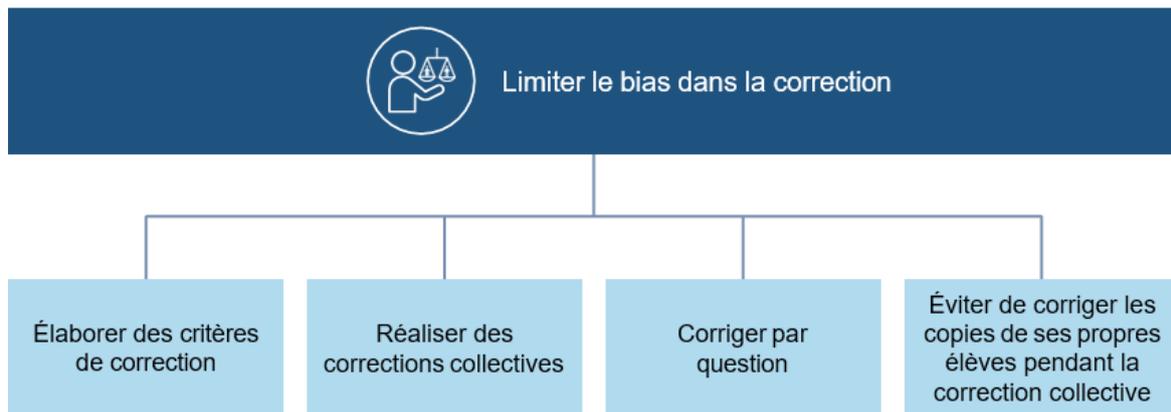
Source : Développé par les auteurs du guide

Il est possible de réduire ces biais de plusieurs façons. Quelques solutions permettant de limiter ces biais, souvent bien involontaires de la part des correcteurs, sont présentées ci-dessous.

1. **L'élaboration de critères de correction.** Pour les questions fermées et les questions ouvertes à réponse brève, la définition des critères de correction ne pose, en principe, pas trop de problèmes, même si c'est un travail qui requiert de nombreuses vérifications : par exemple, la réponse considérée comme correcte est-elle bien la seule qui soit acceptable ? En revanche, pour les questions ouvertes à réponse construite, la définition de critères de correction est plus complexe (voir section sur les grilles de correction et de codage).
2. Des **corrections collectives.** Dès qu'il s'agit d'une évaluation plus large que celle d'un enseignant pour les élèves de sa classe, il est intéressant d'impliquer tous les enseignants concernés pour une correction collective. Ceci permet de partager tous les doutes sur l'adéquation d'une réponse et d'arriver à un consensus sur le résultat à attribuer aux différentes réponses fournies par les élèves.
3. **Corriger par question (ou bloc de questions).** Pour éviter l'effet de contagion et les effets de halo et de stéréotypie, il est intéressant de ne pas corriger la copie complète d'un élève pour passer ensuite à l'élève suivant mais bien de corriger une (ou quelques) question(s) chez tous les élèves, puis de passer à la (les) question(s) suivante(s) pour tous les élèves.

4. Si la correction est collective, il est préférable qu'un enseignant ne corrige pas les copies de ses propres élèves, ou que les corrections s'effectuent « à l'aveugle », c'est-à-dire en masquant, le temps de la correction, l'identification des élèves.

**Graphique 4.2 Solutions pour limiter les biais dans la correction**



Source : Développé par les auteurs du guide

## 4.5 Notes

<sup>1</sup> Ces éléments seront développés dans le guide de l'enseignant sur l'évaluation formative et la différenciation (OCDE, non publié<sup>[57]</sup>) <sup>[48]</sup>. Des exemples de telles corrections seront donnés ou mis en ligne.

<sup>2</sup> Voir OCDE (à paraître) « Cadre de référence pour le contrôle continu au Maroc. Étude préparatoire »

1

Introduction

2

Les grandes étapes de l'élaboration des épreuves

3

La passation des épreuves

4

La correction des épreuves

Section

5

## La communication et l'utilisation des résultats

**5.1** Communication des résultats avec les différents acteurs

**5.1.1** *Communication aux élèves et avec les élèves*

**5.1.2** *Communication avec l'équipe éducative*

**5.1.3** *Communication aux parents et avec les parents*

**5.1.4** *Communication avec l'administration*

**5.2** Communication des résultats à travers le document de suivi du contrôle continu

**5.3** Notes

6

Les biais liés au genre et à d'autres caractéristiques des élèves ou de la classe

L'épreuve étant corrigée, il reste à en communiquer les résultats aux différents acteurs concernés. Ces acteurs sont nombreux : l'enseignant lui-même et ses élèves, puis les parents, l'équipe éducative et l'administration (Voir graphique 5.1). Leurs attentes et leurs besoins sont différents et la communication tiendra compte de ces différences.

**Graphique 5.1 Communication des résultats du CC aux différents acteurs concernés**



Source : Développé par les auteurs du guide

Le fait que la conception de l'épreuve et les premières étapes du processus décrites (élaboration de l'épreuve, passation et correction) auront été suivies en respectant tous les principes décrits dans ce guide n'est pas suffisant. Si la communication ne se fait pas de façon à ce que les résultats soient utiles aux acteurs concernés, l'épreuve ne sera pas valide en termes de conséquences. On verra dans les paragraphes suivants ce que peuvent être ces conséquences pour chaque catégorie d'acteurs concernés.

En général, la communication avec les acteurs concernés doit être constructive. Plutôt que d'annoncer simplement un résultat final, celle-ci devrait se concentrer sur la contextualisation du résultat et l'interprétation de ses conséquences pour l'apprentissage.

« En général, la communication avec les acteurs concernés doit être constructive. [...] Celle-ci devrait se concentrer sur la contextualisation du résultat et l'interprétation de ses conséquences pour l'apprentissage. »

Une trop grande importance accordée à l'évaluation sommative dans l'enseignement comme dans la communication génère du stress chez l'élève et renforce la motivation extrinsèque (venant du seul désir d'obtenir des bonnes notes) au détriment de la motivation intrinsèque (venant du désir d'apprendre) nécessaire à la qualité et à la pérennité des apprentissages.

Pour l'enseignant, le point qui, à juste titre, peut sembler le plus important est celui de la communication avec les élèves. Cependant, à la différence de l'évaluation formative (qui n'est pas l'objet de ce guide), les résultats du contrôle continu concernent aussi l'équipe administrative (direction et conseils de l'établissement) et les instances de contrôle (le ministère de l'Éducation nationale). Nous développerons cela dans la section qui suit.

## 5.1 Communication des résultats aux différents acteurs

### 5.1.1 Communication aux élèves et avec les élèves

La finalité première de l'évaluation étant d'améliorer l'apprentissage des élèves, ceux-ci devraient être placés au cœur du cadre d'évaluation. Ils devraient disposer des moyens nécessaires pour évaluer leurs propres progrès (l'autoévaluation constitue en effet une compétence clé pour la formation tout au long de la vie).

Le retour d'information aux élèves leur fournira des indications précises sur les points à améliorer, mais aussi sur leurs points forts<sup>1</sup>. Ce retour peut être réalisé par écrit (annotation des tests, bilans des élèves), mais aussi oralement. De façon générale, une communication bienveillante et encourageante est recommandée, ce qui n'empêche pas de conserver l'exigence nécessaire.

« Le retour d'information aux élèves leur fournira des indications précises sur les points à améliorer, mais aussi sur leurs points forts. »

Avec les élèves, la communication sur chaque contrôle continu commence avant la passation de l'épreuve. Des informations leur sont communiquées sur la date de l'épreuve et sur le matériel qu'ils devront avoir à leur disposition ainsi que sur les objectifs contrôlés.

On sait que la communication préalable des objectifs est un élément facilitateur pour l'élève ; elle lui permet de mieux préparer l'épreuve et elle le met en confiance. Par exemple, une liste des objectifs qui seront contrôlés peut lui être communiquée lui permettant de pointer les éléments avec lesquels il se sent en confiance et ceux qu'il lui faut revoir avant la passation de l'épreuve.

La communication avec les élèves se fait ensuite via l'exploitation faite en classe des résultats des évaluations pour leur fournir un retour formatif.

La correction de l'épreuve avec les élèves est un moment important de l'évaluation. Longtemps, cette correction a reposé sur le traitement de l'ensemble de l'épreuve devant les élèves par l'enseignant ou par les élèves au tableau. L'élaboration du document de suivi de l'épreuve<sup>2</sup>, complété des relevés des annotations faites au cours de la correction (chapitre 4 du guide) serait de nature à améliorer l'efficacité de cette correction. L'utilisation de ce document permettrait à l'enseignant de passer très vite sur ce qui a été très majoritairement réussi, de s'arrêter davantage sur ce qui a présenté des difficultés pour une partie importante des élèves et pour planifier un soutien

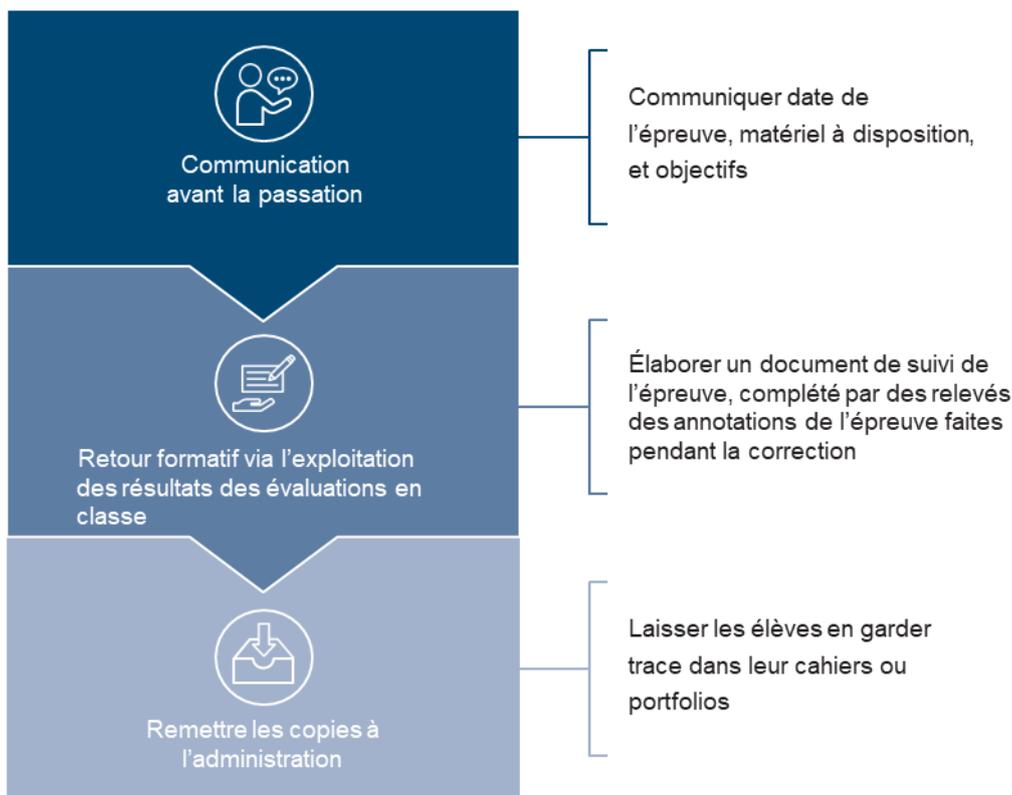
supplémentaire aux élèves qui rencontrent des difficultés persistantes (Voir le guide sur l'évaluation formative et la différenciation (OCDE, non publié<sup>[18]</sup>).

Certains enseignants peuvent préférer corriger l'épreuve avec les élèves avant de leur communiquer leurs résultats et leur permettre de consulter leur copie. C'est là une option raisonnable à condition qu'elle ne soit pas source d'ennui pour les élèves ayant bien réussi et que simultanément cette correction ne se déroule pas trop rapidement pour que les élèves en difficulté puissent en profiter pleinement.

S'il faut remettre ensuite les copies à l'administration, il est important que les élèves en gardent trace dans leurs cahiers, ou mieux dans leurs portfolios\*. La remise en forme d'éléments de la correction peut être une tâche donnée de façon différenciée aux élèves comme travail à la maison.

Il importe donc de faire usage des résultats des élèves aux épreuves de contrôle continu pour améliorer leurs apprentissages et pour planifier la remédiation. L'enseignant analyse les résultats des élèves pour les intégrer dans son enseignement et investir dans des activités de remédiation. On voit bien, ici, en quoi une évaluation sommative peut avoir des retombées formatives importantes.

#### Graphique 5.2 Les principales étapes de la communication avec les élèves



Source : Développé par les auteurs du guide

### 5.1.2 Communication avec l'équipe éducative

En général, le travail des enseignants relatif à l'évaluation est en partie collectif. On distingue le collectif disciplinaire de l'établissement (ensemble des enseignants d'une même discipline) et le collectif interdisciplinaire lié à la classe (l'ensemble des enseignants intervenant dans la classe). La communication des résultats du contrôle continu se fait dans ces deux directions.

**Graphique 5.3 La communication entre les enseignants sur les résultats du contrôle continu**



Source : Développé par les auteurs du guide

La collaboration entre les enseignants d'une même discipline a pu commencer dès l'élaboration d'un contrôle continu. Elle se poursuit alors lors de la correction des épreuves puis de l'analyse des résultats. Loin d'être une perte de temps, cette collaboration permet de mieux rationaliser le travail des enseignants et de répartir entre eux une partie de la tâche. Elle renforce la position professionnelle des enseignants ainsi que la fiabilité de leurs contrôles.

La collaboration entre enseignants d'une même classe se fait en particulier lors de l'élaboration de bilans individuels et collectifs (élève et classe) et des conseils de classe. Pour ces bilans, la prise en compte des résultats des contrôles continus est importante, mais elle ne suffit pas. Il convient alors de contextualiser les résultats en tenant compte des caractéristiques de la classe ainsi que de



chaque élève en particulier. Il ne s'agit bien sûr pas de modifier les notes, mais, dans la mesure du possible, de les éclairer.

Signalons ici que la suite des notes des contrôles ne permet pas toujours d'analyser les progrès des élèves ni ceux de la classe. Les contrôles successifs portent sur des contenus et des objectifs différents et sont de difficultés différentes. Sauf à utiliser des procédures statistiques visant à les rendre comparables<sup>3</sup>, des expressions telles que « a progressé » ou « est en baisse » sont à utiliser avec prudence.

### 5.1.3 Communication aux parents et avec les parents

La communication se fera donc aussi en direction des parents. Là encore il ne s'agit pas que de fournir des notes : il faut aussi les expliquer et fournir des conseils pour impliquer les parents en vue d'une amélioration des apprentissages de leurs enfants. En un mot, il s'agit d'engager un dialogue constructif avec les parents.

« Il ne s'agit pas que de fournir des notes: il faut aussi les expliquer et fournir des conseils pour impliquer les parents en vue d'une amélioration des apprentissages de leurs enfants. »

Les notes sont également communiquées directement aux parents chaque semestre par le biais du bulletin scolaire. D'autre part, les notes sont enregistrées sur la base de données en ligne MASSAR, à laquelle ont accès, de façon différenciée, les élèves, leurs parents et les enseignants.

Si les parents ont besoin d'être informés des notes de leurs enfants liées aux évaluations sommatives, ils apprécient aussi de recevoir des appréciations qualitatives et formatives relatives à leurs performances, leur permettant ainsi de suivre et de contribuer à leurs apprentissages.

Pour cela, les enseignants peuvent fournir aux parents une synthèse des annotations sur l'épreuve avec mention de l'état des acquis par rapport aux attentes du programme, des progrès enregistrés et des efforts à fournir par la suite. L'enseignant pourrait se baser sur le document de suivi de l'épreuve du contrôle continu pour fournir ce retour individualisé (voir section ci-dessous Communication des résultats à travers le document de suivi du contrôle continu).

En plus d'une communication écrite régulière, il est important d'assurer une communication en face à face avec les parents, notamment par le biais de réunions parents-enseignants régulièrement programmées ou de réunions ad hoc en réponse à certains besoins urgents.

Le dialogue avec les parents sera adapté à leur niveau de compréhension et de réceptivité. Il est souhaitable qu'il soit positif et bienveillant. En fin de compte, il est important que les parents puissent comprendre ce qu'ils peuvent faire pour aider leur enfant.

### 5.1.4 Communication avec l'administration

Les résultats des épreuves de contrôle continu sont communiqués à l'administration de l'établissement sous une forme définie par l'établissement ou par le ministère. Il peut s'agir en

particulier de lui remettre au fur et à mesure les fichiers informatiques des documents de suivi des épreuves<sup>4</sup>, de façon à compléter les dossiers rassemblant les résultats des contrôles d'une classe, de plusieurs classes ou de tout l'établissement.

Les notes des contrôles ainsi que d'autres informations à préciser (les épreuves, des jugements synthétiques ou détaillés des résultats) sont transmises à la plateforme MASSAR.<sup>5</sup>

Au niveau de l'établissement, les notes obtenues aux contrôles continus, sont aussi reportées dans le livret scolaire de l'élève. Pondérées par les coefficients correspondants, elles font partie intégrante de la moyenne générale annuelle.

Il est souhaitable que les notes du contrôle continu soient accompagnées d'appréciations mettant en valeur les progrès de l'élève tout en précisant les difficultés rencontrées et les moyens à mettre en œuvre pour les surmonter (remédiation).

Il importe que le livret scolaire permette de suivre l'évolution des résultats de l'enseignement au sens large, sans se limiter aux notes de contrôles. L'observation continue des enseignants leur permet en particulier de porter des jugements sur des compétences transversales. Ces dernières, bien que difficilement mesurables, font partie des attendus curriculaires ; notamment l'esprit d'initiative, la pensée critique, les compétences sociales, l'intérêt et la motivation, la persévérance. Elle leur permet aussi de mieux contextualiser les résultats du contrôle continu.

## 5.2 Communication des résultats à travers le document de suivi du contrôle continu

La communication des résultats des contrôles continus aux différents acteurs peut sembler complexe. Elle peut cependant être largement facilitée par une organisation adaptée. Nous proposons dans les lignes qui suivent un modèle de document qui permettrait de faciliter grandement la communication.

Les enseignants ont besoin de garder une trace aussi complète que possible des contrôles qu'ils ont fait passer ; ceci, pour eux-mêmes aussi bien qu'aux fins de communication externe. Traditionnellement cela se fait d'une part dans le cahier de textes et d'autre part dans les documents individuels de l'enseignant.

À une épreuve donnée se trouve associée la mémoire de ses conditions de réalisation, des objectifs pris en compte, des résultats obtenus par les élèves ainsi que des analyses qui auraient été réalisées a priori et a posteriori.

Un document de suivi d'une épreuve est un instrument qui regroupe l'ensemble des informations associées à cette épreuve. Ce document, totalement ou partiellement informatisé, pourrait contenir : la liste des connaissances et des compétences contrôlées, rapportées aux attentes curriculaires, l'épreuve elle-même et la liste des

« Un document de suivi d'une épreuve est un instrument qui regroupe l'ensemble des informations associées à cette épreuve. »

items prévus pour la prise d'information sur la performance, ainsi que le barème de notation (points attribués à chaque item).

Il contiendrait aussi le détail des notes et appréciations de chaque élève ayant passé l'épreuve, les enseignements tirés de la correction en ce qui concerne les acquis des élèves, les mesures prises pour le retour aux élèves et les opérations de remédiation, ainsi que l'analyse a posteriori qui en aura été faite.

Les possibilités offertes par les TIC (les technologies de l'information et de la communication) permettent de rationaliser la mémorisation et la communication relatives aux évaluations tout en les simplifiant. Un document de suivi du contrôle continu (éventuellement en ligne) pourrait regrouper, pour chaque enseignant l'ensemble de ces informations. Un dossier informatique pourrait rassembler l'ensemble des documents de suivi des contrôles et un fichier par élève relatif à son suivi personnel.

Dans les pays où le travail collectif des enseignants est la règle, ce type de document est créé par les équipes d'enseignants ; ailleurs, chaque enseignant s'organise pour garder la mémoire de ses évaluations en relation avec la tenue du cahier de textes.

Un tel document de suivi de l'épreuve faciliterait la communication entre les différents partis concernés et réduirait considérablement la charge liée aux contrôles continus, certains aspects de la communication se réduisant alors à des copier-coller du document de suivi. De plus, l'enseignant garderait la mémoire des contrôles continus et des enseignements qu'il aura pu en tirer tout au long de l'année et même, d'une année à l'autre.

Nous proposons avec ce guide un modèle de document de suivi qui demande à être adapté au contexte et à son utilisation. Le dossier contenant ce modèle de document de suivi ainsi qu'un exemple d'épreuve se trouve à l'adresse [https://www.dropbox.com/sh/5e6sjr0hx8p8twk/AAC4e1fOFBI\\_9HtuU3W4RDv3a?dl=0](https://www.dropbox.com/sh/5e6sjr0hx8p8twk/AAC4e1fOFBI_9HtuU3W4RDv3a?dl=0).

## 5.3 Notes

---

<sup>1</sup> Le Guide de l'Enseignant sur l'Évaluation Formative et la Différenciation Pédagogique (OCDE, non publié<sup>[18]</sup>) traite de la communication aux élèves et avec les élèves de manière plus approfondie (retour d'information et adaptation de l'enseignement).

<sup>2</sup> Voir la section suivante Communication des résultats à travers le document de suivi du contrôle continu.

<sup>3</sup> Par exemple, ramener toutes les notes à une même moyenne et à un même écart-type.

<sup>4</sup> Voir la section suivante Communication des résultats à travers le document de suivi du contrôle continu.

<sup>5</sup> Cette plateforme est ouverte de façon différenciée aux différents acteurs

1

Introduction

2

Les grandes étapes de l'élaboration des épreuves

3

La passation des épreuves

4

La correction des épreuves

5

La communication et l'utilisation des résultats

Section

6

## Les biais liés au genre et à d'autres caractéristiques des élèves : une menace pour l'équité

**6.1** Les biais liés aux caractéristiques des élèves

**6.2** Biais liés aux caractéristiques de la classe : l'effet Posthumus

**6.3** Notes



## Encadré 6.1

### L'équité : une question de justice

L'équité, ce n'est pas l'égalité, c'est une question de justice (justice méritocratique ou compensatoire).

Se poser la question de l'équité, c'est se demander quelle inégalité peut être considérée comme « juste » ou « acceptable » ?

Par exemple, on peut vouloir apporter un soutien particulier aux élèves défavorisés ou à besoins spécifiques. Dans ce cas il est normal que l'on traite les élèves de manière inégale (en offrant plus de soutien aux moins favorisés, en les évaluant avec moins de sévérité), en leur offrant une évaluation adaptée. Cela peut être généralement considéré comme acceptable, ou comme juste, dans une logique compensatoire (aider ceux qui ont plus de difficulté, comme dans les politiques de discrimination positive). Mais dans une autre logique - évaluer les mérites des élèves - cette inégalité de traitement ne sera pas considérée comme juste. L'équité est d'abord une question de valeurs.

Source : Développé par les auteurs du guide

## 6.1 Les biais liés aux caractéristiques des élèves

Les biais liés au genre, à l'origine socioculturelle ou ethnique, ou à la langue parlée par les élèves constituent une menace pour l'équité de l'évaluation.

Ceux-ci ont fait l'objet de nombreuses études scientifiques en sciences de l'éducation et en psychologie sociale. Il faut comprendre que ces biais sont le plus souvent le résultat de processus inconscients, de croyances ou de stéréotypes, et ne relèvent pas d'une intention de discriminer les élèves.

Ces biais peuvent intervenir au niveau de la conception de l'épreuve et au niveau de la correction des copies d'élèves.

« Les biais liés au genre, à l'origine socioculturelle ou ethnique, ou à la langue parlée par les élèves constituent une menace pour l'équité de l'évaluation. »

Graphique 6.1 Les niveaux des biais liés aux caractéristiques des élèves



Source : Développé par les auteurs du guide

### 1. Au niveau de la conception de l'épreuve

- Le choix des textes ou des thèmes de mise en situation ou de production écrite peut avantager les filles ou les garçons. Il vaut donc mieux éviter les thèmes fortement « genrés », dont on sait par expérience qu'ils intéressent plus les garçons ou les filles.
- Il en va évidemment de même pour d'autres caractéristiques des élèves : il faut éviter des thèmes qui pourraient être inégalement familiers aux élèves en fonction de leur origine sociale, culturelle ou ethnique, comme les sports d'hiver, ou la pratique des sports nautiques, du tennis ou de l'équitation.
- Dans les épreuves de compréhension de textes (lecture), une étude de Lafontaine et Monseur (2009<sup>[19]</sup>) a montré en analysant les données PISA (Programme international pour le suivi des acquis des élèves) de l'ensemble des pays de l'OCDE que :
  - i. Augmenter la part de textes narratifs ou littéraires avantage les filles ; inversement, augmenter la part de textes informatifs et la part de textes discontinus (des documents qui comportent du texte, mais aussi des graphiques, des tableaux et des schémas) avantage relativement les garçons.
  - ii. Augmenter la part de questions à choix multiples (QCM) et de questions à réponse courte avantage relativement les garçons ; inversement, augmenter la part de réponses ouvertes construites avantage les filles.
  - iii. Augmenter la part de questions complexes ou de haut niveau taxonomique avantage les filles.

Il faut être conscient de l'existence de ce type de biais et savoir que lorsque l'on pose des choix méthodologiques pour concevoir une épreuve ou des questions, ceux-ci ne sont pas neutres et peuvent avoir un impact non négligeable sur les résultats respectifs des garçons et des filles.

Les études menées dans d'autres domaines (mathématiques et sciences) n'aboutissent pas aux mêmes résultats. Les effets observés selon le genre dépendent donc du domaine d'apprentissage et sans doute d'autres facteurs contextuels et culturels. D'une manière plus générale, il est établi que les QCM sont mieux réussis par les élèves peu performants dans un domaine et jouent donc en leur faveur. L'inverse est vrai pour les questions ouvertes construites.



## Encadré 6.2

### La menace du stéréotype

La simple manière dont une épreuve ou une tâche d'évaluation est présentée peut jouer sur les différences garçons-filles. L'exemple le plus connu est celui d'une tâche de tracé d'une figure. Selon que la tâche est présentée comme une tâche de dessin ou comme une tâche de géométrie, la réussite de celle-ci en fonction du genre va varier sensiblement. Les filles réussissent mieux quand la tâche est présentée comme un « dessin » que quand la tâche est présentée comme une tâche de géométrie (Huguet et Régner, 2007<sup>[20]</sup>). En psychologie sociale, le phénomène s'appelle la « menace du stéréotype » (Spencer, Quinn et Steele, 1999<sup>[21]</sup>) ; (Steele et Aronson, 1995<sup>[22]</sup>).

1. Source : Huguet & Régner (2007). Stereotype threat among schoolgirls in quasiordinary classroom circumstances. *Journal of educational psychology*, 99, 545-560, <https://psycnet.apa.org/doi/10.1037/0022-0663.99.3.545>
2. Spencer, S., D. Quinn et C. Steele (1999), « Stereotype threat and women's math performance », <https://doi.org/10.1006/JESP.1998.1373> ;
3. Steele, M. et J. Aronson (1995), « Stereotype threat and the intellectual test performance of African Americans », <https://doi.org/10.1037/0022-3514.69.5.797>

## 2. Au niveau de la correction des copies d'élèves

### a. Influence du genre de l'élève sur l'évaluation de copies en mathématiques

Lafontaine et Monseur (2009<sup>[19]</sup>) ont mené une étude expérimentale en Communauté française de Belgique à ce propos. Quarante-huit enseignants de mathématiques (secondaire supérieur/lycée) ont été répartis en deux groupes comparables et devaient évaluer quatre copies fictives de niveaux différents (très faible, faible, assez bonne et très bonne) attribuées dans le premier groupe à une fille, dans le second groupe à un garçon.

Les résultats montrent que les enseignants ont tendance à évaluer de manière différente des copies strictement équivalentes selon qu'elles sont attribuées à des garçons ou à des filles. Premièrement, ils surévaluent en moyenne légèrement les copies des garçons. Deuxièmement, un effet croisé a été mis en évidence : les enseignants jugent plus favorablement les copies bonnes et très bonnes portant un prénom de garçon et moins favorablement les mêmes copies portant un prénom de fille. En revanche, ils sous-estiment les copies faibles et très faibles attribuées à un garçon et surestiment les mêmes copies faibles quand elles sont attribuées à une fille. L'intérêt de cette étude est qu'elle montre qu'aussi bien les filles très performantes en mathématiques que les garçons peu performants



sont sous-notés. On est donc vraiment face à un problème d'équité, dans lequel certaines filles et certains garçons ne sont pas évalués à leur juste valeur.

Le phénomène joue de manière largement inconsciente et ne dépend pas du genre de l'enseignant. Pour lutter contre cet effet des stéréotypes et des attentes différentes selon le genre des élèves, il faut y sensibiliser les enseignants et s'interroger sur son propre fonctionnement en tant qu'évaluateur.

b. Influence d'autres caractéristiques des élèves sur le jugement des enseignant

De très nombreuses études ont été menées dans différents contextes éducatifs sur les facteurs qui influencent l'enseignant lorsqu'il porte un jugement sur une production d'élève (corrige une copie, attribue une note, ou complète un bulletin). Ces études comparent le jugement de l'enseignant (les notes attribuées) aux résultats qu'obtiennent les élèves à une épreuve standardisée externe non corrigée par leur enseignant dans le même domaine.

Ces études permettent donc d'estimer dans quelle mesure le jugement de l'enseignant et les scores obtenus à une épreuve externe concordent et, s'ils diffèrent, par quoi cela peut s'expliquer. On ne parle pas de biais dans ce cas.

La méta-analyse consacrée par Dusek et Joseph (1983<sup>[23]</sup>) et les travaux de Bressoux et Pansu (2003<sup>[24]</sup>) ont mis en évidence qu'à performances égales (mesurées par une épreuve externe), une série de critères influence le jugement de l'enseignant, en particulier l'origine sociale de l'élève et le fait que l'élève ait connu le redoublement. Les élèves d'origine sociale défavorisée et les élèves en retard scolaire sont notés plus sévèrement par leurs enseignants que ne le fait l'épreuve externe. En ce qui concerne le genre, les résultats sont mitigés : en langue maternelle, aucun effet n'est observé ; en mathématiques, en France, dans l'enseignement primaire, les garçons obtiennent un jugement légèrement plus favorable (+0,32 sur une échelle de 0 à 10) que les filles (Bressoux et Pansu, 2003<sup>[24]</sup>) à performances égales à l'épreuve externe.

## 6.2 Biais liés aux caractéristiques de la classe : l'effet Posthumus

Dans les années 1980 (Crahay, 2019<sup>[25]</sup>), plusieurs recherches ont mis en évidence que les enseignants adaptent leurs exigences en fonction du niveau moyen des élèves de leur classe. Ce réflexe professionnel est légitime et va de soi lorsque l'enseignant planifie son enseignement : l'enseignant tient compte du niveau de la classe pour ajuster son enseignement aux besoins et au niveau de ses élèves. Toutefois, quand vient le moment d'évaluer les élèves de manière sommative ou certificative, adapter ses exigences au niveau des élèves peut conduire à un traitement inéquitable des élèves selon la classe qu'ils fréquentent.

Des recherches anciennes (Posthumus, 1947, cité par De Landsheere (1980<sup>[26]</sup>)) ont mis en évidence un phénomène récurrent appelé « effet Posthumus » lorsque les enseignants sont responsables d'une évaluation. Ceux-ci ont tendance à choisir des questions qui vont permettre de bien distinguer les élèves : certaines questions difficiles que seuls les très bons élèves réussiront, quelques questions très faciles qui permettront aux élèves les plus en difficulté d'engranger

quelques points et une proportion importante de questions de difficulté moyenne. Il y a derrière cette tendance une sorte de norme sociale implicite qui peut avoir des effets négatifs en matière d'équité.

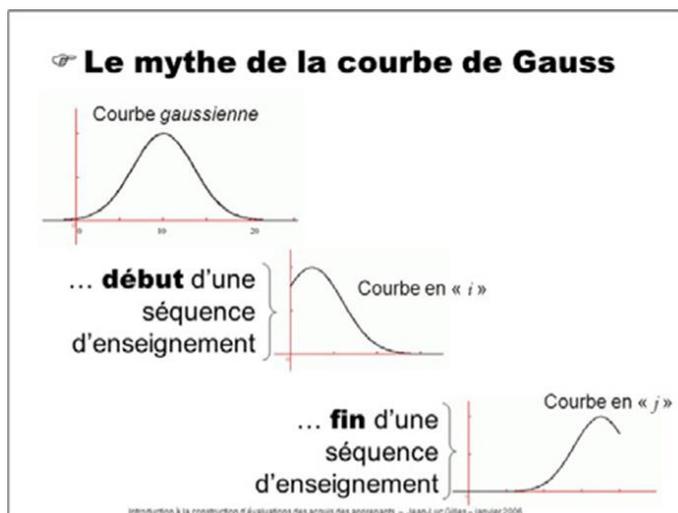
Derrière cette tendance, on retrouve la croyance en une espèce de constante (Antibi, 2003<sup>[27]</sup>) : dans toute classe, il y aurait une majorité d'élèves moyens, une minorité d'élèves forts et une minorité d'élèves faibles. La distribution des connaissances prendrait la forme d'une courbe de Gauss (voir graphique 6.2). Si ceci peut s'avérer exact sur une cohorte d'élèves nombreuse, il n'en va pas de même sur une classe dont les effectifs sont réduits. Tous les enseignants vous diront que certaines années, ils ont une « mauvaise » classe et d'autres années une « bonne classe ».

Croire qu'il est normal que dans une classe les résultats prennent la forme d'une courbe de Gauss, c'est en somme ne pas tenir compte de la qualité de l'enseignement qui a précédé l'évaluation. Comme l'illustre le schéma ci-dessous, ce qui est normal en matière d'apprentissage, c'est qu'avant l'enseignement (au début d'une séquence d'enseignement), une majorité d'élèves ne maîtrisent pas

la connaissance ou la compétence enseignée (courbe en « i » dans le graphique 6.2). En revanche, si un enseignant enseigne de manière efficace, les résultats de l'évaluation devraient à la fin de la séquence d'enseignement prendre la forme d'une courbe en J (voir graphique 6.2). La courbe en J correspond à ce que l'on appelle la pédagogie de la maîtrise : la majorité des élèves maîtrise la matière enseignée et seule une minorité ne la maîtrise pas.

« si un enseignant enseigne de manière efficace, les résultats de l'évaluation devraient à la fin de la séquence d'enseignement prendre la forme d'une courbe en J qui correspond à ce que l'on appelle la pédagogie de la maîtrise : la majorité des élèves maîtrise la matière enseignée et seule une minorité ne la maîtrise pas »

Graphique 6.2 Le mythe de la courbe de Gauss



Source : (Gilles, 2006<sup>[28]</sup>)



Devant de tels résultats, il n'est pas rare que les enseignants ou les parents réagissent en disant que le contrôle ou l'examen était trop facile, comme si un bon enseignant se devait d'être sévère et de mettre en échec un certain nombre d'élèves. Or, il est normal qu'une majorité d'élèves réussissent un contrôle portant sur un contenu qui a été bien enseigné. C'est le signe d'un enseignement de qualité.

Inclure dans un contrôle de type sommatif des questions très difficiles qui dépassent parfois la matière vue relève d'une logique qui ne devrait pas normalement s'appliquer à des contrôles menés en classe. S'il est essentiel dans le cadre d'une évaluation standardisée visant à fournir des indicateurs pour le système (Programme international pour le suivi des acquis des élèves [PISA], Progress in International in Reading Literacy Study [PIRLS], étude comparative Trends in Mathematics and Science Study [TIMSS]) de disposer d'épreuves et d'items qui ont un bon pouvoir de discrimination<sup>1</sup>, cela ne s'applique pas nécessairement aux épreuves qui visent à vérifier si les objectifs d'apprentissage sont atteints. Si, par exemple, une épreuve externe porte sur un socle de compétences minimal, il est parfaitement normal et attendu que 80 % des élèves la réussissent.

Ajuster le contenu des épreuves d'évaluation au niveau de la classe peut conduire à ce que deux élèves de niveau a priori équivalent réussissent plus ou moins bien l'épreuve selon son degré de difficulté. En cas d'épreuves décidant de la réussite de l'année ou du passage en classe supérieure, tels les contrôles continus dans le système éducatif marocain, l'un de ces deux élèves pourrait échouer et l'autre réussir, selon la classe qu'il fréquente. Ces situations seraient profondément inéquitables. Ce risque est davantage présent dans les systèmes éducatifs où les décisions de passage de classe se prennent sur la base d'épreuves conçues par les enseignants avec un cadre de référence peu contraignant, voire inexistant, et ceci encore plus si les différences de niveau entre établissements scolaires et classes sont importantes. Ce phénomène a été particulièrement bien documenté pour la Belgique francophone par (Grisay, 1984<sup>[29]</sup>) et (Crahay, 2019<sup>[25]</sup>).

C'est principalement ce type de constat qui a justifié la mise en place d'évaluations externes standardisées pour les niveaux d'études où intervient une certification. Seules celles-ci peuvent faire en sorte que les élèves soient vraiment mis sur un pied d'égalité et traités de manière équitable. Ceci étant dit, en l'absence d'un tel dispositif au niveau du système, un dispositif plus local à l'échelle d'une région, d'une municipalité, ou d'un groupe d'établissements scolaires peut déjà contribuer à réduire l'ampleur du problème d'inéquité. Tout petit pas dans la direction pour rendre plus communs le contenu ou les critères de correction est une avancée vers plus d'équité. Même la collaboration entre enseignants au sein d'un établissement, qui implique de se mettre d'accord sur ce que l'on évalue et comment, permet de rendre l'évaluation plus équitable.

## 6.3 Notes

---

<sup>1</sup> Des questions et items qui permettent par exemple de discriminer de manière fine les élèves très performants (excellents) des élèves simplement performants (bons élèves).

# Annex A. Glossaire

## Alignement

Les études telles que celles menées par l'OCDE (2013<sup>[30]</sup>), ainsi que différentes recherches, constatent que les niveaux dans lesquels intervient l'évaluation des acquis des élèves, notamment l'évaluation externe, les examens et les évaluations en classe (sommativ, formative, diagnostique), sont souvent assez peu coordonnés.

L'alignement de ces niveaux, pour faciliter la cohérence de l'ensemble des situations d'évaluation, constitue une recommandation constante. Cela rejoint l'objectif de développement dans les systèmes éducatifs d'une culture d'évaluation susceptible, en particulier, de contribuer à l'harmonisation des critères par rapport auxquels sont évalués les apprentissages des élèves. L'opposé de l'alignement est la dissonance, qui fait perdre de l'efficacité à l'ensemble du système. Cette dissonance est constatée dans de nombreux pays.

## Autoévaluation et évaluation mutuelle

À la suite de (Allal, 1988<sup>[31]</sup>), (Vial, 1997<sup>[32]</sup>) définit l'autoévaluation comme « un processus cognitif qui intervient dans l'autoguidage de toute conduite ». Il précise que « l'auto-évaluation est une dimension inhérente aux situations d'apprentissage, c'est pourquoi il est nécessaire de faire émerger l'auto-évaluation : elle doit être conçue comme un objectif d'apprentissage ».

Depuis, la référence à l'autoévaluation s'est généralisée dans les curricula de nombreux pays, en liaison avec la métacognition comme processus réflexif sur ses propres actions.

De plus en plus de curricula insistent sur l'intérêt d'encourager l'autoévaluation chez les élèves. L'autoévaluation devient une compétence qu'il s'agit aussi d'évaluer, de même qu'est encouragée l'évaluation mutuelle, c'est-à-dire l'évaluation par les pairs. Linda Allal (1999<sup>[33]</sup>) propose « trois modalités distinctes, mais liées, de l'implication de l'élève dans l'évaluation : l'autoévaluation au sens strict, l'évaluation mutuelle entre pairs et la co-évaluation stipulant une confrontation des évaluations réalisées par l'enseignant et par l'élève » (Allal, 1999<sup>[33]</sup>).

L'autoévaluation et l'évaluation mutuelle impliquent que les élèves réfléchissent régulièrement aux processus qu'ils mettent en œuvre pour effectuer des tâches, à leurs difficultés et à leurs progrès, de façon formelle et informelle pour éclairer leurs futurs objectifs d'apprentissage.

Cela aide les élèves à assumer la responsabilité de leur apprentissage et à prendre conscience de leurs stratégies d'apprentissage, ce qui est important dans les approches fondées sur les compétences (Maghnoij et al., 2018<sup>[34]</sup>).

## Banque d'items

L'Institut suisse de recherche et de documentation pédagogique, qui bénéficie d'une longue expérience dans ce domaine, définit une banque d'items comme « un répertoire (liste, catalogue, recueil) d'items qui, généralement, ont déjà été utilisés à des fins d'évaluation et qui sont mis à la

disposition de personnes (chercheurs ou enseignants) souhaitant y avoir recours (accès libre ou soumis à certaines conditions). » (IRDP, 2016<sup>[35]</sup>).

Les items inclus sont souvent classés en fonction des 1) « domaines de compétence, contenus d'enseignement et objectifs pédagogiques auxquels ils se réfèrent » ; et 2) « le(s) degré(s) ou l'âge des individus (élèves, étudiants) à qui ils sont destinés ». Dans certains cas, des informations utiles sur le niveau de difficulté, le pouvoir de discrimination et/ou la sensibilité des items, ainsi que des conseils d'utilisation et des suggestions peuvent aussi être inclus dans la fiche technique d'un item.

Il existe aussi des cas où des banques d'items sont réservées à la conception d'examens, de tests standardisés, d'évaluations nationales ou internationales. Ces banques ne sont pas ouvertes pour d'autres usages et protègent leur sécurité. Toutefois, elles donnent accès à des items dits « libérés » pour l'information et la formation des enseignants et des élèves.

La plupart des systèmes modernes se sont munis de banques d'items et d'épreuves ou de tests d'évaluation. C'est ainsi que des banques d'items ont été développées dans de nombreux pays : Australie, France, Irlande, Suisse par exemple. Certaines sont très développées et se prêtent à de nombreux usages et utilisateurs :

- mise à la disposition des enseignants d'items et d'instruments d'évaluation critériés, analysés et validés
- réception et conservation des items et des épreuves élaborées par les enseignants, les collectifs d'enseignants, ainsi que celles déjà utilisées dans les examens et les évaluations externes
- réception de données résultant de l'utilisation de ces items et épreuves, permettant leur calibrage
- une partie de la banque d'items peut être sécurisée pour être utilisée dans le cadre des examens, une autre dans le cadre des contrôles continus, une autre peut être ouverte et accessible aux enseignants, aux élèves et aux parents.

## Compétences

Les compétences sont définies internationalement comme « un ensemble de connaissances, de savoir-faire et d'attitudes appropriées au contexte » (UE, 2006<sup>[36]</sup>). L'OCDE précise que : « Le concept de compétence ne se réduit pas à l'acquisition de connaissances et d'aptitudes, il suppose la mobilisation de connaissances, compétences, attitudes et valeurs pour répondre à des besoins complexes » (OCDE, 2018<sup>[37]</sup>). Par exemple, l'aptitude à communiquer efficacement est une compétence qui peut s'appuyer sur les connaissances linguistiques d'un individu, sur ses compétences pratiques en informatique et sur son attitude à l'égard des personnes avec lesquelles il communique.

On distingue en général plusieurs types de compétences :

- Les compétences générales. Exemple : savoir prendre des notes

- Les compétences disciplinaires. Exemple : en mathématiques, être capable de modéliser une situation problématique
- Les compétences spécifiques. Exemple : être capable de monter une expérience mettant en évidence l'hydrolyse
- Les compétences transversales. Exemple : être capable de travailler en groupe.

Le terme « compétence » reste cependant polysémique et difficilement opératoire dans le domaine de l'éducation. Qu'il s'agisse d'enseignement ou d'évaluation, définir des compétences est une opération complexe qui ne peut se réduire à rebaptiser compétence ce qui était auparavant nommé savoir-faire. Une compétence ne peut qu'être associée à un type de tâches, et non à une tâche particulière. Par exemple, monter une expérience mettant en évidence l'hydrolyse de l'eau est une tâche particulière ; monter une expérience de dissociation moléculaire est une compétence.

Tous les systèmes éducatifs insistent sur l'acquisition de compétences. Il nous semble important de préciser que, stricto sensu, les compétences ne s'apprennent pas : elles se développent. Il est donc légitime de parler d'état ou de niveau de développement de telle ou telle compétence, et donc de considérer ces niveaux comme des acquis possibles.

Insistons sur le fait que, dans cette étude, nous parlons de compétences et d'évaluation des compétences et non d'enseignement ou d'approche par compétence, ce qui est une autre question qui sort totalement du sujet de cette étude.

Le Maroc utilise une classification des tâches qui évoque des degrés de compétence :

- Premier degré : application directe
- Deuxième degré : application indirecte
- Troisième degré : complexité inédite.

Il s'agit là d'une caractérisation des niveaux de complexité des tâches, aussi appelée taxonomie.

## Connaissances

Les connaissances sont les acquis théoriques et techniques qui sont assimilés par formation, expérience professionnelle ou extraprofessionnelle. Les syllabus précisent ainsi les connaissances que les élèves sont censés acquérir à un niveau donné de leur scolarité. Par exemple, en mathématiques, une connaissance peut être de connaître la somme des angles d'un triangle. Il ne s'agit pas ici de définir « la connaissance », objet complexe tant du langage courant que du langage philosophique, mais les « éléments de connaissance ».

On a parfois tendance à opposer connaissances et compétences. En réalité, la plupart des compétences que l'on peut définir se nourrissent de connaissances. Vouloir évaluer des compétences ne signifie pas qu'il faille faire l'impasse sur l'évaluation des connaissances.



Par exemple, en mathématiques, la compétence « savoir modéliser une situation issue ou non de la vie courante », essentielle, suppose des connaissances et des savoir-faire qu'il est souvent nécessaire d'évaluer indépendamment de la compétence proprement dite. Ainsi, si un élève échoue à modéliser une situation de proportionnalité, il faudra se demander s'il/elle possède bien les connaissances et les savoir-faire nécessaires.

## Critères

Les critères sont les caractéristiques attendues d'une performance qui sert de référence à l'évaluation. Les critères sont les dimensions explicites, choisies par l'évaluateur (ou l'institution évaluative) pour analyser la production de l'élève.

Sur chaque critère, on fixe le niveau d'exigence (ou seuil de réussite) et on repère les niveaux de performance de l'élève.

## Curriculum

Le curriculum est défini comme « une description de ce que les élèves doivent apprendre, pourquoi, comment, jusqu'à quel point, et ce, de manière systématique et intentionnelle.

Le programme scolaire n'est pas une fin en soi, mais plutôt un moyen de favoriser un apprentissage de qualité » (UNESCO-IBE, 2013<sup>[38]</sup>). Les chercheurs étendent en général ces définitions à l'ensemble des expériences susceptibles d'être vécues dans la vie scolaire des élèves et à tout ce qui, dans le système éducatif, conditionne ces expériences. En particulier et de façon non exclusive : programmes d'enseignement, objectifs et standards, organisation du temps et de l'espace scolaire, formation des enseignants, évaluation des élèves, contrôle des acquis et examens.

Cependant, le terme est largement polysémique et a de « nombreuses définitions, allant d'un "programme d'études" planifié à une vue d'ensemble qui comprend toutes les expériences d'apprentissage dont l'établissement scolaire est responsable » (UNESCO-IBE, 2013<sup>[38]</sup>).

Certains décrivent même le programme comme un pacte social et politique qui reflète une vision commune de la société fondée sur les besoins locaux, nationaux et mondiaux. Cette définition implique la participation des principales parties prenantes, qui peuvent avoir des points de vue différents, dans le développement du curriculum.

Il y a une tendance d'aller vers une conception plus inclusive des curricula. Conformément aux conceptions initiales relatives au curriculum (John Dewey dès le début du vingtième siècle), les curricula officiels modernes cherchent à prendre en charge une vision plus holistique de la formation des élèves, allant jusqu'à y intégrer les activités parascolaires menées hors la classe et des compétences qui, jusqu'à présent, étaient rarement évoquées dans les curricula : développement de l'imagination, de la créativité, de l'esprit critique, de la capacité à s'autoévaluer et à travailler en groupe.

Les curricula sont de plus en plus souvent bâtis, au moins partiellement, en termes de compétences. Cela implique que les évaluations doivent porter en partie sur les compétences, et non simplement sur les connaissances et les savoir-faire. Certains curricula, ou parties de curricula nationaux, portent sur la partie jugée essentielle de la formation. C'est le cas en France et au Maroc avec le «

socle commun de connaissances et des compétences » ; c'est le cas dans d'autres pays avec le core curriculum, comme aux États-Unis ou en Finlande.

## Évaluation

La langue anglaise a deux termes pour évoquer la question qui nous intéresse ici : *evaluation* et *assessment*. La distinction entre ces deux termes n'est pas toujours assurée, même dans le monde anglophone. En général, le terme *evaluation* est plutôt utilisé lorsqu'il s'agit d'évaluer des systèmes ou des dispositifs, tandis que le terme *assessment* est davantage employé pour évaluer la qualité d'un produit, d'une personne ou d'un objet par rapport à des normes ou critères bien définis. Une *evaluation* se conclut ainsi plutôt par un jugement tandis qu'un *assessment* se conclut plutôt par un constat.

Si l'on se limite au domaine de l'éducation, les termes sont définis comme suit (UNESCO-IBE, 2013<sup>[38]</sup>) :

- *Evaluation* : Un processus systématique visant à juger de l'efficacité de tout programme d'enseignement et d'apprentissage
- *Assessment* : Le processus par lequel les progrès et les réalisations d'un ou plusieurs apprenants sont mesurés ou jugés en conformité avec des critères de qualité spécifiques.

Dans le contexte de cette étude, nous nous limitons à l'évaluation des élèves, donc *assessment* en anglais. Pour l'OCDE, le terme *assessment* désigne les jugements sur les progrès individuels des élèves et la réalisation des objectifs d'apprentissage (OCDE, 2013<sup>[30]</sup>). Il couvre les évaluations en classe ainsi que les évaluations externes à grande échelle et les examens.

La littérature sur l'évaluation fait une distinction entre l'évaluation sommative et l'évaluation formative, en particulier lorsqu'il est question de l'évaluation faite dans le cadre de la classe. Cette distinction a été introduite à la fin des années 1960 par les travaux de Michael Scriven (1967), suivis par ceux de Benjamin Bloom et al. (Eisner, 1972<sup>[39]</sup>), mais ne s'est généralisée dans les curricula scolaires qu'à partir des années 1990 (NCTM, 2000<sup>[40]</sup>).

Il importe de veiller à ce que les différents types et objectifs d'évaluation soient équilibrés afin de créer l'interaction constructive entre les enseignants et les élèves qui permet aux élèves de progresser et d'être motivés pour apprendre. Dans les deux types d'évaluation, sommative et formative, une évaluation efficace fournit des informations aux élèves, aux enseignants et aux décideur(se)s sur le niveau des connaissances et des compétences des élèves et permet d'identifier les améliorations à apporter aux stratégies d'apprentissage, à l'enseignement ou aux politiques éducatives.

La relation complexe entre l'évaluation formative et l'évaluation sommative constitue une préoccupation constante. Le consensus général, par exemple Bennett (2011<sup>[41]</sup>) ; Harlen et James, (1997<sup>[42]</sup>) ; Harlen, (2007<sup>[43]</sup>) ; Black et al. (2003<sup>[44]</sup>), est que ces deux types d'évaluation sont liés et doivent être pensés en cohérence. Toutefois, l'articulation entre l'évaluation formative et l'évaluation parfois qualifiée de sommative présente des difficultés pratiques que les systèmes éducatifs cherchent à résoudre, en particulier par une formation importante et adaptée des enseignants.

Notons encore qu'un consensus existe pour considérer que l'évaluation sommative peut être utilisée de manière formative. La réciproque ne s'applique pas dans tous les pays au même niveau ; l'évaluation développée pour servir une fonction sommative peut être utilisée de manière formative — quoique certaines évaluations formatives puissent ne pas être appropriées pour une utilisation sommative (Baird et al., 2014<sup>[45]</sup>).

## Évaluation diagnostique

L'UNESCO définit l'évaluation diagnostique comme « une évaluation visant à identifier les forces et les faiblesses d'un ou une apprenant(e) en vue de prendre les mesures nécessaires pour améliorer l'apprentissage » (UNESCO-IBE, 2013<sup>[38]</sup>). Elle est également utilisée avant le processus d'enseignement et d'apprentissage pour évaluer le niveau de préparation ou de réussite de l'apprenant.

L'évaluation diagnostique est un type d'évaluation formative, qui a souvent lieu au début d'une unité d'étude afin de trouver un point de départ. L'évaluation diagnostique peut également servir à identifier les élèves qui risquent d'échouer, à découvrir les sources de leurs difficultés d'apprentissage et à planifier les mesures à prendre. L'évaluation diagnostique cherche, au-delà (ou en deçà) des compétences manifestées par les élèves, les causes des erreurs ou des difficultés d'apprentissage. Si l'évaluation diagnostique est parfois distinguée de l'évaluation formative, cette étude considère l'évaluation diagnostique comme un aspect de l'évaluation formative.

## Évaluation et contrôle continu

Les évaluations continues relèvent de la prérogative des enseignants en classe et ont pour but d'identifier le niveau d'apprentissage de chaque élève (et parfois d'une classe ou d'un autre groupe) selon les standards et les normes (dont le programme scolaire). L'évaluation continue fonctionne avec des formes sommatives et/ou formatives (UNESCO-IBE, 2017<sup>[46]</sup>).

Aujourd'hui, de nombreux pays valorisent l'évaluation continue, sous le nom de *continuous assessment*, pour l'opposer aux tests standardisés dont l'abus a souvent été jugé contre-productif, notamment en Corée.

L'évaluation continue présente les caractéristiques suivantes (UNESCO-IBE, 2017<sup>[46]</sup>) :

1. Elle relève principalement de la responsabilité de l'enseignant de la classe.
2. Elle couvre généralement l'apprentissage de l'ensemble des matières scolaires.
3. Elle permet aux éducateurs d'explorer plus en profondeur les capacités des élèves à appliquer les leçons académiques, y compris l'exécution de tâches basées sur le contexte local des élèves et qui peuvent comporter des éléments « pratiques » et s'étendre dans le temps.
4. Elle peut être très précise, nuancée et complète dans l'évaluation et l'information sur la façon dont évoluent les compétences personnelles des élèves, y compris les savoir-faire, les attitudes et les comportements tels que la collaboration, la créativité, la pensée critique, la confiance, la persévérance, la curiosité, la planification, et bien d'autres.

5. Elle se concentre sur les élèves individuels et les salles de classe d'une manière qui se rapporte typiquement aux décisions, et aux actions qui affectent directement les apprenants, que ce soit immédiatement ou à plus long terme.

Au Maroc comme en France, le contrôle continu est depuis longtemps la norme au niveau de l'enseignement, indépendamment de l'existence d'examens. Il s'agit tout simplement de l'évaluation menée traditionnellement par les enseignants dans leur classe en fonction de leurs besoins ou des demandes de l'administration. Cependant, au Maroc, le contrôle continu est surtout lié à la certification. Ses aspects plus formatifs sont souvent relégués au second plan.

## Évaluation formative

Selon l'OCDE, l'évaluation formative de l'élève, ou évaluation de l'apprentissage, vise à identifier les aspects de l'apprentissage au moment où il se développe afin d'approfondir et de déterminer l'apprentissage ultérieur (OCDE, 2013<sup>[30]</sup>).

D'un point de vue pédagogique et didactique, cette partie de l'évaluation, nommée ainsi ou autrement (en anglais, par exemple, *ongoing assessment*), parfois même non nommée comme un allant de soi de l'action pédagogique, fait intégralement partie du processus d'enseignement et en constitue une composante essentielle.

Son objectif n'est pas de contrôler des acquis, mais d'accompagner l'élève dans ses apprentissages en lui fournissant des indications pour l'aider à surmonter les obstacles. Elle fournit aussi à l'enseignant les informations lui permettant d'adapter son enseignement aux besoins de l'élève.

L'étude de l'OCDE (2018<sup>[37]</sup>) sur le Maroc rappelle que : « Au niveau international, l'évaluation formative a tendance à ne pas être accompagnée de notes chiffrées, de sorte que les élèves ne craignent pas de commettre des erreurs et de révéler des lacunes dans leurs connaissances, ce qui constitue un élément clé du processus d'apprentissage. » (Maghnoij et al., 2018<sup>[34]</sup>).

De ce fait, l'évaluation formative est en général différenciée et, généralement, ne conduit pas à des notes pérennes dans la mesure où l'action pédagogique qui la suit a justement pour objectif de traiter les difficultés observées. Moyennant quelques précautions, elle peut cependant, pour une part, transiter vers la certification. C'est en particulier le cas pour les compétences transversales dont les compétences personnelles et comportementales (les *soft skills*) et pour les compétences disciplinaires qui restent difficilement évaluables, comme la maîtrise de l'oral.

Notons que tous les types d'évaluation sont appelés à contribuer à l'évaluation formative, ne serait-ce que par le retour (*feedback*) qui devrait en être fait vers les enseignants et les élèves. Cela est en particulier valable pour les évaluations nationales et internationales.

## Évaluation interne et évaluation externe

L'évaluation faite en classe est un dispositif d'évaluation interne par rapport à la classe s'il est conçu et organisé par l'enseignant, de préférence en concertation avec ses collègues. Elle est externe si les épreuves ou tests sont imposés par un organisme extérieur (par exemple, l'administration), même si ces tests ou épreuves sont passés en classe sous la conduite de l'enseignant.

L'évaluation externe est définie comme un « processus et une méthode d'évaluation élaborés et utilisés par un organisme ou une agence d'examen autre que l'établissement scolaire de l'apprenant » (UNESCO-IBE, 2013<sup>[38]</sup>). Cela prend souvent la forme de tests standardisés dont le but est de noter les candidats en vue d'autres possibilités d'éducation et/ou à des fins de certification.

L'évaluation est dite interne lorsqu'elle est organisée dans le cadre d'un dispositif concerné par l'évaluation, en général la classe ou l'établissement (ce qui ne signifie pas qu'elle ne doit pas respecter certaines normes édictées par l'institution). En particulier, l'évaluation formative des élèves est interne à la classe. De ce point de vue, le contrôle continu au Maroc est un dispositif d'évaluation interne.

L'évaluation interne dite continue (ou le contrôle) ne porte pas nécessairement sur les mêmes compétences que l'évaluation externe (examens ou évaluations nationaux). Plusieurs des curricula étudiés reconnaissent en effet que l'évaluation externe ne peut pas observer les mêmes compétences que l'évaluation faite en classe par les enseignants (Hong Kong, Singapour notamment). Cela est particulièrement vrai dans le cas d'un curriculum basé sur les compétences.

La recherche met en évidence la complémentarité possible des deux modes d'évaluation et justifie une partie des différences (observées dans tous les pays) par des divergences dans les éléments évalués (Al-Maskari, 2015<sup>[47]</sup>). Il est souvent instructif de comparer les résultats de ces deux modes d'évaluation et de chercher à expliquer les différences observées. Il ne faudrait pas, pour autant, chercher à les réduire à tout prix, ce qui reviendrait à réduire la portée du curriculum.

## Évaluation sommative

Selon l'OCDE, l'évaluation sommative de l'élève, ou évaluation de l'apprentissage, vise à résumer l'apprentissage qui a eu lieu, afin d'enregistrer, de noter ou de certifier les réalisations (OCDE, 2013<sup>[30]</sup>). Plus généralement, on parle d'évaluation sommative lorsque l'on cherche à évaluer un champ de connaissances, de savoir-faire ou de compétences. Dans ce cas, ce qui détermine le caractère sommatif de l'évaluation est l'ampleur de l'objet sur lequel porte l'évaluation ; par exemple, un cours entier, un domaine d'étude, une compétence générale.

Bien que ne s'opposant pas à l'évaluation formative par le biais de l'exploitation des résultats avec les élèves, l'évaluation sommative intervient en général dans la partie de l'évaluation qui suit l'élève et qui a des enjeux (passages de classe, orientation scolaire et examens). Dans certains cas, le terme d'évaluation sommative est réservé à l'évaluation « qui compte », c'est-à-dire qui contribue à produire des notes ou des bilans qui vont jouer un rôle pour l'avenir des élèves. C'est cette acception que nous retiendrons dans cette étude.

## Examens

Dans les systèmes scolaires, les examens sont constitués « d'observations ou d'épreuves destinées à apprécier le savoir ou la culture des candidats en vue de la délivrance d'un diplôme, de l'entrée dans un établissement scolaire ou de l'aptitude à une fonction déterminée » (Foulquie, 1971<sup>[48]</sup>).

Seuls sont concernés par notre étude les examens de fin de cycle collégial dans la mesure où ils intègrent, au moins en partie, les évaluations internes menées dans le cadre de la classe, ce qui, au Maroc, est le cas de l'examen de fin de 3<sup>e</sup> année secondaire.

## Fiabilité

Parmi les qualités attendues d'une évaluation, la fiabilité concerne à la fois la validité et la fidélité (Rémond, 1980<sup>[1]</sup>), (Hivon, 1989<sup>[2]</sup>). En ce qui concerne l'évaluation sommative, la fiabilité est essentielle en raison des enjeux élevés de la progression des élèves. Elle ne peut se passer de normes précises et du contrôle.

La fidélité signifie que les résultats du test sont cohérents et stables dans différentes situations de test. Le degré de fidélité d'un instrument peut être affecté par un certain nombre de facteurs. Une évaluation hautement fidèle garantit que l'évaluation est précise et n'est pas influencée par l'évaluateur ou les conditions de l'évaluation. Les tests standardisés sont appréciés pour leur fidélité, particulièrement si le codage est automatisé.

La validité consiste à évaluer vraiment ce que l'on prétend évaluer. La validité concerne la pertinence des déductions, des utilisations et des conséquences de l'évaluation. Une évaluation hautement valide garantit que tous les aspects pertinents de la performance de l'élève sont couverts par l'évaluation. La validité est plus facilement assurée par des épreuves à questions ouvertes.

Cela est d'autant plus vrai lorsque l'on veut évaluer des compétences et pas seulement des connaissances ou des savoir-faire. Dans ce cas, un test en QCM, voire une épreuve en question ouvertes, mais courtes, suffit rarement aux besoins. Plusieurs pays comme la Corée, Hong Kong ou l'Irlande ont ainsi diminué l'usage des tests standardisés pour laisser plus de place à l'évaluation conduite dans les classes par les enseignants (OCDE, non publié<sup>[49]</sup>).

Notons que le mot « fiabilité » n'est pas un mot technique de l'évaluation. Toutefois, c'est le terme utilisé au Maroc et il résume bien les soucis de validité et de fidélité.

## Modération — harmonisation

La modération est utilisée pour s'assurer que les enseignants portent des jugements cohérents sur les normes (par exemple, les programmes scolaires), afin que les résultats d'évaluation portés sur un élève soient précis, équitables et comparables à ceux portés sur tous les autres élèves (NERF, s.d.<sup>[50]</sup>).

L'un des inconvénients de l'évaluation laissée aux enseignants et/ou aux établissements est le manque de fiabilité des notes ou des attributions de niveaux. Dans de nombreux pays où les résultats de l'évaluation en classe interviennent dans l'examen ou la certification, des dispositifs sont mis en place pour assurer la fiabilité des notes.

Il existe de nombreux instruments de modération, allant d'activités informelles (par exemple, une discussion avec un(e) assistant(e) d'enseignement sur le travail d'un(e) élève) à un processus externe formel (par exemple, des modérateurs des autorités locales examinant les jugements des enseignants en collectant des échantillons d'évaluations notées ou en menant des études statistiques).

Les enseignants et les autres professionnels de l'éducation doivent travailler ensemble, avec le soutien des activités de modération externe organisées par l'autorité locale. Dans certains cas, des commissions composées d'enseignants (éventuellement d'autres établissements) et d'inspecteurs



sont chargées d'harmoniser les critères et d'intervenir dans la régulation du jugement. Dans certains pays, ce rôle est confié à un organisme indépendant.

## Notes, échelles numériques, échelles catégorielles

Une note est définie comme l'appréciation d'un travail (un devoir par exemple) ou d'un comportement, exprimée conventionnellement par un chiffre ou une lettre. Noter un produit ou les résultats d'un élève, c'est, traditionnellement, le placer sur une échelle définie a priori.

L'échelle peut être numérique (échelle de 0 à 20, 0 à 10, 1 à 5, etc.) : c'est la notation traditionnellement utilisée en France et, plus généralement, dans les pays latins. L'échelle peut être littérale, par exemple A, B, C, D, E, comme dans de nombreux pays, notamment les pays scandinaves et Singapour. Des systèmes utilisent conjointement les deux types d'échelles.

L'échelle des notes est en général associée à une échelle de valeurs. Ainsi, en France, dans l'enseignement secondaire, la note 15 est associée à « bien », la note 10 à « moyen », etc. Noter, c'est donc décider de la valeur par rapport à des attentes.

Elle peut aussi être de nature qualitative ou catégorielle. Dans ce cas, l'échelle est en général composée d'un ensemble de niveaux qui, chacun, porte une information différente sur le niveau atteint par l'élève par rapport à une norme donnée. L'échelle peut même ne pas être linéaire, comme c'est le cas pour les grilles de correction qui ne présupposent pas un parcours unique.

Lorsque l'analyse d'une production d'élève conduit à repérer des indices de réalisation de critères prédéfinis, comme cela se passe souvent en évaluation formative ou en recherche, on parle de codage. Autrement dit, noter n'est pas la seule façon de rendre compte du contenu d'une production (devoir, test, performance écrite ou orale, etc.). L'analyse d'une production peut se réduire au codage et donc au repérage de ses qualités et défauts. Le codage peut aussi être une étape intermédiaire pour la production d'une note.

Les recherches docimologiques (étude de la fiabilité des notes) ont depuis près d'un siècle montré le peu de fiabilité des notes numériques, peu valides et peu fidèles. Elles ont aussi montré que le recours à des standards et à des barèmes précis pour la notation n'améliore que partiellement la situation. Selon des recherches plus récentes, les échelles catégorielles ne résolvent pas totalement la question de la fidélité (Leclercq, Nicaise et Demeuse, 2004<sup>[51]</sup>).

Au Maroc, au collège, les épreuves de contrôle continu sont notées sur une échelle de 0 à 20.

## Outils d'évaluation des élèves

Du classique questionnaire composé de QCM au portfolio contenant des traces de travaux effectués tout au long de l'année, en passant par l'observation directe, l'interrogation orale et la rédaction d'un texte, les outils utilisés pour évaluer les acquis des élèves sont extrêmement variés.

Les outils classiques sont bien connus et restent utiles : tests, devoirs et épreuves de tous types, quiz, interrogations orales et tests standardisés. Cependant, la volonté d'évaluer des compétences et pas seulement des connaissances et des savoir-faire amène à varier encore davantage les instruments d'évaluation et à imaginer des moyens inédits en fonction de la compétence à évaluer.

Comment par exemple évaluer l'esprit critique, la créativité, la capacité à travailler en groupe, le savoir apprendre, la citoyenneté ? Cela suppose une problématisation des besoins par les enseignants, l'expérimentation de nouvelles situations d'évaluation, l'analyse critique des résultats. De plus, un outil donné ne se prête pas toujours aussi bien à l'évaluation formative qu'à l'évaluation sommative, mais cela dépend surtout du contexte et de l'accompagnement qui lui est donné.

## Portfolio

La technique des portfolios se généralise dans de nombreux systèmes éducatifs. Le portfolio d'un(e) élève contient un ensemble de travaux que l'élève a effectué au cours de l'année.

L'élève peut ainsi présenter son portfolio à une personne ou à un jury ; le portfolio devient alors un élément de l'évaluation sommative.

Mais le portfolio est surtout un instrument d'évaluation formative. Il favorise le dialogue de l'élève avec l'enseignant autour de ses apprentissages. Objet d'échanges entre élèves, il favorise l'autoévaluation comme l'évaluation mutuelle.

## Score

Les tests standardisés comme les quiz, très utilisés dans l'évaluation en classe, ne produisent pas des notes, mais des scores, exprimés en pourcentage de réussite. Rien n'empêche de les convertir en note si nécessaire, mais on doit alors convenir d'une méthode de conversion.

## Standards et normes

Ces deux termes sont synonymes. En français, on utilise plutôt « norme », mais l'usage de « standard » se généralise. D'une façon générale, les standards sont des documents officiels définissant les qualités que doit posséder un produit ou une action : standards ou normes pour l'évaluation ou le contrôle, pour la formation, pour les manuels scolaires, etc.

S'agissant de l'évaluation scolaire, les standards (ou normes) sont les documents officiels précisant ce que les élèves doivent savoir et être capables de faire à chaque niveau. Lorsque de tels standards existent, le contrôle des acquis des élèves se fait alors en référence à ces standards.

Le concept de progression de l'apprentissage est lié à celui de norme. Selon l'OCDE, « les progressions de l'apprentissage décrivent la manière dont les élèves progressent habituellement



dans l'apprentissage de différentes matières. Elles peuvent servir de feuille de route aux enseignants pour identifier l'ensemble des compétences et des connaissances que les élèves doivent maîtriser pour devenir compétents dans des domaines plus complexes du programme scolaire. Ces progressions d'apprentissage peuvent être décrites dans le programme scolaire ou dans un document distinct » (OCDE, 2013<sup>[30]</sup>).

De nombreux pays de l'OCDE ont élaboré des normes d'apprentissage pour aider les enseignants à développer des évaluations fiables et, par conséquent, à émettre des jugements plus fiables. Des exemples notables de normes d'apprentissage peuvent être consultés sur les sites officiels de la Nouvelle-Zélande, du gouvernement californien, de l'Écosse et de l'Irlande.

Une norme étant donnée, il reste à l'enseignant à produire ou à choisir une situation d'évaluation (en général non réduite à un item) qui soit valide par rapport à cette norme. Cette action, qui peut être complexe, est ce que l'on appelle « opérationnalisation de la norme » (ou de l'objectif).

## Taxonomie

Une taxonomie est un système de classification hiérarchisé qui peut être utilisé pour le classement des objectifs pédagogiques (c'est le cas de la taxonomie de Bloom, bien connue) ou pour déterminer la complexité cognitive des tâches proposées aux élèves. Certains curricula ou instructions concernant l'évaluation en classe proposent ou imposent l'utilisation d'une taxonomie. C'est le cas du Maroc pour les contrôles continus.

La taxonomie de Bloom a été reconnue peu adaptée aux disciplines scientifiques. La taxonomie SOLO (Structure of the Observed Learning Outcomes) est très utilisée en Australie et en Nouvelle-Zélande. La taxonomie Gras-Bodin (Bodin, 2007<sup>[52]</sup>), particulièrement valide pour les mathématiques, est utilisée en France et dans d'autres pays de la francophonie.

L'utilisation d'une taxonomie, éventuellement associée à une table de spécification, oblige à diversifier les niveaux des demandes cognitives faites aux élèves et donc de ne pas rester aux plus bas niveaux (connaissance des faits, application de procédures).

Lorsque l'on élabore une situation d'évaluation, on peut décider a priori de la répartition des niveaux taxonomiques des items de l'évaluation, ou se voir imposer une telle répartition, comme cela se fait au Maroc. Cette répartition se fait dans un tableau que l'on nomme « table de spécification ».

## Test standardisé

Les tests standardisés sont, d'une part, construits et notés selon des normes précises ; d'autre part, ils portent sur des contenus bien définis tels que les programmes et les instructions officielles. Ces tests, construits de façon externe, peuvent être utilisés dans le cadre de l'évaluation en classe. Ils sont de plus en plus souvent passés sur ordinateur et corrigés sans intervention humaine.

Outre leur facilité de passation, de correction et leurs résultats immédiats, ces tests ont l'avantage d'être parfaitement fidèles et d'assurer une harmonisation tout aussi parfaite entre les classes et établissements faisant passer le même test.



Cependant, ces tests peinent à assurer la validité, ce qui explique que beaucoup de pays qui en faisaient un usage intensif pour l'évaluation externe tendent maintenant à en limiter l'usage (États-Unis et pays asiatiques). En revanche, d'autres pays qui résistaient à cette forme d'évaluation commencent à les utiliser d'une façon mesurée. Dans les deux cas, il s'agit d'établir un équilibre entre l'évaluation faite par les enseignants et l'évaluation externe.

## Annex B. Édumétrie : pour aller plus loin

Tout au long de ce guide, nous nous sommes concentrés sur ce que l'enseignant peut faire pour améliorer la qualité, la fiabilité et l'équité d'une épreuve d'évaluation sommative telle que le contrôle continu.

Une fois que l'épreuve a été passée et que des données sur les performances des élèves sont disponibles, il est possible de vérifier si la qualité, la fiabilité et l'équité sont satisfaisantes.

Une série de procédures existent pour valider une épreuve, vérifier ses qualités édumétriques, sa cohérence, sa fidélité, le pouvoir discriminatif des items etc.

Ces procédures - nous insistons sur cela - ne peuvent s'appliquer aux contrôles continus dans leur forme actuelle : ceux-ci comportent bien trop peu de questions et n'ont été administrés qu'aux élèves d'une ou de quelques classes. Les conditions pour appliquer ces procédures ne sont donc pas remplies.

Un bref aperçu des procédures les plus courantes est fourni ci-après à titre informatif afin d'illustrer comment les responsables d'épreuves externes standardisées s'y prennent pour valider une épreuve d'évaluation.

### Mise à l'essai des items (prétest)

L'élaboration d'une épreuve standardisée obéit aux mêmes principes que ceux décrits au chapitre 2 de ce guide. Une étape essentielle est la réalisation d'un prétest auprès d'un échantillon suffisamment nombreux et diversifié d'élèves de façon à estimer un ensemble de paramètres statistiques associés à l'épreuve et à chacun de ses items, en particulier les indices de difficulté et de discrimination. L'objectif du prétest est donc d'évaluer la qualité des items et des grilles de correction alors que le test définitif sert à évaluer les acquis des élèves à l'aide de l'épreuve validée.

Les indices de discrimination reflètent la capacité de l'item à bien distinguer entre les performances des élèves plus ou moins forts (Rocher, 2015<sup>[53]</sup>) ; (IRDP, 2017<sup>[54]</sup>). Dans une approche classique, la difficulté d'un item est exprimée par le pourcentage de réponses correctes, ou taux de réussite (Rocher, 2015<sup>[53]</sup>). Cette approche comporte une sérieuse limite puisque l'estimation de la difficulté d'un item dépend des aptitudes de l'échantillon qui a passé l'épreuve. Pour arriver à estimer le degré de difficulté d'un item indépendamment de l'échantillon, il faut recourir à d'autres modèles édumétriques que l'approche classique, les modèles de la réponse à l'item (voir plus loin).

### Estimer la cohérence interne : l'alpha de Cronbach

Pour estimer la cohérence interne d'une épreuve (de son score total ou d'un sous-score), on utilise un indice appelé alpha de Cronbach pour les items concernant une compétence, une capacité ou un ensemble de compétences relatives à un même domaine de savoir.

Cet indice, qui, mesure la moyenne des corrélations entre items est classiquement utilisé pour voir si le test mesure bien la compétence, la capacité ou le domaine de connaissances qu'il est censé évaluer. La valeur de cet indice statistique varie entre 0 et 1. La cohérence interne d'un test est

d'autant plus élevée que la valeur prise par l'indice est proche de 1. Pour considérer que les items groupés d'un test cognitif mesurent bien une même compétence (on dit aussi trait ou construct), la valeur du coefficient doit être au moins égale à 0,80. Pour des échelles autres (contextuelles, par exemple), on peut être plus souple et accepter 0,70.

La valeur d'un alpha est affectée par le nombre d'items. On ne calcule pas d'alpha sur des épreuves cognitives comportant peu d'items.



## 1. Exemple 1

Scores à l'ensemble du test et par domaine

	Score moyen	Alpha
<b>Ensemble du test (122 items)</b>	<b>48%</b>	0.95
<b>Domaines des nombres (75 items)</b>	<b>50%</b>	0.93
<b>Domaines des grandeurs (47 items)</b>	<b>44%</b>	0.89

Les valeurs de l'épreuve (0.95) et pour les deux grands domaines qui la composent, les nombres (0.93) et les grandeurs (0.89).

Sous-scores par compétence

	Score moyen	Alpha
Calculer : <i>construire des tables</i> (21 items)	<b>57%</b>	0.85
Calculer : <i>identifier et effectuer des opérations dans des situations variées</i> (25 items)	<b>50%</b>	0.85
Opérer, fractionner : <i>résoudre des problèmes simples de proportionnalité directe</i> (12 items)	<b>37%</b>	0.77
Comparer, mesurer : <i>construire et utiliser des démarches pour calculer des périmètres, des aires et des volumes</i> (17 items)	<b>36%</b>	0.72

En ce qui concerne les sous-scores par compétence, les alphas sont très bons pour les deux premières sous-compétences, mais en dessous de 0.80 pour les deux dernières. On remarquera que celles-ci comportent moins d'items (12 et 17 items). Ceci montre la difficulté de construire des scores valides au départ d'un nombre limité d'items.

Source : Développé par les auteurs du guide dans le cadre d'un prétest. Non diffusé.

### **Le pouvoir discriminatif des items**

Le  $r_{bis}$  ou corrélation bisériale de point est l'indice classiquement utilisé pour apprécier le pouvoir discriminatif des items. Il varie entre +1 et - 1.

Cette statistique permet de vérifier si les élèves les meilleurs au test ont majoritairement choisi la réponse correcte à chacun des items. Il est logique que les bons élèves à l'ensemble du test

fournissent davantage la réponse correcte à chaque item que les élèves peu performants. Si à un item, les élèves les plus performants au test choisissent massivement un distracteur (réponse erronée), ou omettent de répondre, c'est un signe clair qu'il y a un problème dans la question.

Pour considérer qu'un item fonctionne bien, toujours pour la compétence visée, il faut qu'indépendamment du résultat, le r.bis de la réponse correcte soit supérieur à 0,20 et que les r.bis des réponses ou propositions incorrectes soient négatifs, proches de zéro ou au minimum bien contrastés par rapport à la réponse correcte. Quand le r.bis de la réponse considérée comme correcte est en dehors des valeurs acceptables, cela signifie, entre autres, que les élèves qui réussissent bien à l'ensemble de l'épreuve n'ont pas choisi la réponse correcte à l'item considéré.

L'examen du r.bis permet donc de détecter une incohérence entre le résultat à un item donné et l'ensemble du test, ainsi que d'analyser la qualité des solutions proposées.

Associé aux pourcentages d'élèves qui ont choisi chaque proposition, le r.bis aide à détecter un problème de formulation ou de correction (proposition erronément renseignée comme correcte) et peut conduire à la décision de supprimer une question, ou à la reformuler.



## 2. Exemple 2 QCM

### Exemples d'analyse de questions à l'aide du r.bis

Voici un exemple d'un item (QCM), avec le pourcentage des réponses pour chaque possibilité de réponse et sa valeur r.bis correspondante.

Quatre amis vont au parc d'attractions. L'entrée coûte 15 €. Deux d'entre eux possèdent un bon de réduction de 5 €. Que paient-ils ?	<input type="checkbox"/> $(4 \times 15 \text{ €}) - (2 \times 5 \text{ €})$
	<input type="checkbox"/> $(15 \times 4 \text{ €}) + (2 \times 5 \text{ €})$
	<input type="checkbox"/> $(15 \text{ €} : 4) - (1 \times 2 \text{ €})$
	<input type="checkbox"/> $(4 \times 15 \text{ €}) + (2 \times 5 \text{ €})$

Source : (Fédération Wallonie-Bruxelles, s.d.<sup>[6]</sup>)

It	Parc d'attraction	% d'élèves	r.bis
31			
1	$(4 \times 15 \text{ €}) - (2 \times 5 \text{ €})$	58%	0.44
2	$(15 \times 4 \text{ €}) + (2 \times 5 \text{ €})$	11%	-0.22
3	$(15 \text{ €} : 4) - (1 \times 2 \text{ €})$	13%	-0.29
4	$(4 \times 15 \text{ €}) + (2 \times 5 \text{ €})$	13%	-0.05
5	Plusieurs réponses	1%	-0.11
9	Omission	4%	-0.12

Source : Développé par les auteurs du guide dans le cadre d'un prétest. Non diffusé.

Le tableau de résultats ci-dessus donne les pourcentages de réussite et les valeurs du r.bis pour la réponse correcte, les distracteurs, les élèves qui ont coché plusieurs cases et ceux qui n'en ont coché aucune. Dans ce cas, le r.bis de la réponse correcte est de 0.44, supérieur donc à la valeur seuil de 0.20. Les r.bis des autres réponses sont tous négatifs. L'item fonctionne bien et a un bon pouvoir discriminatif.



### 3. Exemple 3 QCM

Le tableau de résultats ci-dessous, extrait de la préparation d'un test standardisé portant sur l'addition, indique que non seulement la réponse correcte n'est choisie que par 5 % des élèves, mais surtout, le r.bis de la réponse correcte (0,07) est très largement insuffisant, ce qui a conduit à écarter cette question. Le pourcentage d'omission (43 %) est, lui aussi, inquiétant. Ces résultats sont sans doute liés à la présence d'une négation dans la question ("il n'est pas utile").

It 63	Entoure l'opération où il n'est pas utile d'associer des termes	% d'élèves	r.bis
1	$2361 + 339 + 956$	27%	0.05
2	$3254 + 437 + 534$	5%	0.07
3	$11,7 + 3,4 + 8,3 + 6,6$	25%	-0.01
9	Omission	43%	-0.06

Source : Développé par les auteurs du guide dans le cadre d'un prétest. Non diffusé.



### 4. Exemple 4 question ouverte

Les r.bis peuvent aussi être utilisés dans le cas de questions ouvertes, comme dans l'exemple ci-dessous où des codes ont été attribués aux différentes réponses fournies par les élèves (1 pour la réponse correcte et de 2 à 5 pour les erreurs les plus fréquentes). Le tableau montre un r.bis de 0,62 pour la réponse correcte et des r.bis largement négatifs pour les autres modalités de réponse. Il s'agit donc d'une bonne question, dont le pouvoir discriminatif est excellent.

Item 6	Quel personnage raconte l'histoire ?	% d'élèves	r.bis
1	Agnès, la petite fille	42 %	0,62
2	Nicolas Ancion	25 %	-0,20
3	Le professeur de musique	6 %	-0,12
4	Le directeur	2 %	-0,17
5	Jason et/ou Nathan	8 %	-0,11
	Autre	12 %	-0,27
9	Omission	5 %	-0,22

Source : Développé par les auteurs du guide dans le cadre d'un prétest. Non diffusé.

## Calculer



Un simple tableur comme Microsoft Excel, permet de faire de nombreux calculs utiles pour le traitement de données d'évaluation. Outre les paramètres classiques, le point biserial peut être facilement calculé, le calcul de l'alpha de Cronbach suppose juste de savoir se servir des formules du tableur. Rappelons que ces techniques ne sont pas utiles pour l'évaluation des élèves au niveau de la classe. Elles n'intéresseront que les enseignants qui décideraient de construire un test standardisé destiné à un regroupement de classes ou qui seraient associés à l'élaboration d'un tel test.

## Les modèles de la réponse à l'item (IRT)

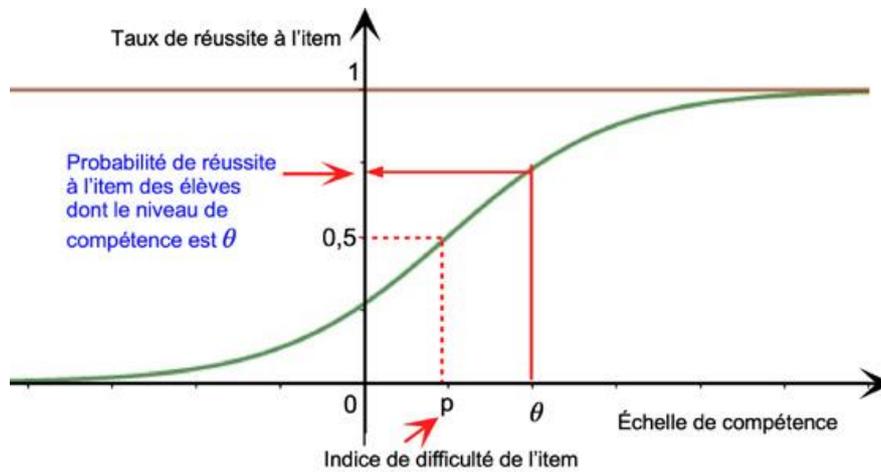
Les techniques de l'analyse des réponses à l'item (Laveault et Grégoire, 2014<sup>[55]</sup>) reposent sur des modélisations probabilistes complexes. Ces modèles décrivent une fonction mathématique qui établit la probabilité d'une réponse correcte à un item comme étant le résultat de deux facteurs : l'aptitude du sujet (considérée comme variable) et les paramètres de l'item (considérés comme fixes).

Elles sont bien adaptées à l'évaluation des compétences pour autant qu'on dispose de données récoltées lors d'un prétest ou d'un test passé par un échantillon d'élèves de taille assez importante (ici encore, un échantillon d'au moins 200 sujets est nécessaire). L'analyse, comme dans le cas de l'analyse classique, permet d'obtenir directement les valeurs de divers paramètres des items et du test lui-même (comme par exemple, les indices de difficulté et discrimination). Dans l'IRT, un item particulier est toujours associé au test dont il est un élément et à la population à laquelle le test est destiné.

L'IRT produit en particulier des courbes de réponses aux items. L'analyse de ces courbes est particulièrement utile ; elle permet en effet de comparer, non pas le comportement des élèves, mais le comportement des items d'un test donné.

La figure ci-dessous montre une telle courbe de réponse. Elle met en évidence la façon dont est défini l'indice de difficulté de l'item (niveau moyen de compétence des élèves ayant une probabilité de 0,5 de réussir à l'item) et la probabilité de réussite des élèves dont le niveau de compétence est  $\theta$ .

### Graphique B-1 Exemple d'une courbe de réponse de l'IRT



Courbe de réponse d'un item

Source : Développé par les auteurs du guide

### Notes

1 En France, le test officiel PIX en ligne pour évaluer, développer et certifier les compétences numériques est un test adaptatif (<https://pix.fr/>).

# Références

- Allal, L. (1999), « Impliquer l'apprenant dans le processus d'évaluation: promesses et pièges de l'auto-évaluation », dans Depover, C. et B. Noël (dir. pub.), *L'évaluation des compétences et des processus cognitifs : modèles, pratiques et contextes*, De Boeck & Larcier. [34]
- Allal, L. (1988), « Peut-on instrumenter l'auto-évaluation? », *Les démarches et instruments de la recherche en évaluation : mise en question(s)*, p. 132. [32]
- Al-Maskari, A. (2015), *Comparison between continuous assessment and final score*, Conference Paper, [https://www.researchgate.net/profile/Azzah-Al-Maskari/publication/327231273\\_Comparison\\_between\\_continuous\\_assessment\\_and\\_final\\_score/links/5b827e3892851c1e123355a6/Comparison-between-continuous-assessment-and-final-score.pdf](https://www.researchgate.net/profile/Azzah-Al-Maskari/publication/327231273_Comparison_between_continuous_assessment_and_final_score/links/5b827e3892851c1e123355a6/Comparison-between-continuous-assessment-and-final-score.pdf). [48]
- Anderson, L. et D. Krathwohl (dir. pub.) (2001), *A Taxonomy for Learning, Teaching, and Assessing : A Revision of Bloom's Taxonomy of Educational Objectives*, Longman. [8]
- Antibi, A. (2003), *La constante macabre ou comment a-t-on découragé des générations d'élèves*, Math'Adore. [28]
- Armstrong, P. (2010), *Bloom's Taxonomy*, Vanderbilt University Center for Teaching, <https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/> (consulté le 13 Juillet 2022). [10]
- Baird, J. et al. (2014), *Assessment and Learning : State of the Field Review*, Norwegian Knowledge Center for Education, [https://www.researchgate.net/publication/263654863\\_Assessment\\_and\\_Learning\\_State\\_of\\_the\\_Field\\_Review](https://www.researchgate.net/publication/263654863_Assessment_and_Learning_State_of_the_Field_Review). [46]
- Bennett, R. (2011), « Formative assessment : A critical review », *Assessment in Education: Principles, Policy & Practice*, vol. 18/1, pp. 5-25, <https://doi.org/10.1080/0969594X.2010.513678>. [42]
- Black, P., C. Harrison et C. Lee (2003), *Assessment for Learning : Putting It into Practice*, Open University Press. [45]
- Bodin, A. (2007), « Dissonances et convergences évaluatives », *Bulletin de l'APMEP*, n° 474, Association des Professeurs de Mathématiques de l'Enseignement Public. [53]

Bressoux, P. et P. Pansu (2003), *Quand les enseignants jugent leurs élèves*, Presses Universitaires de France. [25]

Centre national de l'évaluation et des examens (Non publié), *Examen normalisé régional pour l'obtention du certificat du cycle collégial*, Cadre de référence de la langue Française. [4]

Crahay, M. (2019), *Peut-on lutter contre l'échec scolaire ?*, De Boeck. [26]

De Landsheere, G. (1980), *Evaluation continue et examens : Précis de docimologie*, Labor. [27]

Dusek, J. et G. Joseph (1983), « The bases of teacher expectancies : A metaanalysis », *Journal of Educational Psychology*, vol. 31/4, pp. 674-685. [24]

Eisner, E. (1972), « Review of Handbook on Formative and Summative Evaluation of Student Learning, by B. S. Bloom, J. T. Hastings, & G. F. Madaus », *Studies in Art Education*, vol. 14/1, pp. 68-72, <https://doi.org/10.2307/1319918>. [40]

Fédération Wallonie-Bruxelles (s.d.), *Évaluations Externes Non Certificatives - Éveil/Sciences - Évaluations Antérieures.*, <http://enseignement.be/index.php?page=25186&navi=3208> (consulté le 26 Janvier 2023). [7]

Fédération Wallonie-Bruxelles (s.d.), *Évaluations Externes Non Certificatives - Formation Mathématique*, <http://enseignement.be/index.php?page=24761&navi=2030> (consulté le 27 Janvier 2023). [14]

Fédération Wallonie-Bruxelles (s.d.), *Évaluations Externes Non Certificatives - Formation Mathématique - Évaluations Antérieures*, <http://enseignement.be/index.php?page=25102&navi=3207#2017-2018> (consulté le 26 Janvier 2023). [6]

Fédération Wallonie-Bruxelles (s.d.), *Évaluations Externes Non Certificatives - Lecture Et Production d'Écrit*, <http://www.enseignement.be/index.php?page=24760&navi=2029#2021-2022> (consulté le 27 Janvier 2023). [13]

Fédération Wallonie-Bruxelles (s.d.), *Les Épreuves du CE1D de Français*, <http://www.enseignement.be/index.php?page=26834&navi=3450#2019> (consulté le 27 Janvier 2023). [16]

Foulquié, P. (1971), *Dictionnaire de la langue pédagogique*, Presses Universitaires de France. [49]

Gilles, J. (2006), *Introduction à la construction d'évaluations des acquis des apprenants*, [http://tecfa.unige.ch/tecfa/maltt/actu\\_recher/Seminaire%20TECFA%20220306/intro\\_construction\\_outils\\_evaluation.pdf](http://tecfa.unige.ch/tecfa/maltt/actu_recher/Seminaire%20TECFA%20220306/intro_construction_outils_evaluation.pdf) (consulté le 27 Janvier 2023). [29]

Grisay, A. (1984), « Les mirages de l'évaluation scolaire », *Revue de la direction générale des études*, vol. 19/5, pp. 29-42. [30]

- Harlen, W. (2007), *Assessment of Learning*, Sage Publications. [44]
- Harlen, W. et M. James (1997), « Assessment and learning : Differences and relationships between formative and summative assessment », *Assessment in Education : Principles, Policy & Practice*, vol. 4/3, pp. 365-379. [43]
- Hivon, R. (1989), « « Dictionnaire actuel de l'éducation » by Renald Legendre » », *Canadian Journal of Education / Revue canadienne de l'éducation*, vol. 14/2, pp. 265-267, <https://doi.org/10.2307/1495356>. [2]
- Huguet, P. et I. Régner (2007), « Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances », *Journal of Educational Psychology*, vol. 99/3, pp. 545-560, <https://doi.org/10.1037/0022-0663.99.3.545>. [21]
- Huitt, W. (2011), « Bloom et al.'s taxonomy of the cognitive domain », *Educational Psychology Interactive*, Valdosta State University, Valdosta, GA, <http://www.edpsycinteractive.org/topics/cognition/bloom.html> (consulté le 13 Juillet 2022). [11]
- IRD (2017), *Indice de discrimination (d'un item)*, Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin, <https://www.irdp.ch/institut/indice-discrimination-item-2116.html>. [55]
- IRD (2016), *Banque d'items*, Institut de recherche et de documentation pédagogique, Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin, <https://www.irdp.ch/institut/banque-items-2019.html>. [36]
- Lafontaine, D. et C. Monseur (2009), « Gender Gap in Comparative Studies of Reading Comprehension : to what extent do the test characteristics make a difference? », *European Educational Research Journal*, vol. 8/1, pp. 69-79, <http://hdl.handle.net/2268/10908C>. [20]
- Laveault, D. et J. Grégoire (2014), *Introduction aux théories des tests*, De Boeck, 3e édition. [56]
- Leclercq, D., J. Nicaise et M. Demeuse (2004), « Docimologie critique des difficultés de noter des copies et d'attribuer des notes aux élèves », dans Demeuse, M. (dir. pub.), *Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation*, Les Editions de l'Université de Liège, [https://iredu.u-bourgogne.fr/images/stories/Documents/Cours\\_disponibles/Demeuse/Cours/p6.4.pdf](https://iredu.u-bourgogne.fr/images/stories/Documents/Cours_disponibles/Demeuse/Cours/p6.4.pdf). [52]
- Maghnoij, S. et al. (2018), *Examens de l'OCDE du cadre d'évaluation de l'éducation : Maroc*, Éditions OCDE, Paris, <https://doi.org/10.1787/9789264301832-fr>. [35]
- Ministère de l'Éducation nationale du Maroc (2020), *Portefeuille des projets de déclinaison de la loi cadre 51.17*, <https://m.dahayas.com/wp-content/uploads/2021/02/Portefeuille-des-projets-de-declinaison-de-la-loi-cadre.pdf>. [3]
- Ministère de l'Éducation nationale du Maroc (2010), *Note 192. Note ministérielle sur le contrôle continu en mathématiques en niveau collège*. [5]

- NCTM (2000), *Principles and Standards for School Mathematics*, National Council of Teachers of Mathematics, <https://www.nctm.org/Standards-and-Positions/Principles-and-Standards/Principles,-Standards,-and-Expectations/>. [41]
- NFER (s.d.), *Moderation of Assessment Judgements*, National Foundation for Education Research, [https://www.nfer.ac.uk/media/3139/moderation\\_of\\_assessment\\_judgements.pdf](https://www.nfer.ac.uk/media/3139/moderation_of_assessment_judgements.pdf). [51]
- OCDE (2018), *Le futur de l'éducation et des compétences : Projet Éducation 2030*, OCDE, Paris, [https://www.oecd.org/education/2030-project/about/documents/French%20version%20-%20OECD%20Education%202030%20Position%20Paper\\_final%20\(07.06.2018\).pdf](https://www.oecd.org/education/2030-project/about/documents/French%20version%20-%20OECD%20Education%202030%20Position%20Paper_final%20(07.06.2018).pdf). [38]
- OCDE (2013), *PISA 2012 Released Items*, [18] <https://www.oecd.org/pisa/pisaproducts/PISA%202012%20released%20items%20FRE.pdf>.
- OCDE (2013), *Synergies for Better Learning : An International Perspective on Evaluation and Assessment*, OECD Reviews of Evaluation and Assessment in Education, Éditions OCDE, Paris, <https://doi.org/10.1787/9789264190658-en>. [31]
- OCDE (non publié), *Étude de benchmarking sur les évaluations en classe*. [50]
- OCDE (non publié), *Guide de l'Enseignant sur l'Évaluation Formative et la Différenciation*. [19]
- OCDE (non publié), *Rapport sur l'évaluation formative et la différenciation*. [57]
- Rakocevic, R. et Y. Fournier (2020), *L'Europe de l'éducation en chiffres*, [15] <https://www.education.gouv.fr/l-europe-de-l-education-en-chiffres-2020-304068>.
- Rémond, G. (1980), « « Dictionnaire de l'évaluation et de la recherche en éducation » de G. de Landsheere », *Communication et langages*, vol. 45/45. [1]
- Rocher, T. (2015), « Méthodes psychométriques utilisées dans le cadre des évaluations des élèves », *Éducation & Formations*, vol. 86-87, pp. 37-60, <https://doi.org/10.48464/ef-86-87-02>. [54]
- Rondal, J. et J. Broonen (2021), *Une réforme radicale de l'orthographe française ?*. [17]
- Spencer, S., D. Quinn et C. Steele (1999), « Stereotype threat and women's math performance », *Journal of Experimental Social Psychology*, vol. 35/1, pp. 4-28, <https://doi.org/10.1006/JESP.1998.1373>. [22]
- Steele, M. et J. Aronson (1995), « Stereotype threat and the intellectual test performance of African Americans », *Journal of Personality and Social Psychology*, vol. 69/5, pp. 797-811, <https://doi.org/10.1037//0022-3514.69.5.797>. [23]
- UE (2006), *Recommandation du parlement européen et du conseil sur les compétences clés pour l'éducation et la formation tout au long de la vie*, Union européenne, <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=celex%3A32006H0962>. [37]

UNESCO-IBE (2017), « Continuous assessment for improved teaching and learning : A [47]  
critical review to inform policy and practice », *Current and Critical Issues in Curriculum,  
Learning and Assessment* 13.

UNESCO-IBE (2013), *Glossary of Curriculum Terminology*, UNESCO International Bureau [39]  
of Education, [http://www.ibe.unesco.org/sites/default/files/resources/ibe-glossary-  
curriculum.pdf](http://www.ibe.unesco.org/sites/default/files/resources/ibe-glossary-curriculum.pdf).

Université de Poitiers (2014), *Taxonomie de Bloom et roue pédagogique*, [https://blogs.univ- \[9\]  
poitiers.fr/t-roy/2014/05/09/taxonomie-de-bloom-et-roue-pedagogique/](https://blogs.univ-poitiers.fr/t-roy/2014/05/09/taxonomie-de-bloom-et-roue-pedagogique/) (consulté le 13 Juillet  
2022).

University of Central Florida (s.d.), *Bloom's Taxonomy*, [https://fctl.ucf.edu/teaching- \[12\]  
resources/course-design/blooms-taxonomy/](https://fctl.ucf.edu/teaching-resources/course-design/blooms-taxonomy/) (consulté le 13 Juillet 2022).

Vial, M. (1997), « L'auto-évaluation: Entre auto-contrôle et auto-questionnement ». [33]

### **Pour plus d'informations**

OCDE (non publié), *Cadre de référence pour le contrôle continu au Maroc.  
Étude préparatoire*.

OCDE (non publié), *Étude de benchmarking sur les évaluations en classe*.