

# Évaluation et pédagogie d'investigation dans l'enseignement scientifique :

## De la politique à la pratique

**Wynne Harlen**

**Comité de rédaction:**

**Derek Bell, Jens Dolin, Pierre Léna, Shelley Peers,  
Xavier Person, Patricia Rowell and Edith Saltiel**

**Traduit de l'anglais par Zheng Ma**

**Édition française présentée par Edith Saltiel et Pierre Léna**

**IAP Global Network of Science Academies / Réseau mondial des Académies des sciences  
Programme Éducation à la science**

La démarche d'investigation et la pédagogie qui la met en œuvre auprès des élèves sont maintenant reconnues en France comme de solides supports pour un enseignement scientifique de qualité à l'école primaire et au collège, dans l'esprit du socle commun de connaissances, de compétences et de culture. Depuis une décennie, de très nombreux projets pilotes, de par le monde et dans des pays d'une grande diversité, se sont également engagés dans cette voie, où les élèves expérimentent, raisonnent, exercent imagination et créativité, développent leur usage de la langue et leurs capacités d'écoute de l'autre, prennent confiance dans leur intelligence quelle qu'en soit la forme. Il est unanimement reconnu que les modalités d'évaluation des élèves, que celle-ci soit formative ou sommative, ne peuvent être traités indépendamment de la pédagogie sur laquelle elles influent fortement. Le développement, en France, d'un enseignement fondé sur l'investigation requiert donc de porter une grande attention aux modalités d'évaluation. C'est le propos de ce petit ouvrage que d'apporter aux professeurs, aux formateurs et aux étudiants quelque éclairage sur celles-ci, en s'appuyant sur l'expérience acquise dans nombre de projets pilotes, tant dans l'espace anglo-saxon qu'en France autour de La main à la pâte, et ailleurs dans le monde.



# **Évaluation et pédagogie d'investigation dans l'enseignement scientifique :**

De la politique à la pratique

**Wynne Harlen**

Comité de rédaction :

**Derek Bell, Jens Dolin, Pierre Léna, Shelley Peers,  
Xavier Person, Patricia Rowell and Edith Saltiel**

Traduit de l'anglais par Zheng Ma  
Édition française présentée par Edith Saltiel et Pierre Léna

**Publié par**

**IAP Le Réseau mondial des Académies des sciences/**

**Programme Éducation à la science (Science Education Program SEP)**

[www.interacademies.net/activities/projects/12250.aspx](http://www.interacademies.net/activities/projects/12250.aspx)

TWAS-Strada Costiera, 11-34151, Trieste, Italy

ISBN : 978-1-291-58452-3

**© Wynne Harlen 2013**

Les copies et traductions peuvent être faites sans frais ni autorisation préalable à condition de mentionner dûment leur provenance

# Évaluation et pédagogie d'investigation dans l'enseignement scientifique:

## De la politique à la pratique

### Sommaire

#### Préface

<b>Introduction</b> _____	<b>1</b>
Le contexte du livre _____	1
Les points principaux dégagés lors de la Conférence _____	2
L'objet du livre _____	5
<b>1 Clarification des termes</b> _____	<b>7</b>
Évaluation (des élèves), évaluation (systémique) et appréciation (des éducateurs) _____	7
Les tests et autres méthodes d'évaluations _____	8
La validité, la fiabilité, les ressources et la capacité de gestion _____	9
<b>2 La pédagogie d'investigation dans l'enseignement scientifique: motif et objectifs</b> _____	<b>13</b>
L'investigation dans l'enseignement des sciences _____	13
Le motif _____	15
Les objectifs _____	16
<b>3 Les objectifs et les utilisations de l'évaluation des élèves</b> _____	<b>19</b>
Les objectifs et les utilisations _____	19
L'évaluation formative des élèves _____	20
L'évaluation sommative des élèves _____	24
La relation entre l'évaluation formative et l'évaluation sommative des élèves _____	26
<b>4 L'évaluation des élèves, la pédagogie et les programmes</b> _____	<b>31</b>
Les impacts de l'évaluation des élèves _____	31
La responsabilisation _____	35
L'évaluation des élèves et les théories de l'apprentissage _____	37
Le rôle de l'évaluation des élèves dans la saisie des grands problèmes mondiaux _____	40

<b>5</b>	<b>La mise en œuvre d'une évaluation formative des élèves pratiquant l'ESFI</b>	<b>43</b>
	L'évaluation formative des élèves et l'ESFI	43
	Le dialogue en classe	44
	Les questions des enseignants	48
	Les commentaires fait aux élèves	51
	Les commentaires dans l'enseignement	53
	L'auto-évaluation de l'élève et l'évaluation par les pairs chez les élèves	54
<b>6</b>	<b>La mise en œuvre de l'évaluation sommative des élèves dans l'ESFI</b>	<b>57</b>
	L'évaluation de la compréhension et des compétences en investigation	57
	Les méthodes d'une évaluation sommative des objectifs de l'ESFI	59
	Le potentiel et les limites des tests pour évaluer les objectifs de l'ESFI	60
	Les alternatives aux tests	70
	La mise en œuvre d'une évaluation par les enseignants	71
	Amélioration de la fiabilité d'une évaluation réalisée par les enseignants	73
	Quelques exemples d'évaluation des élèves reposant sur les enseignants	76
	Les rapports issus de l'évaluation sommative des élèves: le pour et le contre de l'utilisation de niveaux	80
	Les avantages et les inconvénients des tests et d'une évaluation faite par les enseignants	82
<b>7</b>	<b>Changer les pratiques de l'évaluation des élèves</b>	<b>83</b>
	Les modalités du changement de pratiques dans l'enseignement	83
	Quelques exemples de changement de pratiques dans l'évaluation des élèves	85
	Les implications de l'évaluation des élèves au sein de l'ESFI	92
	<b>Tableaux</b>	
1	Les principes de la pratique de l'évaluation des élèves en classe	95
2	Les principes destinés aux équipes de gestion de l'école	96
3	Les principes destinés aux inspections locales et nationales	97
4	Les principes destinés à la formulation des politiques nationales	98
	<b>Bibliographie de la version originale en anglais</b>	<b>99</b>
	<b>Bibliographie ajoutée pour l'édition française</b>	<b>103</b>

**Encadrés**

1	Les pays représentés à la conférence d'Helsinki _____	2
2	Le compromis entre la fiabilité et la validité _____	11
3	Les activités des élèves: apprentissage par l'investigation _____	18
4	Les activités des élèves: apprentissage par les méthodes de transmission _____	18
5	Les composants essentiels de la mise en oeuvre d'une évaluation formative des élèves _____	23
6	Les composants essentiels de la mise en oeuvre d'une évaluation sommative des élèves _____	26
7	Une recherche systématique portant sur l'impact de tests à grande échelle _____	33
8	Les écoles rendent compte de leurs activités _____	36
9	Un exemple d'apprentissage des sciences par l'investigation _____	39
10	Expliquer l'effet de la pression de l'air _____	45
11	Dialogue à propos d'œufs _____	47
12	Les questions pour découvrir et encourager le développement des idées _____	49
13	Les questions pour découvrir et encourager le développement des compétences d'investigation _____	49
14	Les questions pour encourager la collaboration, le partage des idées, la réflexion et l'évaluation _____	50
15	Communiquer aux élèves de l'école primaire les critères d'évaluation de leur travail _____	55
16	L'utilisation de l'auto-évaluation par des élèves du secondaire, afin d'améliorer leur travail _____	56
17	L'inégalité des chances dans les tests _____	67
18	Les sources de questions et de tâches utilisées dans des tests d'évaluation _____	68
19	Exemple de critères d'évaluation des élèves en investigation scientifique _____	75
20	Les critères de notation des portfolios liés à la planification _____	77
21	Les principes de la pratique de l'évaluation des élèves _____	94

**Figures**

1	L'évaluation des élèves liée à des objectifs formatifs _____	21
2	L'évaluation des élèves liée à des objectifs sommatifs _____	24
3	Les interactions entre les divers aspects de l'ensemble du programme _____	31



## Préface

En 1996, Georges Charpak, avec le soutien unanime de l'Académie des sciences, lance le programme *La main à la pâte*, en association avec le ministère de l'éducation nationale. Dix-sept années plus tard, le programme est bien vivant, il a développé la curiosité de millions d'enfants, inspiré nombre de professeurs des écoles, d'inspecteurs, de responsables du ministère, de parents d'élèves. En 2005, le *Science Education Program* du Réseau mondial des Académies des sciences (IAP) tout juste créé sous la vigoureuse impulsion de Jorge Allende, biochimiste et ancien Président de l'Académie des sciences du Chili, lance au sein des Académies et de leurs partenaires ministériels dans un grand nombre de pays, une réflexion qui promeut un enseignement fondé sur l'investigation (ESFI) à l'école et au collège. *La main à la pâte*, avec les Dix principes de sa pédagogie formulés dès 1998 et ses nombreuses coopérations internationales, s'est pleinement reconnue dans cet ESFI. La conférence internationale, organisée par l'IAP à Helsinki en 2012 et, à l'origine de ce livre, a rassemblé des participants de cinquante pays allant des plus riches aux plus pauvres, autour du thème de *L'évaluation des élèves dans un enseignement fondé sur l'investigation*. La conférence a unanimement conclu qu'au sein de l'ESFI, l'évaluation des élèves est un enjeu essentiel, tant pour ces derniers que pour la qualité recherchée de l'enseignement scientifique. Elle a souhaité qu'un ouvrage, approfondissant les échanges et les questions posées, soit publié pour guider et stimuler l'indispensable réflexion sur le sujet.

Parler d'une évaluation conduite dans le cadre de l'ESFI suppose de préciser ce que l'on entend par ESFI. Ce souci de précision est nécessaire, compte tenu des interprétations diverses et parfois erronées que l'on peut parfois rencontrer. Ce livre s'y attache et cet aspect ne peut qu'intéresser le lecteur français, puisqu'en France les programmes de l'école primaire ont en 2002 introduit la *démarche d'investigation*, démarche reprise ensuite pour le collège, le lycée et aujourd'hui les classes préparatoires aux grandes écoles.

Toute évaluation des élèves est complexe car elle suppose non seulement d'avoir clairement défini ce que l'on cherche à évaluer mais aussi de se donner les moyens d'en extraire des résultats probants. Ceci est d'autant plus difficile qu'un grand nombre de facteurs entrent en jeu, certains étant difficilement contrôlables. Par ailleurs, l'ESFI implique de s'intéresser non seulement aux connaissances mais aussi aux compétences scientifiques acquises par les élèves. Cela suppose de profondes modifications tant dans les évaluations elles-mêmes que dans les pratiques pédagogiques. En France, depuis la création en 2006 du Socle commun de connaissances et de compétences, et aujourd'hui de culture, comprenant l'ESFI dans ses objectifs, les enseignants cherchent comment mieux évaluer leurs élèves. Ce livre insiste fortement sur l'importance de l'évaluation formative, objectif repris, par exemple, dans les « *grilles de références pour l'évaluation et la validation des compétences* » de l'école primaire française.

Ce livre, incontestablement inspiré d'abord par les expériences, échecs et recherches au Royaume-Uni, donne également nombre d'exemples issus d'autres pays. Paraissant au moment même où se mettent en place en France les nouvelles *Écoles supérieures du professorat et de l'éducation*, il devrait aider enseignants, inspecteurs, formateurs et cadres du ministère, mais aussi étudiants et futurs professeurs, à faire évoluer les pratiques pédagogiques et les procédures d'évaluation. Le but ultime demeurant, bien entendu, que chaque enfant prenne conscience de ses progrès, de ses potentialités tout en conservant sa curiosité, son avidité de comprendre le monde qui l'entoure, en acceptant également d'écouter l'autre et de le respecter.



Parce que les cultures d'évaluation des élèves diffèrent aujourd'hui assez profondément de part et d'autre de la Manche, sans parler d'autres régions du monde, la traduction en français n'a pas toujours été aisée. Le vocabulaire propre à ce sujet dans le monde anglo-saxon, et singulièrement dans les publications pédagogiques, administratives ou de recherche, ne recouvre pas toujours le vocabulaire auquel sont habitués les professeurs et les cadres de notre éducation nationale. Remerciant ici notre traductrice Madame Zheng Ma pour la qualité de son travail, nous avons veillé à revoir ou expliciter certains termes quand leur correspondance en français faisait question. Nous espérons que le lecteur s'y retrouvera et qu'il explorera ainsi, avec l'ESFI en France et dans le monde, cette difficile mais si importante question de l'évaluation des élèves.

**Edith Saltiel & Pierre Léna**

Fondation de coopération scientifique *La main à la pâte*

Août 2013

---

# Introduction

Depuis le courant de la décennie 1990, nous constatons une expansion rapide d'intérêt en faveur d'un *enseignement des sciences fondé sur l'investigation* (ESFI). La promotion et le développement des pratiques et des matériels, en classe et au laboratoire, au sein de projets pilotes dans le monde encouragent les élèves à prendre des initiatives, à donner un sens aux événements et aux phénomènes du monde autour d'eux. L'adoption d'une pédagogie d'investigation dans l'enseignement démontre son potentiel qui permet aux élèves de développer la compréhension, les compétences, les attitudes et les intérêts, nécessaires à chacun pour vivre dans des sociétés de plus en plus dépendantes des applications des sciences. L'investigation conduit à la connaissance des objets ou des phénomènes à l'issue des études mais surtout, elle aide à construire des concepts généraux ayant un grand pouvoir explicatif et qui permettent la compréhension d'objets ou d'événements nouveaux. Elle engendre également des réflexions sur les processus de pensée et les stratégies d'apprentissage qui sont nécessaires pour continuer à apprendre tout au long de la vie. Cependant, la mise en œuvre de l'ESFI pose un certain nombre de problèmes. Le principal d'entre eux est l'évaluation de l'apprentissage des élèves car celle-ci a une forte influence sur le contenu et la manière d'enseigner.

## Le contexte du livre

La décision d'écrire ce livre fut stimulée par la conférence internationale qui eut lieu à Helsinki du 30 mai au 1<sup>er</sup> juin 2012. La conférence, intitulée *Le développement de la pédagogie d'investigation dans l'enseignement scientifique : nouvelles questions*, fut organisée conjointement par le Réseau global des Académies des sciences (IAP), ALLEA (All European Academies), l'Académie finlandaise des sciences et des lettres et le Centre de l'enseignement scientifique de Finlande (LUMA). Les questions traitées portaient sur deux thèmes liés à l'ESFI : le rôle de l'évaluation et le rapport avec l'industrie. Une grande partie de la conférence fut consacrée aux problèmes liés à l'évaluation, qui font l'objet de ce livre.

Le Programme de l'éducation à la science (*Science Education Program*), et plus précisément de l'enseignement des sciences, de l'IAP, sous la direction du fondateur du programme, le Professeur Jorge Allende (Chili), a tenu des conférences internationales importantes sur les divers aspects de l'ESFI en 2005, 2008 et 2010 en alternance avec des réunions de groupes de travail pour l'organisation et la communication des activités. Le projet a également induit et soutenu des activités à travers quatre réseaux régionaux académiques dans le monde. Après le départ du professeur Allende, le SEP a été dirigé par le professeur Pierre Léna (France) entre 2010 et 2013. Depuis mi-2013, le programme est dirigé par le professeur Dato Lee Yee Cheong (Malaisie).

Au fil du travail de l'IAP et du SEP, il apparaît que le rôle de l'évaluation des élèves est devenu de plus en plus important dans la compréhension et la mise en œuvre de l'ESFI. L'évaluation est un aspect essentiel d'une planification stratégique pour l'évolution de l'enseignement. Avec la diffusion des projets pilotes, des informations sur l'efficacité de l'ESFI sont réclamées pour justifier les ressources requises pour sa mise en œuvre. De bonnes mesures des résultats de l'ESFI sont nécessaires pour cela. Cependant, fournir des informations sur les résultats des élèves ne représente que l'un des rôles de l'évaluation ; le rôle qu'elle peut jouer en facilitant l'apprentissage et développant une compréhension plus profonde des objectifs de l'enseignement scientifique s'est vérifié largement ces dernières années dans le monde. Néanmoins, l'influence indéniable de ce qui est évalué sur le programme et sur la pédagogie montre que l'évaluation peut également restreindre les changements

**Encadré 1: les pays représentés à la conférence d'Helsinki**

Argentine  
 Australie  
 Autriche  
 Brésil  
 Cameroun  
 Canada  
 Chine  
 Colombie  
 Costa Rica  
 Danemark  
 Estonie  
 Finlande  
 France  
 Allemagne  
 Ghana  
 Haïti  
 Hongrie  
 Inde  
 Iran  
 Israël  
 Italie  
 Ile Maurice  
 Mexique  
 Kenya  
 Kosovo  
 Malaisie  
 Monténégro  
 Mozambique  
 Pays-Bas  
 Nouvelle-Zélande  
 Nigeria  
 Norvège  
 Pakistan  
 Sénégal  
 Serbie  
 Slovénie  
 Afrique du Sud  
 Soudan  
 Suède  
 Suisse  
 Tanzanie  
 Tunisie  
 Ouganda  
 Ukraine  
 Royaume-Uni  
 États-Unis  
 Venezuela  
 Viêt-Nam  
 Zambie  
 Zimbabwe

souhaités dans l'enseignement des sciences lorsque les méthodes et le contenu de l'évaluation ne reflètent pas les objectifs de l'ESFI.

Répondre aux questions qui entravent la mise en œuvre de l'ESFI devient également plus urgent, compte tenu de la valeur d'une pédagogie d'investigation pour le développement des compétences requises par le marché du travail actuel et futur. C'est la raison pour laquelle le conseil du programme SEP de l'IAP a décidé, lors de sa réunion en avril 2011 à Paris, que l'évaluation des élèves devait être un axe majeur de la conférence de 2012, d'où découla la décision de produire ce livre.

## Les points principaux dégagés lors de la Conférence

Les présentations lors de la conférence d'Helsinki, notamment les discussions entre les 93 participants venants de 50 pays (Voir encadré 1), ont indiqué des préoccupations majeures dans le domaine de l'évaluation des élèves, résumées en six points:

- La nécessité de clarifier les termes ;
- La compréhension des interactions entre l'évaluation, la pédagogie, les programmes scolaires et les politiques de l'enseignement ;
- Le rôle des tests, des tâches et des enseignants dans l'évaluation de l'ESFI ;
- L'adéquation de l'évaluation à des objectifs différents ;
- Le défi de transformer l'évaluation ;
- La nécessité de développer les recherches dans le domaine de l'évaluation.

### La nécessité de clarifier les termes

Si nous souhaitons faire des progrès dans l'élaboration des politiques et dans la pratique de l'évaluation de l'ESFI, il est très important qu'existe une compréhension commune des mots utilisés. Les discussions lors de la Conférence révélèrent une certaine confusion quant à la signification de termes clés tels que *evaluation*,<sup>1</sup> *tests*, différence entre *évaluation formative* et *évaluation sommative*.

En outre, la signification de l'ESFI continue parfois à être interprétée de manière plutôt restrictive, portant uniquement sur le développement des compétences. Cela est compréhensible étant donné que l'ESFI a été introduit dans

1 Une difficulté de traduction intervient ici, l'anglais disposant de deux termes, respectivement *assessment* and *evaluation*, tandis que le français ne dispose que du terme *évaluation*. Cf. *infra* (N.d.T.)

certains pays comme un antidote à un enseignement basé exclusivement sur des manuels scolaires. Cependant, une vision commune de la signification de l'ESFI dans sa mise en pratique et de ses différences avec les autres approches d'enseignement et d'apprentissage est non seulement nécessaire pour sa mise en œuvre, mais indispensable en tant que plateforme consensuelle pour l'élaboration des modalités d'évaluation. Les participants de la Conférence ont précisé que, en dehors de la signification des termes, il existe des différences culturelles qui sous-tendent le discours et l'évaluation de l'ESFI, conduisant à des visions systémiques différentes qui devraient faire l'objet d'approfondissement et de débat.

### **La compréhension des interactions entre l'évaluation, la pédagogie, les programmes scolaires et les politiques dans l'éducation**

Face à un consensus général selon lequel le contenu évalué influe sur la priorité accordée par les enseignants à différents objectifs d'apprentissage, il est donc essentiel que tous les objectifs importants soient inclus dans l'évaluation. Il est suggéré que ce système serait facilité si les programmes scolaires et l'évaluation étaient conçus ensemble par les mêmes organismes afin d'éviter des objectifs ambitieux convertis en une série d'épreuves déconnectées. Un autre point précisé par plusieurs participants affirme que l'évaluation doit rester cohérente avec la théorie d'apprentissage qui est celle de l'ESFI. La majorité pense que la connexion entre une évaluation réussie, la mise en œuvre de l'ESFI et l'élaboration des notions scientifiques clés doit être précisée. Il est nécessaire de décrire ce que l'on entend par «qualité» dans l'apprentissage des sciences et en quoi cela est différent à différents âges, au fur à mesure que les élèves avancent dans leur apprentissage de notions *de* science et portant *sur* la science. L'évaluation doit contribuer effectivement au développement de bons citoyens, offrir des connaissances et des compétences nécessaires pour aborder les grands problèmes mondiaux. Les politiques d'enseignement établies sur une ambition de performances de haut niveau dans les tests et les enquêtes internationales doivent être conciliables avec les préoccupations d'une éducation de haute qualité pour tous.

### **Le rôle des tests, des tâches proposées et des enseignants dans l'évaluation de l'ESFI**

Les participants à la conférence admettent que la nature de l'ESFI contient beaucoup de défis conceptuels, logistiques et techniques pour l'évaluation des élèves. Sur le plan conceptuel, les procédures de l'évaluation devraient permettre une meilleure compréhension de ce que signifie "être un bon élève en science". Ceci implique une clarification des objectifs de l'enseignement scientifique et de la façon dont l'ESFI y contribue. Sur le plan technique, le défi consiste à assurer la fiabilité sans compromettre la validité. L'objectif énoncé, consistant à développer des apprenants confiants, autonomes et collaboratifs, est difficile à évaluer directement et nous devons rechercher des mesures de substitution reflétant ces qualités. Nous constatons qu'il est quasiment impossible de proposer des tests d'une durée raisonnable, accessibles à tous les élèves, qui puissent fournir la richesse d'information nécessaire pour évaluer les objectifs de l'ESFI. De plus, la production et l'administration des tests impliquent un coût élevé en temps ainsi qu'en autres ressources. Les participants affirment que les enseignants doivent être impliqués dans la gestion de l'évaluation de certaines ou de toutes les étapes de la scolarité. Il est suggéré que des commentaires sur les résultats de l'évaluation et l'appréciation des enseignants puissent avoir un rôle dans le renforcement de leurs compétences d'évaluation. L'impact de la culture locale de l'école doit également être considéré. Il se peut que les enseignants soient contraints d'adopter de nouvelles approches de l'enseignement et de l'évaluation par des mesures de type purement comptable, ayant une vision étroite de l'enseignement et de l'apprentissage des sciences.

## **L'adéquation de l'évaluation à des objectifs différents**

Afin que l'évaluation soit utilisée pour faciliter l'apprentissage, les enseignants doivent intégrer des stratégies d'évaluation formative dans le cadre de leur pédagogie au lieu de cumuler une série de mini-évaluations sommatives. Pour l'évaluation sommative, le test est la méthode couramment utilisée, afin de contrôler les performances à la fin d'un thème ou d'une leçon, et pour faire à intervalles réguliers un rapport sur les progrès. De l'avis général, la plupart des tests utilisés par les enseignants ne reflètent pas les objectifs de l'ESFI. Puisque les tests ont un impact important sur ce qui est enseigné, il est important de considérer des alternatives à la plupart des formes actuelles de tests afin d'obtenir des informations plus fiables sur l'apprentissage produit par l'ESFI. Certains participants ont affirmé que l'utilisation des résultats des tests dans une logique strictement comptable inhibe l'apprentissage. Nous avons noté que les programmes de tests à grande échelle, tels TIMSS et PISA, utilisent des méthodes qui n'aboutissent pas à des conclusions fiables sur la performance de l'ESFI au niveau d'un système national. Toutefois, bien que nous ayons reconnu que ces résultats d'enquêtes jouent un rôle, ils doivent être considérés comme un des indicateurs de performance du système et doivent être analysés en tenant compte d'informations obtenues sur d'autres aspects.

## **Le défi de changements dans les politiques et la pratique de l'évaluation**

Les changements les plus importants et les plus grands défis sont perçus par les participants à la Conférence comme étant en étroite relation avec la participation des enseignants dans l'évaluation. Ces derniers ont besoin d'aide pour élaborer leurs «compétences d'évaluation». Certains objectifs de l'ESFI sont mieux évalués par les enseignants. Les possibilités fournies aux élèves pour apprendre en classe avec l'ESFI offrent toutes aussi l'occasion pour les enseignants d'évaluer et de noter les progrès portant sur la compréhension et les compétences de leurs élèves. Les enseignants ont également besoin d'outils pour les aider à leurs évaluations formative ou sommative. Dans de nombreux pays, la participation des filles à l'étude des sciences est un défi, mais il n'y a aucune preuve de différences entre les sexes en matière de compétence dans l'ESFI. Il est important de peser sur la formation initiale des enseignants, afin qu'ils entrent dans la profession en connaissant l'ESFI tout en sachant comment l'incorporer dans leur pratique de l'évaluation. Les participants à la conférence suggèrent également la nécessité d'éduquer les responsables politiques, les administrateurs, les parents et le public en général sur les significations, les buts, les points forts et les limites de l'évaluation.

## **La nécessité de recherches approfondies portant sur l'évaluation**

Les travaux de recherche présentés lors de la conférence ont fourni des preuves frappantes de la nécessité de repenser certains aspects de la manière dont les élèves sont évalués, afin de leur donner des occasions de montrer ce qu'ils savent et peuvent faire. La confiance dans l'exactitude de certains résultats de l'évaluation est souvent ébranlée par des recherches montrant que les méthodes avec lesquelles les élèves abordent une question ou une épreuve, ainsi que leur processus de pensée, ne correspondent pas aux effets attendus au moment de l'interprétation des résultats. Plus d'études sont nécessaires afin d'améliorer la validité des outils et des procédures d'évaluation et d'écarter des espoirs irréalistes de précision. Nous savons que même des changements mineurs dans la formulation et le fonctionnement de questions écrites peuvent influencer les performances des élèves, mais nous en savons moins sur leur impact différentiel selon le sexe, l'expérience ou le contexte. Mentionnons plusieurs autres domaines où des recherches sont nécessaires : l'impact des méthodes offrant une alternative aux tests ; la combinaison de différents types de données qui peut fournir une meilleure représentation de certains types de performances des élèves ; la façon d'augmenter la confiance dans une évaluation faite par les enseignants ; la formation des enseignants à utiliser l'évaluation formative ou sommative ; la combinaison des évaluations formative et sommative au niveau des systèmes d'évaluation nationaux et régionaux.

## L'objet du livre

Dans ce livre, comme lors de la conférence, nous distinguons deux types de questions à traiter:

- les processus et les modalités de *l'évaluation de l'apprentissage des élèves*
- la valeur d'une pédagogie d'investigation dans l'enseignement scientifique et *l'évaluation de son efficacité*.

Le premier sujet concerne les différentes questions, résumées ci-dessus, qui découlent de la reconnaissance du rôle de l'évaluation des élèves en tant que partie intégrante de leur expérience d'apprentissage. L'évaluation était autrefois considérée comme une étape postérieure à l'apprentissage et tout à fait distincte du processus d'apprentissage. Ce point de vue n'est plus tenable ; l'évaluation des élèves est maintenant reconnue comme un élément central de l'enseignement, jouant un rôle certain dans sa contribution à l'apprentissage autant que dans l'appréciation de sa qualité. Nous savons que la manière dont les résultats de l'évaluation des élèves sont utilisés a une forte influence, positive ou négative, sur le contenu et les méthodes d'enseignement. Ainsi, la nature de cette évaluation, et en particulier la mesure selon laquelle elle permet aux élèves de montrer ce qu'ils savent et peuvent faire par rapport aux objectifs d'apprentissage visés, constituent des facteurs clés dans l'éducation des élèves. Dans le contexte de l'ESFI, nous constatons avec préoccupation que les outils et les procédures d'évaluation des élèves les plus courants sont insuffisants face à ce qui est nécessaire pour assurer une bonne appréciation de la réussite des objectifs de l'ESFI.

En ce qui concerne le second point, la valeur de l'ESFI ne peut pas être déterminée uniquement par des preuves empiriques ; il s'agit d'un jugement de valeur qui veut démontrer que les compétences, la compréhension, les intérêts et les attitudes qui constituent les objectifs de l'ESFI sont tout à fait utiles et nécessaires dans une éducation moderne. L'évaluation d'un programme d'enseignement *peut* montrer dans quelle mesure les élèves atteignent ces objectifs grâce à une pédagogie fondée sur l'ESFI. L'évaluation des élèves a un rôle dans l'évaluation d'un tel programme mais beaucoup d'autres facteurs interviennent. En particulier, les résultats des élèves sont uniquement informatifs par rapport aux résultats de l'ESFI s'il existe par ailleurs des résultats probants qui montrent que les étudiants apprennent réellement avec une pédagogie d'investigation et que les données fournies par l'évaluation permettent de tirer des conclusions sur leur compréhension de la science et sur leurs compétences d'investigation scientifiques qui sont les objectifs-mêmes de l'ESFI. La nature actuelle de l'évaluation des élèves mise en œuvre dans l'apprentissage en sciences est l'un des facteurs principaux qui freine l'implantation d'un programme d'ESFI.

Il est sans doute nécessaire de répondre aux questions posées concernant l'impact sur la réussite des élèves d'un enseignement fondé sur l'investigation, mais cela ne peut se faire sans outils adéquats. Ainsi, en examinant le premier point ci-dessus, portant sur l'évaluation de l'apprentissage des élèves, nous offrons non seulement la perspective de fournir certains outils nécessaires à l'évaluation du programme d'enseignement, mais surtout celle de considérer une évaluation adéquate des élèves comme faisant intégralement partie d'un « apprentissage fondé sur l'investigation »



# Chapitre 1

## Clarification des termes

La plupart des termes employés au sujet de l'évaluation ont une utilisation à la fois technique et courante, comme dans le cas de nombreux mots utilisés en science. Le manque de clarté dans la signification et la cohérence entrave la communication et la compréhension des distinctions et des connexions entre les différents concepts. [Pour le lecteur francophone, une double clarification est apparue nécessaire, car la traduction du texte original en anglais peut ne pas couler de soi. Les termes anglais *evaluation, assessment, test, standards, criteria, validity, reliability*, largement utilisés par les publications dans cette langue, n'ont pas nécessairement un équivalent strict en français, et peuvent même avoir de faux amis. Ce chapitre permettra de clarifier le vocabulaire employé en français, en faisant au mieux le lien avec le sens recherché dans le texte original en anglais (N.d.T.).]

Le terme *inquiry*, traduit par *investigation*, ne pose pas de problème similaire. Il est discuté au chapitre 2 et nous verrons au chapitre 3 d'autres mots, qui sont en rapport avec les objectifs, les fonctions et les utilisations de l'évaluation.

### Évaluation (des élèves), évaluation (systémique) et appréciation (des éducateurs)

L'OCDE, dans ses rapports destinés à ses pays membres, distingue deux formes d'évaluations (*assessment* et *evaluation*). Il établit une distinction claire et nécessaire entre l'évaluation des élèves (*assessment*), l'évaluation des écoles, du système éducatif ou d'une politique éducative (*evaluation*), l'appréciation (*appraisal*) des enseignants tout en affirmant que ces éléments doivent travailler de concert dans les politiques qui visent à améliorer les résultats d'apprentissage :

Le terme *évaluation des élèves* est utilisé pour désigner les jugements portés sur la performance et la réussite de chaque élève individuel par rapport aux objectifs d'apprentissage. Il comprend les *contrôles* faits en classe ainsi que les *tests* et les *examens* externes faits à grande échelle. Le terme *appréciation des éducateurs* est utilisé pour désigner les jugements portés sur la performance des professionnels de l'école, par exemple, les enseignants et les directeurs d'école. Enfin, le terme *évaluation systémique* est utilisé pour désigner les jugements portés sur l'efficacité des écoles, des systèmes éducatifs et des politiques d'enseignement.<sup>2,3</sup>

L'évaluation **des élèves** et l'évaluation **systémique** décrivent toutes les deux un processus de création et d'interprétation de preuves<sup>4</sup> en vue d'un certain objectif. Ces deux formes d'évaluation impliquent l'une et l'autre des décisions sur le choix des preuves à utiliser, la production et la collecte des preuves

2 Nusche, D. et al (2012) *OECD Reviews of Evaluation and Assessment in Education: New Zealand 2011*. OECD Publishing, p24.

3 Dans cette traduction en français, nous avons conservé les définitions et leur objet précis, puis choisi le terme français qui nous paraît le mieux leur correspondre. Ces termes seront désormais ceux que nous utiliserons dans la traduction de la suite de cet ouvrage (N.d.T.).

4 La traduction de l'anglais *evidence* fait surgir une autre difficulté. Le français *évidence* n'ayant pas tout à fait le même sens. Nous traduisons donc par le terme *preuve*, bien que celui-ci ait un caractère très contraignant dans son emploi le plus courant en français. Il ne faut donc pas lui donner dans cette traduction le sens d'une rigoureuse démonstration mathématique ou logique, mais plutôt un sens d'élément de preuve ou d'élément convaincant...



d'une manière systématique et planifiée, l'interprétation des preuves en vue d'en déduire un jugement, ainsi que la communication et l'utilisation de ce jugement. Il est à noter que les preuves, quelle qu'elles soient, ne sont jamais qu'une indication ou qu'un échantillon d'un éventail plus large qui pourrait être utilisé.

Dans ce livre, d'après la convention de l'OCDE, l'expression **évaluation des élèves** (*assessment*) est partout utilisée pour désigner le processus de collecte et d'utilisation de preuves sur les résultats de l'apprentissage, le plus souvent celui des élèves, mais également celui des autres acteurs tels que les enseignants.

L'expression **évaluation systémique** (*evaluation*) est utilisée en référence à la création et à l'utilisation de preuves sur les systèmes, les matériels, les procédures et les processus. L'évaluation systémique des écoles, des systèmes ou des méthodes d'enseignement peut faire usage de preuves obtenues sur l'apprentissage des élèves, mais le jugement porte sur la valeur ou le succès d'autres éléments tels que les politiques et le programme scolaire plutôt que sur l'apprentissage des élèves, bien que ce dernier puisse faire partie des preuves utilisées dans l'évaluation systémique.

## Les tests et autres méthodes d'évaluation

Bien que les termes **évaluation des élèves** d'une part, **contrôles** ou **tests** d'autre part, soient parfois utilisés de façon interchangeable, ils méritent d'être soigneusement distingués. Les *tests* peuvent être considérés comme une méthode particulière pour collecter des données en vue de l'évaluation des élèves et par conséquent, cette dernière recouvre plus que les tests et peut comprendre d'autres méthodes de collecte et d'interprétation des données.

Un regard plus attentif sur l'évaluation des élèves peut aider à clarifier cette relation et à identifier d'autres aspects de cette évaluation, impliquant des termes tels que *norme* (*standard*) et *critère* (*criteria*).

Toute évaluation de niveau des élèves implique la production, l'interprétation, la communication et l'utilisation de données en vue d'un certain objectif. Dans cette perspective, il est possible d'avoir une vaste gamme de types d'activités différentes sous certaines conditions : *a)* les élèves sont engagés dans une activité ; *b)* la collecte de données portant sur cette activité est effectuée par un agent ; *c)* le jugement porté sur les données est fait par leur comparaison à certaines normes ; *d)* il faut certains moyens pour décrire et communiquer le jugement. Chacune des composantes de l'évaluation des élèves peut ainsi prendre des formes différentes :

- a) Les activités dans lesquelles les élèves sont engagés peuvent être, par exemple :
  - leur travail habituel ;
  - certaines tâches écrites ou pratiques créées par l'enseignant dans le but de l'évaluation ;
  - des tâches écrites ou pratiques créées à l'extérieur.
- b) Les données peuvent être collectées par :
  - L'enseignant ;
  - Les élèves ;
  - L'enseignant avec les élèves ;
  - Un agent extérieur (jury d'examen, les autorités de qualification, concepteur des tests).
- c) Les données peuvent être jugées par rapport à :
  - des normes. Ici la règle de comparaison se réfère à la performance des autres élèves jugée d'après les normes ;
  - des critères. Ici la règle de comparaison est une description des divers aspects de la performance de l'élève, jugée d'après les critères.
  - la performance passée de l'élève. Ici la performance actuelle d'un individu est jugée par rapport à ses performances antérieures ou dans un champ différent.

- d) Les jugements peuvent être communiqués sous forme de :
- un commentaire écrit ou oral par l'enseignant ;
  - une note, un score ou un pourcentage ;
  - un profil de réussite ;
  - un niveau ;
  - un classement.

Des outils et des procédures différents d'évaluation des élèves sont créés par des combinaisons différentes de ces diverses manières de collecter, de juger et de communiquer des données. Par exemple, un test standardisé est constitué de tâches créées par une agence externe qui a soumis le test pendant son élaboration à un grand échantillon représentatif de la population concernée. Par conséquent, le score d'un individu peut être exprimé pour se comparer avec une *norme* définie sur cette population particulière. Le résultat indiquera si la performance de l'élève est au-dessus ou en dessous de la moyenne, mais pas ce qu'il/elle est capable de faire.

Un *test fondé sur les critères* diffère d'un *test fondé sur les normes*, car il est conçu pour donner des informations sur ce que l'élève peut faire par rapport aux résultats visés. Les exercices sont choisis en fonction de leur pertinence par rapport au programme afin que les résultats puissent être utilisés, non pas pour comparer un élève aux autres, mais pour montrer comment sa performance se compare à la performance attendue. En même temps, le niveau cible de performance est fixé par référence à ce qui peut être attendu de la population considérée. Il y a donc un élément normatif dans la détermination des critères en fonction desquels la performance est jugée.<sup>5</sup> Lorsque des tests externes sont appliqués, les données sont limitées aux résultats des tests ; alors que si l'évaluation des élèves est directement effectuée par les enseignants, la possibilité existe d'utiliser la gamme complète des activités d'apprentissage pour porter un jugement en fonction des critères de la performance attendue par rapport aux objectifs des leçons enseignées.

## La validité, la fiabilité, les ressources et la capacité de gestion

Pour décider quelle est la meilleure façon de mener l'évaluation des élèves dans un cas spécifique, il est nécessaire de réfléchir aux propriétés des outils possibles par rapport aux objectifs et aux utilisations des résultats de cette évaluation. Il est évident que la première propriété souhaitable est que toute évaluation doit être adaptée à son objectif et qu'elle doit bien évaluer ce que l'on veut évaluer. Une autre propriété est qu'elle doit fournir des données fiables. Mais il y a aussi d'autres questions à prendre en compte, en particulier compte-tenu de l'interdépendance des différents composants du système : l'impact sur les autres pratiques d'évaluation, le programme et la pédagogie, ainsi que l'utilisation des ressources. L'évaluation peut être coûteuse en ressources financières, en temps des élèves et des enseignants. La capacité de gestion de cette évaluation doit donc aussi être prise en considération.

### Validité

Il est commun de définir la *validité* d'une évaluation des élèves en appréciant la correspondance plus ou moins bonne entre ce qui est évalué et le comportement ou les résultats attendus de l'apprentissage. Différents types de validité ont été proposés en fonction des informations utilisées dans le jugement de cette dernière. Par exemple, *la validité de contenu* se réfère à la manière dont l'évaluation couvre adéquatement la discipline enseignée. Cette validité est souvent fondée sur le jugement des experts en la matière. Cependant la couverture du contenu n'est pas suffisante pour faire une distinction entre un test classique et une autre forme d'évaluation, destinée à un enseignement scientifique fondé sur l'investigation. *La validité*

5 Black, P (1998) *Testing: Friend or Foe?* London: Falmer Press.

*conceptuelle* est un concept plus large qui reflète la gamme complète des résultats de l'apprentissage dans un domaine particulier. Sa principale exigence veut que l'évaluation touche à tous les aspects, mais surtout aux aspects pertinents pour la réussite des élèves par rapport à l'objectif spécifique de l'évaluation. La présence d'aspects inutiles est autant une menace pour la validité que l'omission des aspects pertinents.

Cependant, cette vision d'une validité qui appartiendrait en propre à une méthode d'évaluation ou à un instrument, indépendamment des circonstances dans lesquelles cette évaluation est utilisée ou de l'usage fait de ses résultats, a été largement contestée. Newton<sup>6</sup> a fait remarquer l'erreur propre à ces deux hypothèses. Dans le cas des conditions d'utilisation de l'évaluation, la précision des résultats en tant que mesure de validité dépend de la façon dont l'évaluation est administrée ainsi que de son contenu. Dans le cas de l'utilisation des résultats, des assertions peuvent être posées sur ces résultats, alors qu'en fait, d'autres facteurs ignorés influent davantage sur ces résultats (par exemple lorsqu'un test de mathématiques requiert une grande capacité de lecture, cela crée une situation confuse car nous ne pouvons décider si c'est la lecture ou la compétence mathématique qui a le plus d'influence sur les résultats).

La notion de *validité* prend en compte non seulement la façon dont l'évaluation teste le contenu, mais aussi les conclusions tirées à partir des résultats. Cela fut formellement exprimé dans une définition de validité donnée par Messick et largement citée:

La *validité* est un jugement évaluatif et intégré du degré avec lequel les preuves empiriques et les raisons théoriques plaident en faveur de l'adéquation et de la pertinence des conclusions tirées et des actions engagées lorsqu'elles se fondent sur les résultats des tests mis en oeuvre.<sup>7</sup>

Adopter cette définition de la *validité* revient à admettre que cette validité peut varier en fonction de facteurs influents sur la performance comme les conditions concrètes des tests, généralement décrites en termes de *fiabilité* de l'évaluation des élèves. Nous reviendrons sur ce point après avoir examiné le sens de ce terme *fiabilité*.

## Fiabilité

La *fiabilité de l'évaluation des élèves* désigne dans quelle mesure les résultats peuvent être considérés comme cohérents et exacts en vue d'un usage spécifique. Cela peut ne pas être le cas si, par exemple, les résultats sont influencés par celui qui donne l'évaluation ou s'ils dépendent de circonstances particulières à un moment donné. Ainsi, la fiabilité est souvent définie et mesurée par l'obtention du même résultat lorsque l'évaluation est répétée. La fiabilité est surtout significative dans l'*évaluation sommative* et en particulier lorsqu'il s'agit de tests. L'utilisation d'une *évaluation formative* (voir chapitre 3) ne concerne que l'élève et son professeur, si bien qu'ici l'idée de porter un jugement qui soit reproductible en traitant tous les élèves de la même manière n'est pas pertinente. Il ne s'agit alors pas de juger sous forme de notes ou de niveaux de classe, mais d'aider l'élève à progresser vers les étapes suivantes de son apprentissage. Par conséquent, la fiabilité au sens strict n'intervient pas ici. Ce qui est important pour l'évaluation formative, c'est « la qualité de l'information recueillie et utilisée en rétroaction sur l'enseignement ».<sup>8</sup>

6 Newton, P. Validity, purpose and the recycling of results. In (ed) J. Gardner *Assessment and Learning*. 2nd edn. London: Sage

7 Messick, S.(1989) Validity, in (ed) R. Linn *Educational Measurement* (3rd edn) American Council on Education , Washington: Macmillan, pp 13-103, p13.

8 Stobart, G. (2012) Validity in formative assessment In (ed) J. Gardner *Assessment and Learning*. 2nd edn. London: Sage p234.

Toutefois, une grande fiabilité *est* nécessaire lorsque les résultats sont utilisés par d'autres et lorsque les élèves sont comparés ou sélectionnés. Il est alors nécessaire que tous les élèves possèdent les mêmes conditions et les mêmes chances, donc que la fiabilité soit élevée. Bien sûr, il n'arrive jamais que les mêmes conditions donnent les mêmes chances à des élèves différents, ou qu'un élève réagisse de la même manière à un test en différentes occasions. Les réponses de chaque élève sous les mêmes conditions varient d'un jour à l'autre et, surtout, le choix des questions posées ; un test sera considéré plus difficile par certains élèves que d'autres de compétence égale. Tous les exercices présentent des questions dans un certain contexte et la recherche montre que les élèves ayant de bons résultats dans un exercice ne feront pas forcément aussi bien dans un autre, évaluant les mêmes concepts ou compétences dans un contexte différent. Il existe un grand nombre d'exercices possibles mais seulement un petit échantillon d'entre eux peut être inclus dans un test de longueur raisonnable. Un choix différent entrainera un résultat différent, donnant naissance à ce qui est décrit comme une *erreur d'échantillonnage*.

**Encadré 2: Le compromis entre la fiabilité et la validité**

Pour toute évaluation d'élèves, il existe une limite à l'optimisation simultanée de la fiabilité et de la validité. Cette difficulté s'applique à toutes les formes d'évaluation, mais elle est plus marquée dans l'utilisation de tests. Lors de la sélection des questions d'un test, afin d'augmenter la fiabilité, il est inévitable que la sélection privilégie les questions qui peuvent être notées de façon systématique ou par l'ordinateur. Cela favorise d'une part les questions qui permettent d'évaluer les connaissances factuelles, d'autre part l'utilisation de questions fermées, contrairement aux questions ouvertes nécessitant la mise en œuvre des connaissances. La limitation du contenu d'un test affecte sa validité. Les tentatives d'augmenter la validité par un élargissement du choix des questions, par exemple en incluant davantage de questions ouvertes qui nécessitent lors de la notation un jugement plus élaboré, conduisent à une réduction de la fiabilité. Il y a donc un compromis à trouver entre la fiabilité et la validité, l'augmentation de la fiabilité diminuant la validité et vice-versa.

L'erreur d'échantillonnage peut être beaucoup plus importante qu'on ne le croit. Par exemple, Wiliam<sup>9</sup> a estimé que, dans le cas des tests nationaux en Angleterre, environ 40% des élèves seraient affectés par ces tests à un niveau de classe incorrect (même si ces niveaux couvrent deux années scolaires). Une façon de réduire cette source d'erreur serait d'augmenter le nombre de contextes pour chaque compétence évaluée et donc le nombre de questions utilisées. Mais la longueur d'un test ne peut pas être considérablement augmenté sans risquer d'autres formes d'erreurs (fatigue des élèves, par exemple), donc plus de questions par compétence ou par concept signifierait moins de compétences et de concepts traités, réduisant ainsi le contenu évalué et la validité du test. Un autre exemple de l'interaction entre la fiabilité et la validité est identifié dans l'encadré 2. L'impact est important concernant les tests individuels des élèves qui comportent les mêmes questions. L'effet est moindre lorsqu'il s'agit d'enquêtes de population où l'erreur d'échantillonnage peut être réduite en utilisant un grand nombre de questions réparties dans des échantillons équivalents d'élèves qui choisissent des groupes de questions différents (voir chapitre 6).

9 Wiliam, D. (2001) Reliability, validity and all that jazz, *Education* 3-13, 29 (3): 17-21.

## Les ressources et la capacité de gestion

Les ressources requises pour mettre en œuvre une évaluation doivent être proportionnelles à la valeur qu'ont, pour les utilisateurs des données, les informations recueillies. Les ressources peuvent être sous forme du temps passé par les enseignants, de l'expertise et des dépenses des écoles, et/ou des organismes externes impliqués dans l'évaluation. En général, un compromis est nécessaire, surtout lorsqu'une grande précision est demandée. La conception et l'exploitation de l'évaluation rencontre des limites de durée ou d'expertise, s'agissant par exemple d'un test externe voulu comme hautement fiable ou d'un examen. Une triple correction de tous les tests apporterait clairement une plus grande confiance dans les résultats ; l'observation de tous les candidats par des spécialistes augmenterait l'éventail des résultats qui peuvent être évalués en externe; une formation de tous les enseignants pour devenir des évaluateurs experts aurait de grands avantages : mais tout cela n'est pas réaliste dans la pratique. L'équilibre entre les coûts et les avantages soulève des questions de valeurs et de possibilités techniques.

Le coût d'une *évaluation formative* est négligeable, une fois qu'elle est incorporée dans la pratique. Le processus de son introduction pourrait néanmoins consommer beaucoup de temps en développement professionnel des enseignants. Une bonne évaluation formative, comme nous le verrons au chapitre 5, nécessite non seulement la maîtrise de certaines stratégies pédagogiques en classe, mais aussi la connaissance des différentes étapes qui permettent à l'apprentissage de se dérouler. Elle requiert aussi des exemples de preuves, tant à la disposition des élèves qu'à celle des professeurs, aidant à identifier les étapes à venir de l'apprentissage. Ces coûts font toutefois partie intégrante des efforts visant à améliorer l'apprentissage.

L'*évaluation sommative* nécessite des ressources en temps des enseignants comme des élèves. Le coût devient considérable lorsque l'on utilise les tests développés par des organismes externes ou des éditeurs commerciaux. Même si les tests et les examens nationaux sont fournis gratuitement aux écoles, le coût doit être pris en charge par le système et peut être étonnamment élevé. Si les coûts directs de production, de distribution, de notation, des tests, etc. sont ajoutés au temps passé pour la préparation, les tests et les examens externes, le total peut s'élever à une proportion importante du budget de l'éducation.<sup>10</sup> Ce point nous conduit certainement à réfléchir sur l'équilibre entre les coûts et les avantages des diverses méthodes à utiliser pour une évaluation sommative.

---

10 Harlen, W. (2007) *Assessment of Learning*. London: Sage p 61/2

## Chapitre 2

# La pédagogie d'investigation dans l'enseignement scientifique : motif et objectifs

Le choix des moyens les plus efficaces pour s'assurer que l'évaluation des élèves soutient la pédagogie d'investigation dans l'enseignement des sciences nécessite une identification claire des résultats d'apprentissage attendus de l'enseignement des sciences fondé sur l'investigation (ESFI). Des modalités d'évaluation fournissant les données les plus fiables sur ces résultats peuvent alors être sélectionnées ou élaborées. Dans ce chapitre, nous examinons de plus près les objectifs de l'ESFI et les raisons de leur importance. Avant de se lancer sur ce sujet, il est important de noter que l'investigation n'est pas du tout la seule approche à utiliser dans l'enseignement des sciences. Certains aspects de l'apprentissage, comme la connaissance des vocabulaires scientifiques, les conventions et l'utilisation des équipements, sont mieux acquis par une instruction directe. Par conséquent, ni l'enseignement des sciences ni l'évaluation des élèves ne se réduisent au contexte spécifique de la pédagogie d'investigation. Cependant, la connaissance des faits et des procédures est un aspect central de l'investigation, afin que les élèves comprennent ce qu'ils apprennent. L'évaluation des élèves doit donc apprécier leur compréhension, leurs aptitudes et leurs compétences, dont l'ensemble constitue l'objectif de l'ESFI.

### L'investigation dans l'enseignement des sciences

L'investigation est un terme utilisé à la fois dans l'enseignement et dans la vie quotidienne pour désigner la recherche d'explications ou d'informations à travers des questions. Elle est parfois assimilée à la recherche, à l'enquête, ou à « la quête de la vérité ». Dans l'enseignement, l'investigation peut intervenir dans des domaines comme l'histoire, la géographie, les arts, mais aussi la science, les mathématiques, la technologie et l'ingénierie. Lorsque des questions sont soulevées, les preuves sont recueillies et les explications possibles sont envisagées. Dans chaque domaine, elle peut susciter l'émergence de différents types de connaissances et de compréhensions. Ce qui distingue *l'investigation scientifique* des autres, c'est qu'elle conduit à la connaissance et à la compréhension du monde naturel et artificiel, par une interaction directe avec ce monde, par la production et la collecte de données, utilisées ensuite comme des preuves venant à l'appui d'une explication des phénomènes ou des événements.

L'investigation est loin d'être un nouveau concept dans l'éducation. Elle est fondée sur la reconnaissance du rôle actif des enfants dans le développement de leurs idées et leur compréhension. Les études de Piaget<sup>11</sup> et les arguments de Dewey,<sup>12</sup> parmi d'autres dans la première moitié du 20<sup>e</sup> siècle, ont attiré l'attention sur le rôle important de la curiosité, de l'imagination, du désir d'interaction et de questionnement dans l'apprentissage des enfants. Plus récemment, le *National Research Council* aux Etats-Unis, l'Académie des sciences en France avec

11 Piaget, J (1929) *The Child's Conception of the World*. New York: Harcourt Brace.

12 Dewey, J. (1933) *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston, MA: D.C. Heath

*La main à la pâte*<sup>13, 14</sup> ont mentionné l'importance de faire participer les élèves, en faisant des observations et posant des questions, recueillant, analysant et interprétant les données, en communiquant les résultats.<sup>15</sup>

De même, la National Science Foundation a défini la pédagogie d'investigation comme conduisant « les élèves à construire leur compréhension des concepts scientifiques fondamentaux grâce à une expérience directe impliquant des matériaux, des documents, des experts, une argumentation et un débat entre élèves ». <sup>16</sup>

Au cours des divers projets pilotes de la dernière décennie, le Programme de l'enseignement des sciences de l'IAP (Réseau mondial des Académies des sciences) a élaboré la définition ci-dessous de la pédagogie d'investigation de l'enseignement des sciences :

L'ESFI recherche un développement progressif d'idées scientifiques fondamentales chez les élèves, grâce à l'apprentissage de l'investigation et de la construction de connaissances, afin de comprendre le monde qui les entoure. Les élèves appliquent les compétences utilisées par les scientifiques telles que le questionnement, la collecte de données, le raisonnement et l'examen des preuves à la lumière des connaissances disponibles, l'établissement de conclusions et la discussion des résultats. Ce processus d'apprentissage est soutenu par une pédagogie d'investigation qui comprend non seulement l'acte d'enseigner, mais aussi ses justifications sous-jacentes.<sup>17</sup>

Certains mots de cette définition méritent d'être soulignés et commentés.

*Le développement progressif d'idées scientifiques fondamentales* souligne l'importance d'identifier quelques idées essentielles qui aident à donner sens aux phénomènes du monde, puis d'assurer que, grâce aux activités d'apprentissage des sciences, les élèves progresseront dans l'élaboration de ces idées.

*L'apprentissage de l'investigation et de la construction de connaissances, afin de comprendre le monde qui les entoure* implique la participation active des élèves dans leur apprentissage, ce qui fait partie de l'évaluation formative (mentionné au chapitre 3) ; cela implique aussi une vision de l'apprentissage comme étant construit par les apprenants (expliqué au chapitre 4).

*L'application des compétences utilisées par les scientifiques* signifie qu'outre les compétences mentionnées, des qualités de rigueur et d'honnêteté sont requises dans la collecte et l'utilisation de données adéquates et pertinentes, afin de tester des hypothèses ou répondre aux questions posées. Les scientifiques vérifient, et répètent la collecte de données si possible, ils interprètent et essaient d'expliquer leurs résultats. Tout au long de leurs investigations, ils tiennent un registre précis et, pour tirer des conclusions, ils consultent les travaux existants et présentent leur travail aux autres par écrit ou dans des conférences afin de partager leurs idées. Cela paraît évident pour les scientifiques, mais en pratiquant l'investigation dans l'enseignement scientifique, il faut souligner que celui qui s'engage dans l'investigation ne connaît pas la réponse à la question ou au problème étudié. Il estime qu'il est important de chercher, est heureux d'essayer et ravi de trouver une réponse ou une solution.

13 Charpak G dir. (1996) *La main à la pâte. Les sciences à l'école primaire*, O Jacob

14 Charpak G ; Léna P., Quéré Y. (2005) *L'enfant et la science, l'aventure de la main à la pâte* O Jacob

15 National Research Council (NRC) (1996) *National Science Education Standards*. Washington DC: National Academy Press

16 National Science Foundation (NSF) (1997) *The Challenge and Promise of K-8 Science Education Reform*. Foundations, 1. Arlington, VA: NSF p7

17 IAP (2012) *Taking Inquiry-Based Science Education into Secondary Education*. Report of a global conference. <http://www.sazu.si/files/file-147.pdf>

*Le questionnement* souligne le fait que les élèves sont engagés à répondre aux questions qui présentent pour eux un réel intérêt et stimulent leur curiosité. Souvent, ces questions sont posées soit par l'enseignant, soit par les autres élèves, soit émergent d'une lecture. Mais quelle que soit l'origine des questions, les élèves les prennent à leur compte dans le cadre de l'investigation avec curiosité et désir de comprendre. S'interroger et répondre aux questions est parfois synonyme de résolution de problèmes, car l'attention se porte toujours sur la recherche d'une solution qui « fonctionne ». Toutefois, trouver la solution unique n'est pas suffisant en science. Le développement des théories et des modèles pour expliquer les phénomènes exige que les idées soient « évaluées face aux explications alternatives et comparées avec des preuves .... Ainsi, savoir pourquoi la mauvaise réponse est erronée peut aider à garantir une compréhension plus profonde et plus forte des raisons pour lesquelles la bonne réponse est juste. »<sup>18</sup>

La discussion des définitions montre clairement que l'apprentissage des sciences par l'investigation est un processus complexe dans lequel les connaissances, la compréhension, les compétences de collecte et d'utilisation des preuves sont reliées entre elles de manière interactive. Les compétences essentielles pour la construction de la compréhension sont à la fois des aptitudes physiques et mentales : la production et l'utilisation des preuves pour vérifier des idées qui peuvent aider à expliquer un événement ou un phénomène étudié.

En même temps, l'utilisation des compétences implique la connaissance et la compréhension : non pas seulement de savoir comment produire, recueillir et interpréter des données, mais aussi de comprendre pourquoi il est important de travailler de manière scientifique. En outre, il existe un aspect affectif, voire émotionnel, lorsqu'on accepte de s'engager dans les différentes actions associées à une investigation, ou de prendre connaissance de résultats qui peuvent nécessiter un changement d'idées préexistantes. Tout cela est ancré dans un contexte culturel qui peut favoriser ou inhiber le développement de la compréhension du phénomène scientifique que l'investigation se propose de faire comprendre

La reconnaissance de cette interdépendance entre les connaissances et les compétences a conduit à proposer que l'investigation serait mieux définie en termes de *pratiques* : des séries complexes d'actions conduisant à une pratique et une vision des sciences comme formant « un ensemble de connaissances bien enracinées dans les faits ». <sup>18</sup> Toutefois, dans ce livre, nous continuons d'utiliser le terme « aptitude » (*skill* en anglais) et « compétence » (*competence* en anglais) de façon interchangeable, en suivant la convention de l'OCDE :

Dans le contexte de la Stratégie de l'OCDE sur les compétences, les termes *aptitude* et *compétence* sont utilisés de façon interchangeable. Par aptitude (ou compétence), on entend le faisceau de connaissances, les attitudes et les capacités qui permettent à un individu d'exercer une activité ou une tâche toujours avec succès au sens large ou étroit ; l'aptitude peut être construite et amplifiée par l'apprentissage.<sup>19</sup>

## Le motif

La pédagogie d'investigation est complexe et n'est pas un choix facile. Nous nous efforçons de la mettre en œuvre parce que nous sommes persuadés qu'elle favorise la compréhension et le développement des compétences requises des élèves pour répondre aux exigences de la vie au XXI<sup>e</sup> siècle. Il est largement accepté<sup>20</sup> que l'enseignement des sciences doit permettre aux élèves de développer des concepts clés des sciences (les *notions-clés* majeures)<sup>21</sup> qui leur permettent de

18 National Research Council (2012) *A Framework for K-12 Science Education*. Washington DC: National Academies Press. p44

19 OECD (2011) *Towards an OECD Skills Strategy*. Paris: OECD. P7 footnote

20 OECD (2003), *The PISA 2003 Assessment Framework* Paris: OECD p132.



comprendre les événements et les phénomènes pertinents de leur vie présente et future. Les élèves doivent également développer une compréhension de la façon dont les idées et les connaissances scientifiques sont obtenues, et des compétences et des attitudes impliquées dans la recherche et l'utilisation des preuves. Les jeunes auront à faire plus de choix que ceux qui ont vécu dans les décennies précédentes. Ils auront besoin de développer les compétences, la volonté, la souplesse de la pensée et de l'énergie nécessaire pour prendre des décisions efficaces. La capacité de continuer à apprendre tout au long de la vie est reconnue comme essentielle pour les générations futures et donc elle doit avoir sa place dans l'enseignement du monde entier, selon l'OCDE :

**Les élèves ne peuvent pas apprendre à l'école tout ce dont ils auront besoin de connaître à l'âge adulte. Ce qu'ils doivent acquérir, ce sont des conditions préalables pour un apprentissage réussi dans la vie future. Ces conditions préalables sont à la fois d'ordre cognitif et motivationnel. Les élèves doivent être capables d'organiser et de réguler leur propre apprentissage, d'apprendre individuellement et en groupe, et de surmonter les difficultés dans le processus d'apprentissage. Cela signifie qu'ils doivent être conscients de leurs propres processus de pensée, des stratégies et des méthodes d'apprentissage.<sup>22</sup>**

En outre, on reconnaît l'importance de développer les compétences, les attitudes, les connaissances et la compréhension, tous ces facteurs étant considérés comme plus importants que l'accumulation de connaissances factuelles en grande quantité. La connaissance du contenu peut être trouvée facilement à partir des sources d'information largement disponibles grâce à l'utilisation des ordinateurs et de l'Internet en particulier. Ce dont les apprenants ont besoin, ce sont les compétences nécessaires pour accéder à ces sources et comprendre comment sélectionner un contenu pertinent en lui attribuant un sens.

L'apprentissage est une activité sociale où le langage joue un rôle clé. L'interaction avec les autres est souvent le résultat d'une compréhension commune des idées où un individu n'aurait pas réussi seul. Les idées des élèves, formées par l'expérience directe, doivent être communiquées, ce qui implique l'utilisation de mots porteurs de sens pour les autres. Le processus dans lequel les idées sont exprimées par le biais de conversation ou d'écriture signifie souvent que les idées doivent être reformulées en tenant compte du sens des mots que d'autres donnent. Il est également nécessaire d'apprendre que la science utilise des mots ayant des significations précises, qui sont différentes de leur utilisation commune dans le langage courant. La science utilise les mathématiques et d'autres symboles abstraits pour quantifier les observations du monde. Nous reviendrons sur le rôle du langage et surtout des échanges oraux dans le chapitre 5.

## Les objectifs

S'il est important pour les élèves d'apprendre les méthodes d'apprentissage et de développement des compétences d'investigation, un équilibre doit exister entre l'apprentissage conceptuel et la manière de s'y prendre pour apprendre. Apprendre la manière de répondre à une question n'est pas suffisant à soi seul, la question doit également trouver une réponse. En outre, trouver la réponse à la question ne suffit pas non plus, car c'est seulement par la manière dont cette réponse est formulée que l'élève pourra apprendre dans de nouveaux contextes.

En résumé, les élèves doivent élaborer à travers l'enseignement des sciences :

21 Bell D. Deves R. Dyasi H. Fernandez de la Garza G. Harlen W. Léna P. Millar R. Reiss M. Rowell P. Wei Yu, *10 notions-clés pour enseigner les sciences de la maternelle à la 3ème*. Le Pommier, traduit de l'anglais Harlen W (Ed) (2010) Principles and Big Ideas of Science Education (2009).

22 OECD (2000) Measuring Student Knowledge and Skills: A new Framework for Assessment. Paris: OECD. p90

- La compréhension des idées scientifiques fondamentales
- La compréhension de la nature des sciences, l'investigation scientifique et le raisonnement
- Les compétences scientifiques de collecte et d'utilisation des preuves
- Les attitudes scientifiques, à la fois dans les sciences et envers les sciences
- Des compétences qui favorisent l'apprentissage tout au long de la vie
- La capacité de communiquer en utilisant un langage et des représentations appropriées qui peuvent être sous forme d'un langage écrit, oral ou mathématique
- L'appréciation de la contribution des sciences à la société et l'application des sciences dans la technologie et l'ingénierie.

Une démarche d'investigation, si elle est conduite de manière efficace, offre la promesse d'atteindre ces objectifs à un degré plus élevé que les approches traditionnelles de l'enseignement des sciences.<sup>23</sup> La réserve, critique, est qu'elle doit être « conduite de manière efficace ». La complexité de l'ESFI, comme indiqué précédemment, lance un défi considérable. Sa mise en œuvre peut exiger un changement fondamental dans plusieurs aspects de la pédagogie, tels que l'agencement de l'espace d'apprentissage (afin que les élèves puissent travailler en collaboration), les questions des enseignants, les commentaires donnés aux élèves, la nature de leur interaction avec les élèves et l'interaction des élèves avec les objets et les phénomènes étudiés. Examinons donc le degré de changement qui peut être exigé en comparant les actions des élèves engagés dans un apprentissage fondé sur l'investigation, dans l'encadré 3, avec ceux de l'enseignement des sciences par transmission, dans l'encadré 4.

L'enseignement en tant que « transmission de faits » était le mode prédominant à l'époque où l'objectif principal de l'enseignement des sciences était de fournir aux futurs scientifiques les connaissances jugées essentielles, au lieu d'offrir à tous les élèves des opportunités pour atteindre les objectifs énumérés ci-dessus. Les sciences ont été transmises aux élèves « prêtes à l'emploi » plutôt qu'avec des expériences et des connaissances construites par l'action. Cependant, nous ne prétendons pas que l'investigation soit la seule forme de pédagogie adaptée à l'enseignement des sciences. Certains sujets, comme la compétence d'utiliser des équipements, des noms, des conventions ou des symboles, sont mieux enseignés directement ; il existe des occasions où l'investigation contribue à donner un sens à l'expérience sans être néanmoins l'approche exclusive. Cependant, lorsque l'objectif porte sur la compréhension de notions scientifiques, l'investigation a un rôle clé dans l'enseignement des sciences.

---

23 Minner, D.D., Levy, A. J, and Century, J. (2010) Inquiry-Based Science Instruction—What Is It and Does It Matter? Results from a Research Synthesis Years 1984 to 2002, *Journal of Research in Science Teaching*, 47 (4) 474-496

*Encadré 3: Les activités des élèves:  
apprentissage par l'investigation*

- Les élèves cherchent les réponses aux questions comme s'il s'agissait de leurs propres questions, même si elles ont été introduites par l'enseignant
- Ils ne connaissent pas la réponse **aux** questions étudiées
- Ils connaissent assez le sujet pour s'engager dans la question
- Ils font des prédictions sur la validité de leurs idées lorsqu'elles émergent
- Ils participent à la conception et l'organisation des investigations engagées pour tester leurs prédictions
- Ils conduisent des investigations par eux-mêmes
- Ils utilisent des sources et des méthodes de collecte de données, qui sont pertinentes pour tester leurs prédictions
- Ils discutent des résultats par rapport aux attentes ou aux prévisions initiales
- Ils tirent des conclusions et essaient d'expliquer ce qu'ils ont trouvé
- Ils comparent leurs résultats et conclusions à ceux des autres
- Ils prennent des notes et constituent des dossiers au cours de leur travail
- Ils s'engagent dans des discussions sur les méthodes employées et sur les résultats de leur investigation

*Encadré 4: Les activités des élèves:  
apprentissage par les méthodes de  
transmission*

- Les activités des élèves suivent une séquence définie par un manuel ou par l'enseignant avec peu d'attention portée à la mise en contexte du sujet étudié par rapports aux questions des élèves
- Ils peuvent lire des descriptions de méthodes d'investigation, mais ils ont peu de chance d'expérimenter le processus eux-mêmes
- Ils peuvent observer des démonstrations faites par les enseignants, mais peuvent ne pas comprendre les raisons de ce qui est fait
- Lorsqu'ils entreprennent des activités pratiques, ils suivent les instructions qui participent peu au choix concernant ce qu'il faut faire
- Les expériences observées et conduites visent à confirmer une conclusion déjà connue : "l'expérience pour montrer que ..."
- Ils ne savent pas toujours pourquoi certaines étapes d'une expérience ou d'une investigation doivent être effectuées
- Ils écrivent des comptes rendus d'investigation sous une forme structurée, souvent copiée à partir d'un livre ou dictée par l'enseignant
- Ils écrivent la «bonne réponse» sans avoir observé ce qui se passe en réalité
- Ils travaillent seuls ou par deux et sont peu enclins à parler de leur travail

## Chapitre 3

# Les objectifs et les utilisations de l'évaluation des élèves

Comme mentionné au chapitre 1, toute évaluation des élèves implique la production, la collecte, l'interprétation et la communication de données. Le processus est semblable, que l'objectif vise à apporter un soutien à l'apprentissage ou qu'il doive conduire à un rapport. Dans ce chapitre, nous explorons ces deux principaux objectifs d'une évaluation des élèves : celui d'apporter un soutien à l'apprentissage et celui de synthétiser sous forme de rapport ou d'appréciation. Nous examinons également s'il serait possible et avantageux d'utiliser les données recueillies dans un but précis tout en ayant un autre présent à l'esprit. La différence entre l'évaluation formative et l'évaluation sommative des élèves est essentiellement déterminée par l'*utilisation* des données. Par exemple, un test en classe peut être utilisé de manière formative pour aider les élèves et les enseignants à identifier les connaissances déjà acquises sur un nouveau sujet, ou de manière sommative pour juger sous forme d'un rapport sur ce qu'ils ont appris. Ainsi, pour être correct, nous devrions parler de « l'utilisation formative d'une évaluation des élèves » et de « l'utilisation sommative d'une évaluation des élèves », mais la commodité autorise les titres plus courts que nous employons ici. Après avoir examiné la nature et l'importance de ces deux objectifs de l'évaluation des élèves, nous réfléchissons à leur relation. Cette discussion se poursuivra dans le chapitre 6, après avoir examiné certaines procédures de la mise en œuvre de l'évaluation formative et l'évaluation sommative des élèves.

### Les objectifs et les utilisations

Voici une question fondamentale à poser pour toute évaluation des élèves : quel est son principal objectif ?<sup>24</sup> Il y a généralement deux réponses à cette question :

- soutenir les élèves pendant leur apprentissage
- découvrir ce qu'ils ont appris à un moment donné

Elles sont décrites comme les objectifs formatifs et sommatifs de l'évaluation des élèves.

L'évaluation formative des élèves a pour but de faciliter l'apprentissage, par conséquent elle est aussi appelée « l'évaluation des élèves *pour* l'apprentissage ». Elle implique des processus de « recherche et d'interprétation de données par les élèves et leurs enseignants afin de déterminer le niveau des élèves dans leur apprentissage, le niveau qu'ils doivent atteindre et les méthodes pour y arriver ».<sup>25</sup>

L'évaluation sommative des élèves a pour but de synthétiser ce qui a été appris à un moment donné sous forme de rapport. Pour cette raison, elle est aussi appelée « l'évaluation *de* l'apprentissage des élèves ». Elle implique des processus de synthèse, en examinant des acquis scolaires au fil du temps, et/ou en les contrôlant par des tests ponctuels.

Dans le cas de l'évaluation formative des élèves, l'utilisation des données vise principalement à les aider dans leur apprentissage. Si les informations recueillies ne sont pas utilisées pour prendre des décisions qui facilitent l'apprentissage, cette évaluation ne peut pas être considérée comme une

24 Stobart, G. (2008) *Testing Times. The uses and abuses of assessment*. London: Routledge

25 Assessment Reform Group (ARG) (2002) *Assessment for Learning: 10 Principles*. [www.assessment-reform-group.org](http://www.assessment-reform-group.org)

évaluation formative des élèves. En revanche, les données de l'évaluation sommative des élèves peuvent être utilisées de plusieurs façons, en les appliquant soit à des élèves individuels, soit aux résultats regroupés de groupes ou de populations d'élèves, sans que la validité en soit assurée. En effet, comme indiqué au chapitre 1, la validité n'est pas une propriété d'un instrument ou d'une procédure d'évaluation des élèves, elle dépend de son utilisation et des conclusions tirées à partir des résultats de cette dernière. Les résultats d'un test de rappel de connaissances n'indiquent pas le niveau dans toute la matière enseignée: un test d'arithmétique ne doit pas être utilisé pour indiquer le niveau de mathématiques ni pour mesurer la qualité de l'enseignement.

Aussi évident que ces points puissent paraître, les pratiques abusives dans l'usage des données issues de tests existent véritablement. Par exemple, Newton<sup>26</sup> a identifié seize utilisations des résultats des tests nationaux en Angleterre. Elles incluent l'évaluation systémique du programme, l'établissement des objectifs pour les élèves et les écoles, le suivi scolaire et la sélection des élèves, le choix de l'école pour les parents et même la valorisation des immeubles dans les quartiers des écoles ayant des résultats plus ou moins élevés. Dans certains cas, il est légitime de recourir à des mesures de performance des élèves pour porter un jugement plus global (par exemple, comme un élément d'évaluation systémique de l'école). Cependant, leur utilisation en tant que mesure unique, en particulier dans un contexte de récompenses et de pénalités, peut conduire à exagérer les performances mesurées. Nous reviendrons sur ce point au chapitre 4 lorsque nous examinerons l'utilisation de l'ensemble des données pour une évaluation systémique de l'école, du système ou de l'identification des objectifs.

Nous allons maintenant examiner la nature et l'importance des deux principaux objectifs de l'évaluation formative et l'évaluation sommative des élèves. Les chapitres 5 et 6 traiteront des méthodes de leur mise en œuvre.

## L'évaluation formative des élèves

### Qu'est-ce que c'est ?

Parmi les différentes définitions de l'évaluation formative des élèves proposées au cours des deux dernières décennies, Wiliam<sup>27</sup> suggère que les principales caractéristiques sont réunies dans la définition suivante :

L'exercice en classe est formatif dans la mesure où les preuves indiquant la performance des élèves sont provoquées, interprétées et utilisées par les enseignants, les apprenants ou leurs pairs, pour choisir les étapes suivantes de l'apprentissage qui seront alors probablement mieux construites qu'en absence de preuves.<sup>28</sup>

La figure 1<sup>29</sup> représente les processus impliqués dans un cycle d'événements. L'évaluation formative des élèves n'est pas occasionnelle, elle fait partie intégrante du processus qui consiste, dans l'enseignement, à prendre en permanence des décisions. Les activités représentées par A, B, et C visent les objectifs de la leçon ou d'une série de leçons sur un sujet donné. Ces objectifs, partagés avec les élèves et l'enseignant, sont exprimés dans des termes *spécifiques*. Par exemple, dans un cours de sciences, ils pourraient être « planifier et mener une investigation sur les conditions favorables des cloportes ». Le travail des élèves dans l'activité A, orienté vers les objectifs, donne à l'enseignant et aux élèves l'occasion d'obtenir des preuves d'une progression vers les objectifs.

26 Newton, P.E. op cit pp270-272

27 Wiliam, D. (2009) An integrative summary of the research literature and implications for a new theory of formative assessment, in (eds) H. L. Andrade and G. J. Cizek, *Handbook of Formative Assessment*, New York: Taylor and Francis

28 Black, P. and Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21 (1), 5-13. p9

29 Adapté de Harlen, W. (2006) *Teaching, Learning and Assessing Science 5-12*. London: Sage. p 87.

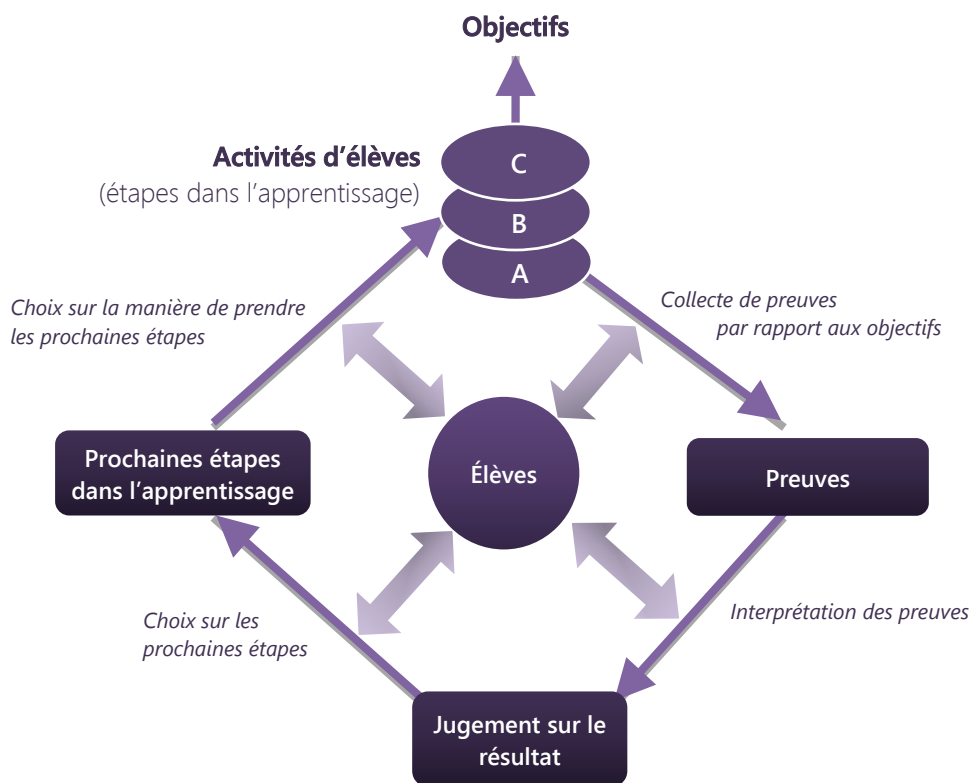


Figure 1 : L'évaluation des élèves liée aux objectifs formatifs (adapté de Harlen, 2006)

Afin d'interpréter les preuves, dans cet exemple, l'enseignant et les élèves ont besoin de connaître la signification d'une « bonne organisation », les élèves doivent donc avoir une certaine compréhension des critères d'évaluation qui seront appliqués à leur travail (l'investigation planifiée tient-elle compte de toutes les variables pertinentes? Quelles sont les preuves et comment les recueillir?). Le jugement conduit au choix des prochaines étapes pertinentes qui peuvent être soit l'intervention, soit simplement la continuation. Comme William le souligne, « l'évaluation formative des élèves n'a pas besoin de modifier les instructions pour être formative ; elle peut simplement confirmer que le plan d'action proposé est en effet le plus approprié »<sup>30</sup>. L'activité B est le résultat de cette décision et la source d'un nouveau cycle d'obtention et d'interprétation de preuves.

Les élèves sont au centre du processus, puisqu'ils sont les acteurs de l'apprentissage. Les flèches à deux têtes reliant les élèves aux différentes parties du cycle d'évaluation des élèves indiquent qu'ils reçoivent des commentaires de l'enseignant et apportent des informations en même temps. Ils participent aux décisions le cas échéant par l'auto-évaluation et l'évaluation des élèves par les pairs.

Dans l'évaluation formative des élèves, les jugements sur les progrès et les choix des étapes suivantes tiennent compte des circonstances de l'apprentissage, des efforts du passé de chaque élève, ainsi que de leur capacité par rapport aux objectifs du travail à un moment donné. Par conséquent, ces jugements sont à la fois basés sur les élèves et sur les critères (Voir Chapitre 1). Cette modalité favorise l'apprentissage (car aucune comparaison n'est faite entre les élèves dans l'évaluation formative des élèves) bien plus que l'application de normes identiques à tous les élèves, une façon de faire qui représente une source de démotivation pour les élèves faibles.

30 Leahy, S. and William D. (2012) From teachers to schools: scaling up professional development for formative assessment, in (ed) J. Gardner *Assessment and Learning*. London: Sage. 49-71, p 51.

Les actions indiquées par les flèches dans la figure 1 ne représentent pas les « étapes » dans une leçon ni nécessairement le résultat des décisions de l'enseignant. Elles représentent la réflexion sur ce que et comment les élèves apprennent et utilisent ceci comme soutien pour un apprentissage ultérieur. Dans certains cas, il se peut que les enseignants et les élèves décident ensemble d'une action immédiate. Dans d'autres cas, l'enseignant peut noter le besoin et y répondre plus tard.

La mise en œuvre de l'évaluation formative des élèves sous-entend que tous les aspects d'une leçon ne peuvent pas être planifiés à l'avance. Par définition, si les idées existantes des élèves doivent être prises en compte, certaines décisions dépendront de la nature de ces idées. Certaines idées peuvent être anticipées à partir de l'expérience des enseignants et des résultats de recherche qui ont été intégrés dans les documents pédagogiques, mais pas toutes. L'enseignant n'a pas besoin d'un contenu prescrit de cours, mais d'un ensemble de stratégies à déployer de manière appropriée dans des circonstances spécifiques. Quelques exemples de ces stratégies sont illustrés dans le chapitre 5.

Le commentaire est un élément essentiel de l'évaluation formative des élèves. Les commentaires à double sens, de l'enseignant aux élèves et des élèves à l'enseignant, impliquent une vision de l'apprentissage comme étant un processus dans lequel la compréhension est activement construite par les élèves :

- Les commentaires de l'enseignant aux élèves offrent aux élèves des informations pour les aider à prendre les dispositions nécessaires en vue d'améliorer leur compréhension ou leurs compétences. L'objet et la forme des commentaires doivent être soigneusement analysés par l'enseignant. L'objet des commentaires a une influence sur l'attention des élèves et les formes sont déterminantes pour améliorer, ou non, l'apprentissage. Si l'apprentissage n'est assimilé qu'à « ce qui est enseigné », les commentaires de l'enseignant aux élèves portent alors sur la qualité ou la réussite du travail plutôt que sur une méthode pour l'améliorer : dans une telle vision de l'apprentissage, l'évaluation formative des élèves ne joue aucun rôle.
- Les commentaires des élèves aux enseignants sont nécessaires afin que les enseignants puissent ajuster les défis proposés aux élèves pour que ces défis ne soient ni trop exigeants et empêchant de réussir, ni trop simples et manquant d'intéresser les élèves. Juger de la capacité des élèves à prendre certaines dispositions en étant aidés (la *zone de proximale de développement*,<sup>31</sup> voir chapitre 4) en utilisant leurs commentaires est une tâche complexe et difficile pour les enseignants. De nombreux enseignants ont besoin de soutien dans ce domaine, s'ils souhaitent utiliser les commentaires de leurs élèves afin de pouvoir revenir dessus et d'optimiser l'apprentissage de leurs élèves.

En résumé, les principales composantes de l'évaluation formative des élèves sont listés dans l'encadré 5. Au chapitre 5, nous examinerons la mise en pratique de cette évaluation.

### **Pourquoi est-elle importante ?**

L'importance de l'évaluation formative des élèves est justifiée par son efficacité à améliorer l'apprentissage. Les études empiriques sur l'évaluation des élèves en classe ont fait l'objet de plusieurs travaux de recherche. L'étude de Black et de Wiliam (1998) a attiré l'attention du monde entier en tentant de quantifier l'impact de l'utilisation d'une évaluation formative des élèves. Depuis, il y a eu un certain nombre d'études et d'investigations qui ont justifié l'affirmation suivante de Leahy et de Wiliam :

Nous constatons que dans des matières scolaires différentes, des pays différents et pour des apprenants d'âges différents, l'utilisation de l'évaluation formative des élèves semble être associée à des améliorations considérables de l'apprentissage. L'estimation de

31 *zone proximale de développement*, telle qu'introduite par Vygotski (N.d.T.)

*Encadré 5 : Les composants essentiels de la mise en œuvre d'une évaluation formative des élèves*

- Les élèves expriment et communiquent leur compréhension et leurs compétences par un dialogue en classe, initié par des questions ouvertes centrées sur la personne
- Les élèves ont une bonne compréhension des objectifs de leur travail ainsi que la notion d'un travail de qualité
- Les commentaires donnés aux élèves apportent des conseils d'amélioration ou de progression en évitant la comparaison avec les autres élèves
- Les élèves sont impliqués dans l'auto-évaluation pour participer à l'identification de ce qu'ils doivent faire pour s'améliorer ou progresser
- Le dialogue entre l'enseignant et les élèves encourage la réflexion sur leur apprentissage
- Les enseignants utilisent en continu les informations sur l'apprentissage, pour adapter leur enseignement afin que tous les élèves aient une réelle occasion d'apprendre<sup>32</sup>

l'ampleur de ces gains pourrait être difficile ... mais il semble raisonnable de conclure que l'utilisation de l'évaluation formative des élèves peut augmenter le taux d'apprentissage des élèves de 50 à 100%.<sup>33</sup>

Stobart<sup>34</sup>, toutefois, a un avis plus prudent en soulignant que, à l'exception d'une étude de Wiliam et al (2004) sur l'impact d'un projet de recherche-action sur la réussite des élèves, « il y a, à ce jour, peu de preuves empiriques directes portant sur l'impact de [l'évaluation formative des élèves] sur leur réussite ». Il note que la plupart des études sur l'évaluation des élèves ont mis l'accent sur l'ampleur des changements dans la pratique des enseignants et les attitudes des élèves en se focalisant sur la participation plutôt que sur l'apprentissage conceptuel des élèves. Néanmoins, on peut faire valoir que ces changements sont des étapes nécessaires pour améliorer l'apprentissage. Par ailleurs, au-delà de subtils changements dans la pédagogie lors de la mise en œuvre d'une évaluation formative des élèves, son influence sur les mesures de l'apprentissage des élèves rendent son impact difficile à détecter. En effet, Wiliam et al (2004) soulignent que les comparaisons sur lesquelles sont fondées leurs affirmations ne sont « pas toutes solides ».

L'importance de l'évaluation formative des élèves dans le cadre de l'ESFI réside dans le fait qu'évaluation formative des élèves et ESFI visent tous deux à une véritable compréhension et à l'acquisition des compétences nécessaires à l'apprentissage. Un enseignement qui vise le développement de la compréhension implique la prise en compte des idées initiales et des compétences existantes des élèves, puis la mise en avant de leurs progrès en ajustant les défis pour prendre en compte ces idées initiales.<sup>35</sup> La pratique de l'évaluation formative des élèves, à travers la collecte de données par les enseignants et les élèves pendant l'apprentissage ainsi que la remontée d'informations pour réguler le processus d'enseignement et d'apprentissage, est clairement en cohérence avec les objectifs et les pratiques d'un apprentissage fondé sur l'investigation. Elle donne aux élèves le sentiment d'une appropriation de leur apprentissage en mettant en avant l'auto-évaluation et la participation aux choix des étapes suivantes ; elle aide donc les élèves à assumer la responsabilité de leur apprentissage à l'école et au-delà.

32 Harlen, W. (2007) op cit p121

33 Leahy, S. and Wiliam, D. (2012) op cit p52

34 Stobart, G. (2008) op cit p154

35 Bransford, J.D., Brown, A. and Cocking, R.R. (Eds) (2000) *How People Learn, Brain, Mind, Experience and School*. Washington, D.C.: National Academy Press.



## L'évaluation sommative des élèves

Nous venons de décrire l'*évaluation formative* des élèves comme ayant un rôle positif dans l'apprentissage, car nous avons tendance à la considérer comme le « bon » côté de l'évaluation des élèves. Quant à l'*évaluation sommative* des élèves qui joue un rôle différent, elle est souvent présentée comme le « mauvais » côté. Ceci est regrettable à plusieurs égards. Tout d'abord, bien que l'évaluation sommative des élèves ne soit pas, à la différence de l'évaluation formative, destinée à avoir un impact direct sur l'apprentissage, elle peut néanmoins être utilisée pour apporter à l'apprentissage un soutien d'une manière moins directe, mais nécessaire, comme par exemple par l'établissement d'une synthèse sur l'apprentissage pour informer leurs futurs enseignants lorsque ces élèves passent d'une classe ou d'une école à l'autre. Ensuite, elle permet aux enseignants, aux parents et aux écoles de garder une trace du parcours des élèves, soit en tant qu'élève individuel, soit en tant que membres de certains groupes (tels que ceux qui sont très performants et ceux qui ont besoin d'une aide spéciale). Troisièmement, elle fournit des données qui, en même temps que les facteurs contextuels, peuvent être utilisées pour l'évaluation systémique et l'amélioration de l'école. La mauvaise réputation de l'évaluation sommative des élèves découle de l'utilisation inappropriée des données qui ne reflètent pas pleinement les objectifs de l'apprentissage. Le danger est considérable dans le cas de l'ESFI en raison de ses objectifs qui sont difficiles à évaluer par des méthodes conventionnelles. Il est donc encore plus important d'examiner les moyens de collecter des informations fiables dans le cadre de l'évaluation sommative des élèves.

### Qu'est-ce que c'est ?

L'évaluation sommative des élèves est le nom donné à une évaluation des élèves effectuée dans le but de rendre compte des acquis à un moment donné. Elle peut avoir, et d'ailleurs a souvent, un certain impact sur l'apprentissage. Ses résultats peuvent être utilisés dans l'enseignement mais ce n'est pas son objectif principal. Nous réexaminerons sa relation avec l'évaluation formative des élèves plus tard. Pour l'instant, le processus est représenté comme la transmission d'informations uniquement pour synthétiser les acquis comme indiqué dans la figure 2. L'emploi du terme « preuve » (en anglais *evidence*) veut marquer ce qui apparaît comme un résultat de la pédagogie, dévoilé de façon convaincante par le test ou l'activité. Dans les paragraphes qui suivent, nous utiliserons le terme *preuve/résultat* pour marquer la différence, qui s'impose, avec une preuve au sens mathématique, dont il n'est évidemment pas question ici.

Les preuves/résultats proviennent des tests, de tâches spéciales ou d'activités habituelles et peuvent être collectées par des moyens divers et variés : des réponses écrites, des objets construits par des élèves, des portfolios, des observations d'actions, des discussions ou des exposés de travail.

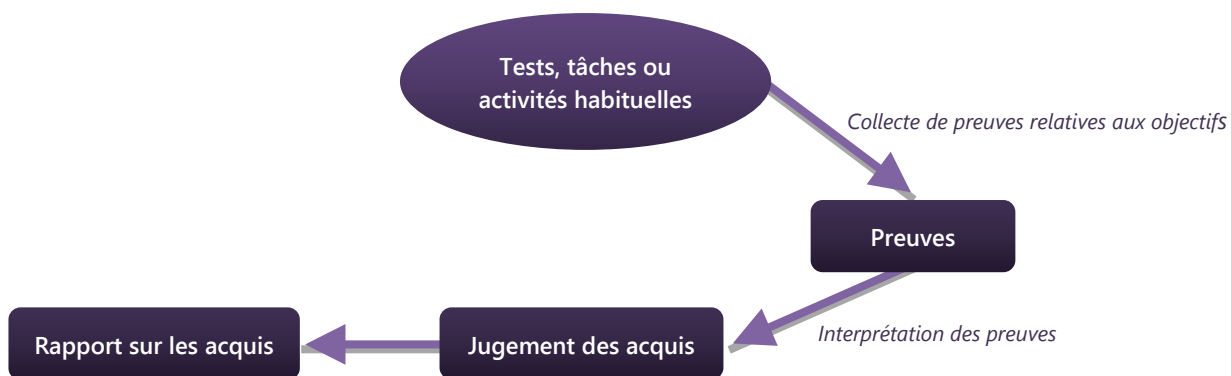


Figure 2: L'évaluation des élèves liée aux objectifs sommatifs (adapté de Harlen, 2006)

La collecte de preuves/résultats portant sur les performances de l'ensemble des compréhensions et des compétences pertinentes est clairement la partie la plus importante du processus, car sans elle, le rapport final sur le niveau atteint risque de ne pas fournir une information fiable sur les résultats des élèves face aux objectifs de l'apprentissage. Les avantages et les inconvénients de l'utilisation de différentes sources de données sont discutés dans le chapitre 6.

Les preuves/résultats sont interprétées en comparant avec des critères ou les normes des objectifs généraux plutôt que ceux des leçons ou des sujets spécifiques, comme dans le cas de l'évaluation formative des élèves. Cette notation peut être effectuée par l'enseignant ou un organisme extérieur, comme pour certains tests et examens nationaux. Les élèves ne jouent un rôle dans ce processus que dans les tests les plus informels en classe. Les élèves sont tous jugés d'après les mêmes critères ou systèmes de notation (rubriques), alors que, comme indiqué précédemment, les critères de l'évaluation formative des élèves peuvent être ipsatifs ou fondés sur les élèves (voir chapitre 1) afin de les aider à reconnaître leur progrès à partir de points de départ différents.

L'interprétation réduit la richesse de la performance réelle à un score, une catégorie ou une note représentative, d'où la perte d'une grande partie de l'information. Selon l'utilisation du résultat, le processus d'interprétation peut inclure certaines procédures pour augmenter la fiabilité du résultat. Lorsque les résultats sont utilisés pour comparer les élèves entre eux, en particulier lorsqu'il s'agit d'une grande sélection ou d'un classement, des mesures sont prises pour vérifier la notation et les jugements modérés par des enseignants ou des examinateurs. Bien que l'évaluation sommative des élèves soit essentiellement effectuée en classe par les enseignants, il est possible que les preuves soient collectées et utilisées sur un large éventail de types de réussite.

L'absence de référence aux élèves dans la figure 2 montre que, d'une manière générale, ceux-ci ne jouent pas de rôle dans ce type d'évaluation. Toutefois, si le processus est ouvert, si les critères d'évaluation sont partagés avec les élèves et les utilisateurs de résultats sans que cela limite le contenu du test, alors les élèves ont la possibilité de jouer un rôle dans le processus, comme par exemple en sélectionnant des éléments dans un portfolio. Il demeure, bien sûr, l'obligation de s'assurer que les jugements sont fiables et basés sur des critères identiques pour tous les élèves. Nous verrons ce point dans le chapitre 6.

La forme du rapport qui résultera de l'évaluation dépend en grande partie de la nature de la tâche, de la base du jugement et de ses lecteurs. Les scores numériques de tests représentent une sommation faite sur un ensemble de questions variées. Le même total peut être obtenu par plusieurs moyens, les scores ont donc peu de signification pour refléter les connaissances et les capacités des élèves. Ils donnent aussi une fausse impression de précision. Les scores peuvent être utilisés directement pour classer les élèves, ce qui est seulement utile dans le contexte d'une sélection car une position dans un classement ne donne aucune indication sur l'apprentissage.

En théorie, un rapport décrivant les performances relatives à des niveaux progressifs, établis d'après les critères, peut fournir une indication plus significative sur le niveau atteint par les élèves (voir chapitre 6). Cependant, pour qu'un rapport soit significatif, la présence d'un profil est préférable à une note unique ou un niveau global combinant plusieurs domaines. L'abréviation de « niveaux », les appellations données aux niveaux progressifs, peuvent être utiles à certaines fins. Mais dans le but d'informer les parents et les élèves, les niveaux doivent être expliqués ou accompagnés par l'explicitation de détails portant sur les capacités des élèves. De plus, comme indiqué plus loin (chapitre 6, page 80), l'utilisation de niveaux peut avoir des conséquences négatives sur la motivation et l'apprentissage des élèves.

Quelques pratiques caractéristiques de l'évaluation sommative des élèves sont résumées dans l'encadré 6.

*Encadré 6: Les composants essentiels de la mise en œuvre d'une évaluation sommative des élèves*

- Les élèves peuvent être impliqués dans des tâches ou des tests particuliers, soit au sein ou en complément de leurs tâches habituelles
- Cette évaluation est pratiquée à certains moments spécifiques et identifiés, se distinguant ainsi du cadre régulier et normal de l'apprentissage
- Elle vise à mesurer la réussite de grands objectifs généraux plutôt que celle de certaines activités d'apprentissage
- Elle vise la réussite de tous les élèves, jugés d'après les mêmes critères ou le même système de notation
- Elle exige des mesures pour garantir la fiabilité
- Elle n'offre que des possibilités restreintes pour une auto-évaluation des élèves

**Pourquoi est-elle importante ?**

Plusieurs raisons ont déjà été données au début de cette section. La raison la plus convaincante pour expliquer son importance est celle de la nécessité. Elle est inévitable car des rapports sur l'apprentissage des élèves doivent être établis à intervalles réguliers. En revanche, l'évaluation sommative des élèves pourrait être envisagée, en un sens, comme une action volontaire, dans la mesure où il est possible d'enseigner sans elle. Mais les enseignants ont besoin de tenir des dossiers ou livrets résumant les résultats des élèves à des moments clés, comme à la fin des grands thèmes ou des semestres, pour les utiliser dans la planification de leur activité. Les parents et les futurs enseignants des élèves, à des moments de transition d'une classe à l'autre ou d'une école à l'autre, ont besoin de dossiers pour connaître les acquis antérieurs. Les directeurs et les gestionnaires des écoles doivent tenir des dossiers pour examiner les progrès des groupes d'élèves afin de réaliser l'auto-évaluation de l'école et la planification de son activité.

Elle est également importante parce que le contenu évalué est de facto considéré comme un signal important sur la qualité de l'apprentissage. Malheureusement, ce qui est évalué est souvent ce qui *peut* être évalué plutôt que ce qui *doit* être évalué. L'écart entre les deux est susceptible d'être particulièrement important dans le cadre de l'ESFI où les objectifs visent à la construction de la compréhension et du développement de « compétences utilisées par les scientifiques ». Nous verrons dans le chapitre 6 les « avantages et les inconvénients » de diverses méthodes d'évaluation sommative des élèves, mais il convient de noter qu'il n'y a pas de modalité ou de méthode « parfaite ». Toute évaluation des élèves n'est qu'un échantillon de contenu étudié, une approximation du niveau acquis et le résultat de plusieurs jugements subjectifs. Une meilleure compréhension du processus d'évaluation des élèves par tous les acteurs, des enseignants aux élèves, des politiques aux employeurs, pourrait aider à réduire la pression exercée par l'évaluation sommative des élèves sur les programmes scolaires. Nous abordons ce point au chapitre 4.

**La relation entre l'évaluation formative et l'évaluation sommative des élèves**

Une des raisons pour laquelle l'évaluation sommative des élèves a une réputation négative est la suivante : lorsque la performance mesurée devient le facteur dominant dans la classe, elle chasse les pratiques d'évaluation formative des élèves. La nature des commentaires donnés par les enseignants aux élèves va au-delà de ce qui serait requis pour franchir l'étape d'apprentissage suivante et modifie l'impression des élèves sur le souci de leurs enseignants à les aider.

Ceci est illustré dans le rapport d'étude suivant :

Roderick et Engel<sup>36</sup> ont montré qu'une école, proposant un niveau élevé de soutien, a été en mesure d'accroître les efforts et d'améliorer les résultats d'élèves en difficulté à un degré beaucoup plus élevé que dans une école comparable, qui fournissait un soutien faible aux élèves en situation similaire. Un soutien solide implique la création d'un environnement d'appui social et éducatif : faire des efforts pour augmenter le sentiment de l'auto-efficacité des élèves; se concentrer sur les objectifs d'apprentissage; rendre les objectifs plus explicites; utiliser l'évaluation des élèves pour aider les élèves à réussir, et créer des cartes cognitives qui mettent au clair les progrès. Les enseignants ont également fait preuve d'un sens aigu de responsabilité envers leurs élèves. En revanche, un soutien faible de la part des enseignants conduit ceux-ci à penser que les élèves ne peuvent atteindre les objectifs, à ignorer la nécessité d'introduire davantage d'activités significatives, à reconnaître insuffisamment les changements et les motivations des élèves, et à négliger la construction de liens personnels avec les élèves, en lien avec les objectifs de l'apprentissage.<sup>37</sup>

Pollard et al<sup>38</sup> ont remarqué que l'introduction de tests nationaux en Angleterre et l'obligation pour les enseignants d'assigner des niveaux aux élèves ont affecté leur rapport aux élèves et l'utilisation de l'évaluation formative des élèves. Les élèves étaient conscients que la réussite des tests comptait plus que les efforts bien que ces derniers étaient pris en compte. Il n'est donc pas étonnant que l'évaluation sommative des élèves ait une mauvaise réputation, même si le problème provient souvent de l'utilisation des résultats plutôt que de la nature de l'évaluation sommative elle-même. Comme nous l'avons noté, certaines évaluations sommatives des élèves sont nécessaires et inévitables : y a-t-il des solutions pour que l'évaluation sommative des élèves soit réalisée et utilisée sans avoir d'impact négatif sur l'évaluation formative des élèves ?

Une approche évidente consiste à réfléchir dans le sens inverse, c'est-à-dire, se concentrer sur l'évaluation formative des élèves pour voir s'il est possible d'obtenir dans le processus formatif les données utiles à un jugement sommatif. Traditionnellement, le système éducatif danois a toujours travaillé de cette façon. En absence de tests nationaux à l'école obligatoire, les enseignants des petites classes apportent des commentaires aux élèves sur leurs progrès dans un processus continu. Dans les classes supérieures, l'enseignant synthétise les progrès des élèves tous les quatre à six mois sous forme d'une note. A la fin de la scolarité obligatoire, la note la plus récente apparaît sur le certificat final. Pour certaines matières, un examen oral a lieu avec un évaluateur externe, généralement un enseignant d'une école voisine, et la note de ces examens est également inscrite sur le certificat final. Le problème ici est de garantir la fiabilité et la protection des élèves contre les enseignants hostiles.

Les deux listes de caractéristiques propres à l'évaluation formative des élèves (encadré 5) et à l'évaluation sommative (encadré 6) et les représentations des figures 1 et 2 montrent quelques différences fondamentales dans l'utilisation des preuves. Cependant, comme nous y avons déjà fait allusion, il est possible que les preuves recueillies à des fins sommatives soient réutilisées dans l'enseignement et l'apprentissage pour identifier les aspects à améliorer. Black et al<sup>39</sup> donnent des exemples d'enseignants qui utilisent des tests en classe pour permettre aux élèves d'identifier leurs

- 
- 36 Roderick, M. and Engel, M. (2001) The grasshopper and the ant: motivational responses of low achieving pupils to high stakes testing. *Educational Evaluation and Policy Analysis* 23: 197-228
- 37 Harlen, W. (2012a) The role of assessment in developing motivation for learning, in (ed) J. Gardner *Assessment and Learning*, pp171-184. p177
- 38 Pollard, A., Triggs, P., Broadfoot, P., Mcness, E. and Osborn, M. (2000) *What pupils say: changing policy and practice in primary education* (chapters 7 and 10). London: Continuum
- 39 Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2003). *Assessment for Learning: Putting it into Practice*. Maidenhead, England: Open University Press.

points faibles et de faire davantage d'effort. En pratique, les enseignants peuvent utiliser cette modalité surtout dans le contexte de tests passés en classe sur lesquels ils ont le contrôle total. Bien que certains tests et examens externes puissent être utilisés de cette manière, par l'examen des copies corrigées et des discussions avec les élèves, le processus risque de passer du développement de la compréhension à « un enseignement fondé sur les tests ».

Un exemple de combinaison de l'objectif formatif et sommatif dans une évaluation des élèves à grands enjeux est le *Senior Certificate* de l'État du Queensland en Australie, qui a adopté cette modalité depuis de nombreuses années. Il est important de mentionner que ce système est conçu pour servir les deux objectifs à la fois par la participation des élèves et par des procédures assurant la qualité. Un portfolio de preuves accumulées pendant les deux années de cours transmet les commentaires aux élèves, leur permettant d'améliorer leurs performances pendant le cours en indiquant ce qu'ils ont réalisé à la fin du cours. Voici l'explication que donne Maxwell :

Pour que cette modalité fonctionne, il est nécessaire d'exprimer les résultats attendus en termes de critères communs à tout l'apprentissage. Ensuite, il peut y avoir débat pour savoir si l'élève est en voie d'atteindre ces objectifs et ce qui doit être fait pour améliorer les performances lors d'une évaluation ultérieure, relativement aux mêmes critères. Puisque c'est l'élève qui construit son portfolio comportant les preuves de sa performance, une évaluation antérieure peut être remplacée par une évaluation ultérieure portant sur les mêmes dimensions sous-jacentes de l'apprentissage. L'objectif est de noter la situation actuelle de l'élève dans son parcours d'apprentissage, et non son niveau de départ ni sa moyenne sur l'ensemble des cours.<sup>40</sup>

Les critères d'évaluation des élèves sont publiés afin que les élèves, les parents ainsi que les enseignants puissent les connaître. Ils décrivent ce que les élèves sont capables de faire dans les différentes catégories et les sous-catégories divisées en cinq niveaux ou normes (voir l'exemple dans l'encadré 20 du chapitre 6). Les preuves du portfolio sont comparées avec les critères de jugements les plus appropriés. Les conditions essentielles de cette modalité consistent, d'une part, au temps des enseignants consacré à une modération qui garantisse la fiabilité des résultats, d'autre part à un respect du professionnalisme des enseignants. Il est également intéressant de noter que la réussite des élèves dans le Senior Certificate est distincte d'une évaluation systémique de l'école et des enseignants. Les écoles sont néanmoins encouragées à utiliser les données du processus de certification pour leur auto-évaluation et leur amélioration.

Ces exemples montrent des données de l'évaluation sommative des élèves qui fournissent des informations de manière formative, nouant ainsi un lien entre évaluation sommative et évaluation formative. Le corollaire consiste dans l'utilisation à des fins sommatives des données recueillies pendant l'évaluation formative. Puisque l'évaluation formative est entre les mains de l'enseignant, les preuves à utiliser dans l'évaluation sommative sont donc recueillies par l'enseignant : cela crée l'occasion de collecter plus de preuves que par le simple examen des activités en classe. Un exemple de double utilisation de données recueillies par les enseignants est donné dans le chapitre 6 (page 79).

Cependant, des différences fondamentales doivent être prises en compte lorsque les preuves recueillies sont utilisées de façon formative ou sommative. Les jugements formatifs sont susceptibles d'être fondés sur les élèves face aux objectifs immédiats du cours, tandis que les jugements sommatifs ne doivent être fondés que sur des critères larges, prenant en compte les objectifs à long terme. Une distinction doit être faite entre les preuves et les jugements, de sorte que les *preuves*,

40 Maxwell, G. (2004) 'Progressive assessment for learning and certification: some lessons from school-based assessment in Queensland' Paper presents at the third conference of the Association of Commonwealth Examination and Assessment Boards, Redefining the Roles of Educational Assessment, March, Nidi, Fiji. p2-3

utilisées dans l'évaluation formative, conduisent à des jugements utilisant des critères plus étendus, tels que définissant les niveaux ou les grades. Nous discutons les mesures additionnelles qui sont nécessaires pour garantir, avec un objectif sommatif, une utilisation fiable de l'évaluation des élèves par les enseignants au chapitre 6.

Un dernier point au sujet de la relation entre évaluation formative et évaluation sommative est de savoir s'il est intéressant de faire une distinction entre elles, ou s'il est préférable de considérer leur relation comme une seule dimension plutôt qu'une dichotomie.<sup>41</sup> Il existe différentes manières d'utiliser l'évaluation formative tout comme il existe différentes façons de collecter des preuves pour l'évaluation sommative. Certaines évaluations formatives impliquent une réponse immédiate aux actions des élèves, tandis que d'autres nécessitent une certaine réflexion et planification. La collecte de preuves fait-elle partie de l'activité d'apprentissage ou nécessite-elle d'être planifiée afin de savoir ce qui a été appris ? L'évaluation formative planifiée pourrait avoir une certaine similitude avec une évaluation sommative reposant sur les enseignants. Sa nature formative provient du mode d'utilisation des informations. Ainsi, il existe un certain flou dans la distinction entre évaluation formative et évaluation sommative: peut-être vaut-il mieux percevoir leur relation comme une seule dimension plutôt qu'une dichotomie. Cependant, l'importance de préserver cette distinction réside dans le rôle de l'évaluation des élèves pour faciliter l'apprentissage, car sans cela, toute évaluation des élèves devient sommative.

---

41 Harlen, W. (2012b) On the relationship between assessment for formative and summative purposes, in (ed) J. Gardner *Assessment and Learning*. London: Sage 87-102. pp 97-100



## Chapitre 4

# L'évaluation des élèves, la pédagogie et les programmes

L'une des principales raisons de la présente publication réside dans les relations bien connues entre l'évaluation des élèves, le contenu des programmes scolaires et la pédagogie (figure 3). Ces relations sont particulièrement importantes dans le cadre du développement de la compréhension et de la mise en œuvre de l'enseignement des sciences fondé sur l'investigation (ESFI). Les méthodes et le contenu traditionnels de l'évaluation des élèves reflètent rarement les grands objectifs de l'ESFI. Par conséquent, l'évaluation des élèves, telle que pratiquée actuellement, a tendance à freiner la mise en œuvre de cet enseignement. Le défi est alors de modifier les pratiques d'évaluation des élèves afin qu'elles aient un rôle de soutien dans ce mode d'enseignement des sciences.

Dans la première section de ce chapitre, nous examinons les impacts potentiellement positifs ou négatifs de l'évaluation des élèves sur le contenu et la pédagogie. Des preuves solides d'un impact négatif émergent d'évaluation poussée d'enseignants et d'écoles. Une telle utilisation est à rejeter, car il est injuste de se servir des résultats de tests comme le seul indicateur de qualité de l'enseignement. Cela ne signifie pas que les enseignants et les écoles ne doivent pas être tenus responsables des résultats de leurs élèves, mais comme nous le verrons dans la deuxième partie, d'autres informations doivent être prises en compte. La troisième partie de ce chapitre analyse le conflit entre un apprentissage fondé sur une évaluation traditionnelle des élèves et l'apprentissage implicite dans l'enseignement fondé sur l'investigation. Enfin, nous verrons que l'évaluation des élèves, si elle est bien conçue et bien élaborée, peut jouer un rôle positif dans l'éducation pour faire face à certains de ses principaux défis face aux problèmes mondiaux contemporains.

### Les impacts de l'évaluation des élèves

Les relations entre l'évaluation des élèves, la pédagogie et le programme sont souvent exprimées sous forme de triangle, en général avec « le programme » placé au sommet. Si le mot « programme » doit désigner tout ce qui est vécu par les élèves à l'école, alors chaque élément de la figure 3 fait

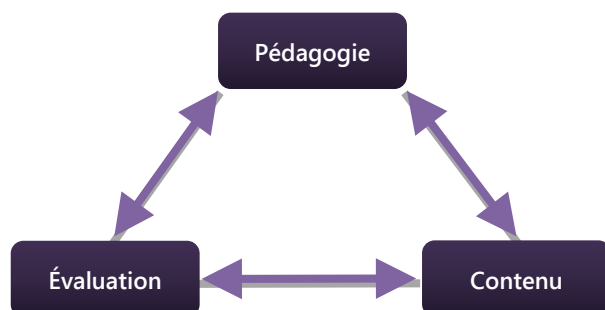


Figure 3: Les interactions entre les divers aspects de l'ensemble du programme

partie de l'un des aspects du programme. Cependant, l'utilisation la plus courante du terme « programme » se réfère implicitement au contenu, en le séparant de la pédagogie et de l'évaluation des élèves. Nous employons ici le terme « contenu du programme », car ce contenu n'est qu'un aspect du programme tel que vécu par les élèves, dont l'ensemble est constitué du contenu, de la pédagogie, de l'évaluation des élèves et des hypothèses sous-jacentes portant sur l'apprentissage.



Dans la figure 3, les flèches représentent ce qui est bien connu : le contenu de l'enseignement est influencé par la façon d'enseigner; le contenu et la méthode de l'évaluation des élèves influencent à la fois le contenu et la pédagogie. Ces interactions sont importantes, car il ne sert à rien de préconiser une mise en œuvre de l'enseignement fondé sur l'investigation avec une évaluation autoritaire des élèves (que ce soit des tests ou des jugements des enseignants) ou un programme surchargé. Il ne sert à rien de suggérer que le contenu doit cibler les principales *notions-clés*, si l'évaluation des élèves nécessite la mémorisation des faits multiples ou si la pédagogie ne forge pas les liens nécessaires pour construire ces notions-clés. Il ne sert à rien de demander aux élèves d'être responsables de leur propre apprentissage continu si l'enseignement ne leur permet pas d'avoir du temps pour la réflexion et la créativité. Nous ne pouvons pas espérer des attitudes positives envers la science si le contenu du programme paraît obsolète ou éloigné des intérêts et des expériences des élèves.

L'impact de l'évaluation des élèves sur le contenu du programme et l'enseignement est loin d'être forcément négatif. Un système efficace d'évaluation des élèves favorise l'apprentissage de différentes manières : en fournissant des commentaires formatifs utiles pour des décisions à court terme sur les activités d'apprentissage, et en apportant des informations aux parents sur la réussite des élèves dans le cadre d'une planification à long terme et de l'auto-évaluation de l'école. À travers l'établissement de critères de réussite ou de propositions de tâches qui illustrent l'utilisation de compétences et de compréhension, l'évaluation des élèves peut aider à clarifier et à communiquer la signification des objectifs d'apprentissage.

Les impacts négatifs surviennent lorsque le contenu évalué ne reflète que les aspects faciles à tester, auxquels s'ajoute l'application de récompenses et de punitions liées aux résultats de l'apprentissage. Lorsque les résultats des tests ont des impacts hautement importants pour les enseignants, la pression est transmise aux élèves, même si les tests ne constituent pas un enjeu majeur pour ces derniers. La recherche montre que, dans ce cas là, les enseignants se concentrent sur le contenu du test, forment leurs élèves à passer les tests et se sentent obligés d'adopter des styles d'enseignement qui ne correspondent pas à ce qui est nécessaire pour développer une véritable compréhension. Il y a maintenant un grand nombre de preuves sur l'impact négatif de l'utilisation intensive des données de l'évaluation des élèves et des tests. L'encadré 7 nous donne une brève indication des résultats d'une étude approfondie sur l'impact des tests auprès des enseignants et les élèves.<sup>42</sup>

Ces résultats soulèvent des questions sur l'équité, puisque les réponses négatives aux tests ne sont pas réparties uniformément sur l'ensemble des élèves. Les élèves peuvent être plus ou moins avantagés selon le sexe, la langue, le milieu familial et l'intelligence.

Menée entre 2007 et 2009, une étude sur l'enseignement primaire en Angleterre, conduite à grande échelle avec des sources variées, a conclu que les tests nationaux des élèves en fin de l'école primaire (11 ans) :

- mettent les enfants et les enseignants sous une grande pression ;
- limitent le programme scolaire ;
- subvertissent les objectifs d'apprentissage ;
- ébranlent l'estime de soi des enfants ;
- vont à l'encontre de l'engagement de l'école pour une éducation complète et équilibrée ;
- transforment la dernière année de l'école primaire en une année de bachotage et de tests.<sup>43</sup>

42 Nordenbo, S. E., Allerup, P., Andersen, H. L., Dolin, J., Korp, H., Larsen, M. S., et al. (2009). *Pædagogisk brug af test - Et systematisk review*. København: Aarhus Universitetsforlag. (In English: Pedagogical use of tests – A systematic review). <http://www.dpu.dk/omdpu/dansklclearinghouseforuddannelsesforskning/udgivelser/paedagogiskbrugaftest/>

43 Alexander, R. (Ed) (2010) *Children, their World, their Education*. Final report and recommendations of the Cambridge Primary Review. London: Routledge. P 316

*Encadré 7: Une recherche systématique portant sur l'impact de tests à grande échelle<sup>42</sup>*

Le Centre Danois pour la Recherche en Éducation a effectué une étude systématique portant sur les questions suivantes:

1. Comment les tests affectent-ils l'enseignement?
2. Comment les tests sont-ils utilisés dans la pédagogie par les enseignants?
3. Comment les tests affectent-ils les élèves?

L'étude s'est appuyée sur une recherche systématique, publiée entre 1980-2008. Les principales conclusions ont révélé des effets négatifs considérables de l'introduction de tests gérés au niveau central:

- Un programme réduit ou déformé expérimenté par des élèves: les enseignants simplifiaient les exigences sur la réflexion des élèves; les faits et les compétences mécaniques sont mis en valeur au détriment des activités créatives et esthétiques ;
- Plus de temps est consacré aux questions incluses dans les tests au détriment de celles qui n'y étaient pas ;
- Un enseignement de plus en plus orienté vers l'enseignement orienté vers les tests et l'apprentissage par cœur.

L'étude a également conclu que, de manière générale, les enseignants n'aiment pas utiliser les tests ni les données des tests à grande échelle dans leur enseignement. En effet, ils ont une attitude sceptique et négative face aux tests à moins qu'ils aient été impliqués dans leur conception. L'existence de tests à grande échelle peut rendre les enseignants plus réticents à faire usage des données de test. Cependant, si les enseignants ressentent un certain sentiment d'adhésion aux tests, ils seront plus disposés à en utiliser les résultats.

Les influences des tests sur les élèves peuvent être dramatiques:

- La simple annonce d'un test suscite des réactions émotionnelles comme la nervosité et la peur, surtout chez les filles ;
- Les élèves se préparent pour le test en apprenant par cœur et en mémorisant des phrases ;
- Ceux qui ont de bons résultats sont plus motivés mais ceux qui ont des résultats faibles perdent leur motivation ;
- Le résultat du test d'un élève peut influencer sa future motivation et son auto-efficacité.

Des effets similaires sont signalés dans d'autres situations où les résultats des tests sont utilisés pour porter des jugements sur les écoles et les enseignants. Toutefois, les recherches menées dans deux Länder d'Allemagne, en Hesse et à Bremen, suggèrent que le rétrécissement du programme et l'enseignement fondé sur les tests ne sont pas simplement dus à l'importance accordée aux résultats, mais par l'existence même des tests. L'étude de Jäger et al sur l'impact des tests à faible enjeu, conduits à l'échelle nationale en fin de l'enseignement secondaire, a révélé que les enseignants se servaient des commentaires de tests pour améliorer les résultats de leurs élèves tout en réduisant le programme. Les chercheurs suggèrent que « cela peut être considéré comme une indication que les enseignants se sentent responsables vis à vis de la performance de leurs élèves, de l'atteinte des objectifs et des normes du programme sans incitations extérieures ». <sup>44</sup>

44 Jager, J.J., Merki, K.M., Oerke, B. and Holmeier, M. (2012) State-wide low-stakes tests and a teaching to the test effect? An analysis of teacher survey data from two German States, *Assessment in Education*, 19 (4) 451-467, p 464

Les conséquences pour la science sont particulièrement graves. Nous constatons que les enseignants ont adopté un style de transmission de savoirs qui est loin d'être ce qu'ils croient être le meilleur pour favoriser la compréhension et le développement des compétences des élèves.<sup>45</sup>

Une autre conséquence sérieuse concerne la pratique de l'évaluation formative des élèves. À la suite de l'introduction du système de test en Angleterre, des chercheurs ont remarqué que les élèves et les enseignants étaient conscients que l'évaluation des élèves en classe, au lieu d'être essentiellement formative, était devenu une série de mini-évaluations sommatives.<sup>46</sup> Les enseignants vérifiaient le « niveau » de performance beaucoup plus souvent que ce qui était nécessaire pour établir une synthèse. Une « culture orientée vers la performance » est développée et laisse peu de place pour mettre en œuvre une évaluation formative des élèves. Cette « culture » a été rendue responsable du comportement d'écoles qui, volontairement, abusaient des évaluations externes des élèves et des jugements sur leur niveau.

L'influence qu'une évaluation sommative des élèves peut avoir sur leur évaluation formative montre clairement qu'il ne suffit pas de prêter uniquement attention à l'évaluation formative. Ceci est confirmé, dans des pays où existent des tests externes à forts enjeux, par l'observation des obstacles rencontrés lors de l'introduction d'une véritable évaluation formative. Ainsi, si l'apprentissage des sciences peut être amélioré grâce à l'ESFI et par l'utilisation d'une évaluation formative, il est également nécessaire de s'assurer que l'évaluation sommative utilisée est compatible avec les objectifs de l'ESFI.

### **Le mythe de relever le niveau par des tests**

Compte tenu de ces effets négatifs, on peut se poser la question suivante : pourquoi des tests nationaux ont-ils un rôle aussi important dans de nombreuses politiques de l'éducation? On dit que « les tests pilotent le niveau », mais peu de preuves soutiennent cette affirmation et beaucoup nous laissent penser que l'augmentation des scores aux tests est due à un entraînement répétitif en vue du test et à un enseignement fondé sur les tests.<sup>47</sup> Les résultats aux tests peuvent s'améliorer, au moins dans un premier temps, mais cela ne donne pas d'information sur les changements obtenus dans l'apprentissage proprement dit. Proposer des tests uniquement dans la gamme restreinte des questions testées fait qu'il est quasiment impossible de percevoir le changement annuel du niveau national à partir des résultats de tests nationaux !

Ce phénomène, où l'impact d'une mesure à grand enjeu est d'occulter sa valeur en tant qu'indicateur de qualité d'un service, est très fréquent dans certains services sociaux. Il est résumé succinctement dans la *loi de Goodhart*<sup>48</sup> :

Lorsqu'une mesure devient une cible, elle cesse d'être une bonne mesure.

Autrement dit, une fois qu'un indicateur social, économique, ou telle autre mesure s'y substituant, est considéré comme une cible dans la politique, il perd sa capacité de donner des informations valables par rapport aux objectifs de départ. Plusieurs exemples existent dans d'autres domaines que l'éducation (par exemple dans les services de santé ou les services de police), où un seul résultat facilement mesurable (tel que le temps d'attente pour un traitement dans un hôpital) est utilisé pour évaluer un processus complexe. Dans le cas de l'éducation, le test est souvent le moyen choisi pour fixer des objectifs. Les enjeux élevés mènent à des pratiques qui augmentent les résultats de tests sans une réelle amélioration de l'apprentissage, par conséquent le test cesse d'être une mesure utile de l'apprentissage. Bien que plus

45 Osborne, J., Simon, S. and Collins, S.(2003) Attitudes towards science: a review of the literature and its implications, *International Journal of Science Education*, 25, 1049-1079

46 Pollard, A and Triggs, P. (2000) *Policy, Practice and Pupil Experience*. London: Continuum International

47 See, for example: Linn, R. L. (2000) Assessments and accountability, *Educational Researcher*, 29 (2) 4-16 and Tymms, P. (2004) Are standards rising in English primary schools? *British Educational Research Journal*, 30 (4) 477-94

48 <http://www.atm.damtp.cam.ac.uk/mcintyre/papers/LHCE/goodhart.html>

souvent illustré en termes de test, le même effet peut être observé dans l'évaluation des élèves où la performance est jugée par rapport aux critères. L'utilisation excessive des résultats restreint l'interprétation des critères qui cessent alors d'indiquer l'ensemble des compétences attendues.

### La responsabilisation<sup>49</sup>

L'impact du fait de juger les enseignants et les écoles en fonction des résultats de tests est maintenant bien reconnu et conduit à affirmer que « la responsabilisation ruine l'apprentissage ». C'est pourtant une erreur, car ce n'est pas la responsabilisation qui pose problème, mais la façon dont elle est exercée. Lorsque les résultats des tests sont considérés comme une mesure de qualité de l'enseignement, ceci entraînera inévitablement des conséquences pour ceux qui sont jugés. Certaines de ces conséquences peuvent être considérées comme involontaires, par exemple lorsque les enseignants interprètent de façon trop stricte ce qu'ils considèrent comme étant important à tester, confondant les sujets testés et les sujets d'enseignements importants. Les enseignants concernés peuvent d'ailleurs affirmer qu'« ils n'ont pas besoin d'enseigner pour préparer aux tests » tout en acceptant la nocivité de certain impact des tests. Pourtant, la pratique continue. On se doute bien que le fait de rendre les enseignants responsables afin d'améliorer les scores est une utilisation délibérée de l'impact de l'évaluation des élèves sur le contenu des programmes. Les tests apportent un moyen direct de contrôler ce qui est enseigné. Comme le dit Stobart<sup>50</sup>, il s'agit d'un moyen, plus rapide et moins coûteux, de modifier le contenu du programme et de la pédagogie, plutôt que d'améliorer ces programmes et de financer le perfectionnement professionnel des professeurs. Ce moyen direct fonctionne en fixant des objectifs pour les écoles, les districts ou le pays entier pour améliorer les résultats des tests.

Au niveau du système scolaire entier, l'ambition de nombreux pays d'améliorer leur positionnement dans le « tableau de classement » de PISA les a conduit à des actions qui ont obligé les districts et les écoles à changer de pratiques.<sup>51</sup> Au niveau national, l'utilisation de tests peut sembler obtenir l'effet désiré pendant quelques années. Voici un exemple qu'illustrent les tests nationaux pour les élèves en fin de l'école primaire en Angleterre. Les tests en anglais, en mathématiques et en sciences pour les élèves de onze ans ont été introduits en 1995. Pour les cinq premières années, les scores ont augmenté d'année en année. Mais à partir de 2000, il n'y avait plus de changement notable. Une étude exhaustive<sup>52</sup> des autres données sur le changement dans la performance durant cette période (y compris les données des enquêtes TIMSS) n'a révélé aucune augmentation de performance. La conclusion montre que les explications les plus probables des changements initiaux fut le résultat de l'enseignement des techniques de test (les tests formels étaient nouveaux pour les élèves de cet âge en 1995) et la mise en place d'un enseignement strictement axé sur le contenu testé.

Des constats et conclusions similaires ont été signalés dans d'autres systèmes de test, par exemple aux USA, où Linn a trouvé un schéma typique de « premiers gains suivis d'une stabilisation »<sup>53</sup> dans les Etats ayant les tests à forts enjeux. Stobart décrit la stabilisation comme la « demi-vie » des tests mesurant la responsabilité, d'environ quatre ans :

Les tests ont été pris en main, les enseignants se sont familiarisés avec le test et les élèves savaient à partir des derniers devoirs ce qu'ils devaient faire, alors l'apprentissage est de plus en plus réduit à rechercher la maximisation des notes.<sup>54</sup>

49 Nous traduisons le terme anglais *accountability* par *responsabilisation* : le fait, pour un individu ou une organisation, d'être tenu pour capable de justifier ses actions ou décisions.

50 Stobart, G. (2008) op cit p138

51 Ertl, H.(2006) Educational standards and the changing discourse on education: The reception and consequences of the PISA study in Germany. *Oxford Review of Education*, 32 (5) pp 619–634

52 Tymms, P. (2004) Are standards rising in English primary schools? *British Educational Research Journal*, 30 (4): 477-494.

53 Linn, R.L. (2000) Assessment and accountability, *Educational Researcher*, 29 (2), 4-16, p6

54 Stobart, G. (2008) op cit p134

Au niveau de l'école, l'obligation de respecter les objectifs sous forme de scores signifie qu'une plus grande attention est accordée aux sujets testés. Pour les écoles primaires, la lecture, l'écriture et les mathématiques ont toujours été ciblées pour le test national, ce qui a créé de graves conséquences pour les sciences et d'autres sujets. En effet, les tests de sciences en fin de l'école primaire ont été supprimés en Angleterre en 2010 pour des raisons liées à la faible validité des tests passés par écrit sans aucune exigence d'expérimentation concrète<sup>55</sup>. Le temps et l'intérêt sont exclusivement consacrés à l'anglais et aux mathématiques, lesquels sont toujours l'objet de tests à enjeux élevés. Il est important d'éviter l'utilisation inappropriée des résultats des élèves, comme par exemple en les considérant comme le seul indicateur dans l'évaluation systémique des enseignants et des écoles. L'explication est pourtant simple : la réussite des élèves n'est pas seulement le résultat de leurs expériences passées à l'école.

### Les limites de la responsabilisation

Ici, la responsabilisation signifie le fait d'être responsable de ses actions et d'être en mesure d'expliquer aux intervenants pourquoi et comment certaines choses ont été faites ou pas. Les enseignants ne peuvent être tenus responsables que pour des actions ou des résultats qu'ils contrôlent directement, comme par exemple les possibilités d'apprentissage et l'aide qu'ils apportent aux élèves en classe. Ils ne sont pas nécessairement responsables des résultats prescrits en externe, car cela dépend de nombreux facteurs au delà de leur contrôle, tels que le niveau antérieur des élèves, les influences et les conditions extra-scolaires qui affectent leur apprentissage. Ces facteurs doivent être pris en compte à la fois par les enseignants dans leur planification pour atteindre les objectifs, et par ceux qui considèrent les enseignants comme étant responsables de la qualité de l'éducation des élèves. Il résulte de ces arguments que les informations utilisées dans la responsabilisation doivent inclure, outre des données sur les résultats des élèves, des informations sur le programme, les méthodes d'enseignement, les aspects pertinents du contexte des élèves et leur historique d'apprentissage.

#### *Encadré 8 : Les écoles rendent compte de leur activité*

Une approche positive de la responsabilisation permet aux écoles et aux enseignants d'exprimer leurs propres objectifs et réussites. Les écoles sont désormais censées communiquer et expliquer leur philosophie, objectifs et politiques à l'ensemble de la communauté. C'est l'occasion de présenter les informations sur la performance des élèves en tenant compte des nombreux facteurs influents. Le processus est facilité dans certains pays par des instructions,<sup>56</sup> adressées aux écoles pour les aider à formuler la façon dont elles rendent compte. Ces instructions sont conçues pour favoriser l'auto-évaluation des écoles et jouent un rôle plus formatif dans l'ensemble du processus. Dans certains cas, ce processus est vérifié par les inspecteurs ; dans d'autres cas, il est disponible pour un examen public.

Les explications de l'école sont constituées des données sur la performance des élèves notée dans les rapports des enseignants et des évaluations des élèves relatives à un ensemble d'objectifs d'apprentissage, sans être limité aux scores de tests, pour montrer le résultat aux parents et à tous ceux qui souhaitent la connaître. Les explications doivent également indiquer, dans le cadre de l'évaluation systémique interne de l'école, comment des données agrégées sur la performance des élèves sont utilisées pour vérifier les progrès des groupes d'élèves : comme ceux qui ont des langues maternelles différentes, des niveaux plus ou moins faibles, et des sexes différents. Tout cela dépend des mesures efficaces reliées aux objectifs d'apprentissage, en mettant les résultats dans le contexte des actions et des systèmes mis en œuvre pour la réalisation de ces objectifs.

55 En France, ces tests de science n'existent pas à l'école primaire.

56 SEED (Scottish Executive Education Department) (2002) *How Good is Our School? Self evaluation using quality indicators*. Edinburgh: HMIE.

## L'évaluation des élèves et les théories de l'apprentissage

L'alignement de l'évaluation des élèves avec le contenu du programme et la pédagogie est important pour justifier la pertinence de l'évaluation des élèves. La discussion du chapitre 1 montre que la validité est un jugement dont les conclusions sont soutenues par des preuves et des justifications théoriques. Un point fréquemment exprimé à la Conférence IAP d'Helsinki (voir l'Introduction) concerne la nécessité d'une évaluation des élèves qui reflète la théorie de l'apprentissage de l'ESFI. Nous devons donc apporter un regard théorique en discutant l'impact de l'évaluation des élèves sur le contenu du programme et la pédagogie.

### Les théories de l'apprentissage

Les théories d'apprentissage diverses et variées peuvent être regroupées en trois catégories : le « comportementalisme », le « constructivisme cognitif » et le « socioconstructivisme » (Dans certains ouvrages américains<sup>57</sup>, « constructiviste » est décrit comme « cognitive » et « socio » comme « situé », mais les significations restent essentiellement les mêmes.) Une simple formulation de ces termes par Watkins<sup>58</sup> exprime leur signification :

- Comportementalisme : « l'apprentissage est enseigné »
- Constructivisme cognitif : « l'apprentissage est la compréhension individuelle »
- Socioconstructivisme : « l'apprentissage est l'acquisition des connaissances en faisant intervenir des interactions avec d'autres personnes. »

*Le comportementalisme* décrit une vision de l'apprentissage dans lequel les comportements sont construits par un système de récompenses et de punitions. Selon l'hypothèse, les comportements récompensés seront renforcés et ceux qui sont punis disparaîtront; l'apprentissage peut alors être contrôlé de l'extérieur et la motivation est presque entièrement extrinsèque. Une autre caractéristique du comportementalisme, particulièrement pertinente pour l'évaluation des élèves, est que les comportements complexes sont décomposés en plusieurs parties qui peuvent être enseignées, pratiquées et évaluées séparément. Ce point de vue est donc compatible avec les tests sur des idées et des compétences déconnectées, où le temps est bref et où les réponses sont soit bonnes, soit mauvaises.

Deux principales caractéristiques de la vision du *constructivisme cognitif* sur l'apprentissage affirment, d'une part, que les apprenants construisent leur propre compréhension en développant des modèles mentaux, et d'autre part, que les connaissances existantes ont un rôle important dans ce développement. L'objectif consiste à comprendre ce qui arrive lorsque une nouvelle expérience dans un modèle existant ou nouveau. La participation active des élèves est primordiale, car « ce sont eux qui apprennent ». Les idées constructivistes de l'apprentissage renforcent le rôle d'une évaluation formative des élèves : d'abord « situer le positionnement des apprenants dans leur apprentissage » afin de décider « où ils doivent aller et comment y arriver » (voir page 21). Peu d'exemples d'évaluation sommative des élèves sont basés sur la vision constructiviste, il existe toutefois quelques tentatives à travers des tests informatisés et la schématisation conceptuelle sur écran.<sup>59</sup> James conclut que « beaucoup de tests formels s'appuient encore fortement sur les approches comportementalistes ». <sup>60</sup>

57 Discussed in Pellegrino, J.W., Chudowsky, N. and Glaser, R. (Eds) (2001) *Knowing what Students Know - The Science and Design and Educational Assessment*. Washington, DC: National Academy Press.

58 Watkins, C. (2003) *Learning: A Sense-Maker's Guide*. London: Association of Teachers and Lecturers.

59 Osmundson, E., Chung, G., Herl, H., Klein D. (1999) Knowledge-mapping in the classroom: a tool for examining the development of students' conceptual understandings. Los Angeles, California: National Centre for Research on Evaluation and Student Testing, University of California. [www.cse.ucla.edu/Reports/TECH507.pdf](http://www.cse.ucla.edu/Reports/TECH507.pdf)

60 James, M. (2012) Assessment in harmony with our understanding of learning: problems and possibilities, in Ed J.Gardner *Assessment and Learning*, 2nd edn. London: Sage pp187 – 205.

Les perspectives *socioconstructivistes* de l'apprentissage accentuent l'idée de la compréhension, mais en donnant « un sens à de nouvelles expériences avec les autres » plutôt que par un travail individuel. Dans ces situations, l'individu prend (internalise) d'abord dans une expérience partagée ce qui est nécessaire pour faciliter sa compréhension, puis extériorise le résultat sous forme de contribution à la discussion de groupe. Il y a un constant va-et-vient d'un individu à un groupe puisque le savoir se construit collectivement par l'interaction sociale et le dialogue. Les ressources physiques et linguistiques ont des rôles importants, comme l'indique James:

Dans cette optique, l'apprentissage se fait dans les interactions entre l'individu et l'environnement social. La pensée est menée à travers des actions qui modifient la situation et la situation qui change la pensée : elles sont constamment en interaction. Il est particulièrement important de noter que l'apprentissage est une activité médiée dans laquelle des objets culturels jouent un rôle crucial. Ceux-ci peuvent être soit des objets physiques tels que des livres et du matériel, soit des outils symboliques comme le langage. Car le langage, étant au cœur de notre capacité de penser, s'est développé à travers les rapports entre les individus et les relations sociales qui précèdent l'apprentissage (Vygotsky, 1978). Ainsi, l'apprentissage est une activité sociale et collaborative dans lequel les individus développent ensemble leur pensée.<sup>61</sup>

Certaines conséquences profondes sur l'évaluation des élèves découlent de l'opinion de Vygotsky<sup>62</sup> : pour tout apprenant, il existe une zone située juste au-delà de la compréhension actuelle (lorsqu'une personne contrôle les idées consciemment et sait qu'il/elle les utilise) où les idées plus avancées peuvent être utilisées avec de l'aide. Vygotsky appelle cette région la « zone proximale du développement ». C'est en substance ce que nous avons appelé l'« étape suivante » que l'élève peut identifier lorsqu'est pratiquée son évaluation formative. « L'échafaudage » (*scaffolding*) est un terme approprié pour décrire le processus qui veut aider les élèves à franchir cette nouvelle étape dans la compréhension, en introduisant de nouvelles idées ou de meilleures pratiques scientifiques avec un vocabulaire adapté, ce qui leur permet d'exprimer leurs idées avec plus de précision. Il admet ainsi que, avec la compagnie des autres apprenants, l'élève peut dépasser ce qu'il peut comprendre et faire tout seul en remettant en question son « vrai » niveau de performance. S'agit-il du niveau de « performance indépendante » ou de « performance assistée » dans le contexte social ? On a pu ainsi faire valoir que le niveau de performance, avec l'assistance et les nouveaux outils fournis par les autres individus, donne une meilleure évaluation des élèves que l'administration des tests de performance sans assistance.<sup>63</sup>

### Les implications pour l'évaluation des élèves dans l'ESFI

Dans l'encadré 9, l'ESFI est décrit comme un processus de construction de la compréhension, à travers la collecte de preuves pour tester les explications et les idées possibles, constituant ainsi une approche d'apprentissage à la fois constructiviste et socio-culturelle. Si la validité dépend de la façon dont l'évaluation des élèves reflète l'approche de l'ESFI, comment une évaluation sommative d'un enseignement des sciences fondé sur l'investigation doit-elle être réalisée? Certainement pas avec des élèves assis séparément dans une salle d'examen. L'évaluation des élèves, conforme à l'idée d'un apprentissage soutenu par l'ESFI, doit se fonder sur ce que les élèves peuvent faire en interaction avec les autres. Lorsque les activités en classe ou en laboratoire impliquent la construction d'un échafaudage de la part des enseignants et des interactions entre les élèves, ces activités offrent

61 James, M. (2012) op cit p192/3

62 Vygotski L Pensée et langage (1933). La dispute, Paris 1997

63 Grigorenko, E. (1998) Mastering tools of the mind in school (trying out Vygotsky's ideas in classrooms), in Eds R. Sternberg and W. Williams *Intelligence, Instruction and Assessment: Theory and Practice*. Mahwah, NJ: Erlbaum.

*Encadré 9: Un exemple d'apprentissage des sciences par l'investigation*

Dans l'apprentissage fondé sur l'investigation, le développement de la compréhension émerge de la curiosité d'un phénomène ou d'un événement nouveau pour celui qui apprend et soulève des questions qui retiennent son attention. L'exploration initiale peut révéler des caractéristiques, rappelant une idée venant de l'expérience antérieure, qui suggèrent une explication possible ou une réponse à une question. Cela peut être l'idée d'un élève, le résultat d'un *brainstorming* avec les autres élèves ou d'autres sources d'information. Travailler de façon scientifique consiste à faire une prédiction basée sur une idée, à collecter des données pertinentes pour trouver des preuves à l'appui de la prédiction et à l'application de l'idée. L'investigation pourrait donc être longue en impliquant des expériences contrôlées ou simplement des observations prolongées.

Lorsque les preuves correspondent à la prédiction et que cette idée de départ fournit une bonne explication, cette idée devient « plus ample » car elle explique un ensemble de phénomènes. Même si cela ne semble pas « fonctionner » parfaitement, quelque chose a quand même été appris dans le champ d'application. Mais pour trouver une explication qui « fonctionne », des idées alternatives doivent être utilisées et testées. Cela peut provenir du *brainstorming* initial ou ultérieur, utilisant des résultats trouvés. L'utilité des idées développées ainsi dépend de la collecte et de l'utilisation des preuves avec la méthode scientifique. Par conséquent, la capacité d'utiliser ses compétences d'investigation scientifique est une partie essentielle du développement de la compréhension, et le résultat d'une réflexion partagée sur le choix des données et la manière de les collecter et interpréter.

également la possibilité d'évaluer la « performance assistée » des élèves en donnant un argument supplémentaire aux enseignants qui jouent un rôle central dans l'évaluation des élèves.

Un exemple d'évaluation des élèves, informée par l'interaction sociale et l'utilisation des objets, est illustré par Dolin et Krogh dans leur étude qui réexamine les réponses des élèves à certaines questions de PISA (voir page 70). Un entretien est mené avec chaque élève pendant environ 30 minutes pour explorer leurs connaissances et leur raisonnement. Ces entretiens individuels n'ont clairement qu'un champ d'application limité lorsqu'un grand nombre d'élèves doit être évalué. Cependant, ceci peut présenter un problème de fiabilité. Une situation plus contrôlée serait de donner aux groupes d'élèves une tâche et du temps pour faire un remue-méninges (*brainstorming*) sur la façon de s'y prendre et, si possible, mener une investigation. Chaque élève ferait ensuite un travail individuel qui sera évalué à l'aide d'une série de critères.

Une approche différente pourrait consister dans une « interaction » à distance. Les élèves mènent une investigation et reçoivent en retour des commentaires à ce sujet. Les réponses des élèves aux commentaires contribuent à l'évaluation globale de ces élèves. Ce type d'approche, qui n'a encore que peu d'exemples dans la pratique, défie les idées conventionnelles du processus et du sens de l'évaluation des élèves. Comme le mentionne James : « Il y a encore beaucoup de travail à faire pour trouver des moyens de mieux aligner l'évaluation des élèves avec quelques unes des idées les plus puissantes de la théorie moderne de l'apprentissage ». <sup>64</sup> La question des éléments déclencheurs pour un tel changement est abordée au chapitre 7.

---

64 James, M. (2012) op cit p196



## Le rôle de l'évaluation des élèves dans la saisie des grands problèmes mondiaux

Affirmer qu'une évaluation adéquate des élèves peut aider à développer leur compréhension des grands problèmes mondiaux tels que le réchauffement climatique, la perte de la biodiversité, la famine et la pauvreté causés par l'activité humaine pourrait paraître surestimer le rôle de cette évaluation. Pourtant, l'éducation joue certainement un rôle essentiel pour permettre des changements face à ces problèmes. Par conséquent, l'évaluation des élèves, ayant une implication décisive dans l'éducation, doit aussi avoir son rôle à jouer.

Un objectif crucial de l'éducation est de préparer les élèves à devenir des citoyens informés qui comprennent la raison des problèmes créés par l'activité humaine et les éléments nécessaires pour trouver les solutions tout en étant motivés à participer à des actions responsables. De nombreux aspects sont à examiner, mais voici deux questions importantes :

- Quelles sont les idées ou les concepts essentiels qui doivent être compris?
- Comment engendrer la compréhension (par opposition à une connaissance superficielle des faits) de ces concepts essentiels?

Pour répondre à la première question, nous devons identifier les idées pertinentes et puissantes qui peuvent faciliter la compréhension du monde et de son fonctionnement, ainsi que leurs interactions? Comment l'intervention humaine peut ou ne peut pas influencer notre environnement mondial? Il s'agit d'identifier les notions-clés *des sciences* et *sur les sciences* (c'est-à-dire, le fonctionnement de la science, ses forces et ses limites) et de veiller à ce que l'enseignement des sciences soit conçu pour développer la compréhension de ces notions-clés. La meilleure réponse que nous avons à la seconde question, portant sur la manière de mener les élèves à une véritable compréhension, est de proposer aux élèves un apprentissage par l'investigation. Nous avons décrit, au chapitre 2 et dans l'encadré 9, la définition d'un tel apprentissage, en affirmant qu'il permet un apprentissage actif (physiquement et mentalement) à travers lequel ceux qui apprennent comprennent le monde qui les entoure grâce à l'investigation. Cela ne signifie pas qu'ils doivent apprendre ou tout faire par eux-mêmes. Au contraire, comme indiqué plus haut, les opinions actuelles sur le processus du développement des idées reposent sur le rôle de la collaboration, de la discussion et du dialogue avec les autres. Les idées progressent par la réflexion collective en s'appuyant sur ce qui a déjà été trouvé par les scientifiques du passé. Cela signifie que les apprenants développent une compréhension partagée, qui devient plus significative pour eux que les autres visions alternatives qu'eux-mêmes (ou d'autres) aient eu auparavant.

Alors pourquoi l'évaluation des élèves entre-t-elle dans cet argument? En bref, si le principe selon lequel « toute évaluation des élèves doit contribuer à l'apprentissage » (voir l'encadré 21 au chapitre 7) est traduit dans la pratique, son rôle devient alors clair. Plus précisément, l'utilisation de l'évaluation formative des élèves régit l'enseignement et l'apprentissage afin que la compréhension soit bien soutenue. Elle agit pour optimiser le défi des nouvelles expériences des élèves afin qu'elles ne soient ni trop éloignées de leurs idées et compétences existantes, ni trop familières pour qu'il y ait un peu de changement dans ce qu'ils peuvent faire ou savoir. En ce qui concerne l'évaluation sommative des élèves, elle peut être utilisée pour apporter un soutien à l'apprentissage par plusieurs moyens. Par exemple, comme mentionné au chapitre 3, la cartographie de la progression d'élèves individuels et de groupes d'élèves implique la nécessité d'agir pour garantir un apprentissage pour tous, pour les filles et les garçons, pour le plus capable et le moins, pour les personnes défavorisées ainsi que les plus fortunées. Le contenu de la carte doit montrer que l'objectif est que tous les élèves fassent des progrès vers les objectifs de l'ESFI : des idées solides et des compétences d'investigation scientifique. (Plus tard, dans le chapitre 6, nous mentionnerons quelques implications dans la manière de spécifier les résultats des élèves). Le progrès des élèves envers ces objectifs est un facteur clé dans l'auto-évaluation des écoles quant à leur mise en pratique d'un apprentissage par la compréhension.

L'évaluation des élèves n'est qu'un des facteurs divers qui influent sur l'apprentissage des élèves. Clairement, ce que les élèves peuvent atteindre dépend du contenu du programme et de la pédagogie, mais les interactions indiquées dans la figure 3 et la discussion au début de ce chapitre montrent le rôle non-négligeable de l'évaluation des élèves. Celle-ci doit en outre contribuer à apporter une meilleure compréhension des objectifs en explicitant pourquoi ils doivent être atteints. L'évaluation des élèves doit donc être incluse dans toute discussion sur la meilleure façon de préparer les élèves aux défis mondiaux.



## Chapitre 5

# La mise en œuvre d'une évaluation formative des élèves pratiquant l'ESFI

Dans le chapitre 3, nous avons examiné, d'une manière générale, la signification et l'importance de l'évaluation des élèves *pour aider à l'apprentissage* (formative) et pour *juger de l'apprentissage* (sommative). Dans ce chapitre et le chapitre suivant, nous regardons les méthodes de la mise en place d'une évaluation des élèves à ces fins, plus particulièrement dans le contexte des objectifs de l'ESFI. Dans le chapitre 2, six objectifs fondamentaux de l'enseignement des sciences ont été identifiés pour préparer les élèves à vivre dans un monde en pleine transformation, tandis que la définition de l'ESFI de la page 14 met en lumière sa contribution spécifique au développement de *la compréhension* et des *compétences utilisées par les scientifiques* (ce que nous appelons *les compétences en investigation scientifique*). Ainsi, l'objet principal de ce chapitre porte sur le rôle de l'évaluation des élèves lorsqu'il s'agit de promouvoir le développement des deux objectifs suivants : les connaissances scientifiques et les compétences en investigation scientifique. Le chapitre 6 examinera les méthodes de l'évaluation sommative des élèves et quels progrès sont acquis dans ce but.

### L'évaluation formative des élèves et l'ESFI

La mise en œuvre de l'évaluation formative des élèves n'est pas une tâche facile. Sans requérir des documents ni des techniques particulières, elle nécessite cependant des compétences et des connaissances de la part des enseignants. L'objectif de l'évaluation formative des élèves est d'aider les élèves à franchir une nouvelle étape dans le développement de leur compréhension et de leurs compétences. Pour mettre en œuvre l'évaluation formative des élèves, les enseignants ont besoin, d'une part de connaissances sur la manière de recueillir et d'utiliser des preuves qui indiquent les progrès des élèves pendant l'apprentissage, d'autre part, d'un savoir sur la façon de créer des environnements d'apprentissage efficaces permettant d'obtenir de nouveaux progrès. Certains chercheurs, d'après Sadler<sup>65</sup>, décrivent ceci comme des actions permettant à l'élève de « réduire l'écart » entre l'état actuel de ses connaissances et compétences et l'objectif de l'apprentissage. Toutefois, l'appellation d'« étapes suivantes », plutôt que celle de « réduction de l'écart », exprime mieux une vision du progrès dans l'apprentissage comme étant un processus continu. De plus, l'évaluation formative des élèves exige surtout un changement de la part des enseignants en leur donnant un nouveau rôle à jouer dans le processus d'apprentissage. Lorsque l'apprentissage est vu comme quelque chose que *les élèves font*, et non pas comme quelque chose qu'*ils subissent*, le rôle de l'enseignant est de concevoir des environnements dans lesquels les élèves peuvent être activement engagés dans la construction de leur compréhension et le développement de compétences.

Il est évident que l'évaluation formative des élèves est essentielle à la mise en œuvre de l'ESFI. L'apprentissage par l'investigation est un processus de développement des connaissances qui tient compte de la meilleure façon d'apprendre pour les élèves, c'est-à-dire par leur activité physique et mentale. Elle est fondée sur le fait que les élèves construisent les idées, les connaissances et la

65 Sadler, D. R. (1989) Formative assessment and the design of instructional systems, *Instructional Science*, 18, 119-44.

compréhension avec leur propre réflexion et expériences. Les études actuelles sur l'apprentissage nous dévoilent que ceci est réalisé lorsque les activités des élèves leur permettent de développer la compréhension, c'est-à-dire, lorsqu'ils travaillent dans la zone située entre les idées existantes et les notions ou les compétences plus avancées, ou dans la zone proximale de développement (voir page 38). L'évaluation formative des élèves est la conception stratégique de ces activités et des environnements d'apprentissage en vue de garantir la possibilité de progrès dans l'apprentissage. Lorsque les élèves apprennent par l'investigation, ils développent leur compréhension à travers les activités indiquées dans l'encadré 3 et dans la définition de l'ESFI : faire des observations, soulever des questions de recherche, planifier et mener des investigations, examiner des preuves à la lumière de ce qui est déjà connu, tirer des conclusions et communiquer, discuter avec les autres pour partager les idées, expliquer et défendre son point de vue. Il est également reconnu que la participation de ces activités dépend du degré d'intérêt, de la pertinence et du plaisir voire de l'excitation chez les élèves.

Pour l'enseignant, l'évaluation formative des élèves de l'ESFI signifie l'utilisation de stratégies identifiées dans la description du chapitre 3 :

- promouvoir le dialogue en classe ;
- utiliser les questions pour générer les preuves et faciliter le développement des idées et des compétences des élèves ;
- fournir un feedback aux élèves ;
- utiliser les commentaires des élèves pour ajuster l'enseignement ,
- encourager les élèves à participer à l'évaluation de la qualité de leur propre travail.

Nous examinons d'abord les différentes manières de mettre en pratique les aspects de l'évaluation formative des élèves par les enseignants. Puis, dans le chapitre 7, nous reviendrons sur comment amener les changements nécessaires dans l'enseignement en vue d'implanter ces pratiques.

## Le dialogue en classe

C'est par le langage que nous développons une compréhension partagée des idées. Les idées formées à partir de l'expérience directe doivent être communiquées en essayant de trouver les mots qui transmettent aux autres le sens que nous leur donnons. Dans ce processus, nos propres idées doivent souvent être reformulées pour tenir compte du sens des mots perçus par les autres. La valeur de la parole lors de l'apprentissage a été établie dès l'Antiquité, mais à l'époque moderne, c'est l'étude de Douglas Barnes, lors des années 1970, qui a attiré l'attention sur l'importance des discussions informelles ou « exploratoires »<sup>66</sup>. Dans ce type de discussion, les élèves s'interrompent, se répètent, hésitent et reformulent leurs phrases. Selon Barnes, les élèves ne s'engagent dans ce type de discussion qu'en l'absence de l'enseignant parce qu'il n'y a aucune source d'autorité à qui les élèves puissent s'adresser. Cependant, il a aussi montré que le rôle de l'enseignant consistait à s'assurer de l'approfondissement de la pensée, à remettre en question les imprécisions, et à s'assurer que les affirmations des élèves soient réfléchies. L'enseignant doit donc s'engager sur une voie délicate située entre la domination de la discussion, qui risque de décourager les élèves à développer et exprimer leurs propres idées, et l'abstention qui laisse les élèves seuls, sans aucune stimulation de leurs idées.

L'extrait de l'encadré 10 est une transcription de la discussion d'un groupe de trois filles observées par Barnes avant et pendant l'intervention de l'enseignant. Barnes souligne qu'en posant des questions telles que « Quelle pression de l'air ? », « Pourquoi cela ne s'est pas produit plus tôt ? », « À quel moment s'est-t-il arrêté ? », l'enseignant incite ces jeunes filles à réfléchir de manière systématique sur

66 Barnes, D. (1976) *From Communication to Curriculum*. Harmondsworth: Penguin.

le processus, avec l'aide de questions qu'elles-mêmes n'avaient pas posées. Cependant, les échanges initiés par l'enseignant illustrent un fait commun dans le discours en classe dans lequel :

- l'enseignant pose une question (*quelle pression de l'air ?*)
- l'élève répond (*à l'intérieur de la bouteille*)
- l'enseignant fait un commentaire évaluatif (*C'est bien...*)
- puis pose une autre question (*pourquoi cela ne s'est pas produit plus tôt?*).

#### Encadré 10 : Expliquer l'effet de la pression de l'air

Après quelques leçons sur l'air, les élèves (âgés de 12 à 13 ans) ont été invités à exercer des activités simples sur la pression de l'air. Parmi les équipements, il y a un flacon contenant de l'eau. Il est fermé par un bouchon traversé par une paille qui va à peine en dessous de la surface de l'eau du flacon. Les élèves (T, B et C) ont été invités à souffler fort dans la paille puis à réfléchir : ils devaient discuter et essayer d'expliquer ce qui s'est passé.

Discussion en absence de l'enseignant

- T Ah! L'eau monte dans la paille.  
 C Je me demande comment... Je me demande pourquoi... Comment tu l'arrêtes ?  
 ? ...ça s'arrêtera

*Une autre fille souffle dans la paille*

- B Pr...probablement c'est à cause de la pression de l'air  
 T Oui...ça doit être la pression de l'air et...  
 C ... c'est la pression de l'air  
 C ça met la pression sur l'eau.

Discussion avec l'intervention de l'enseignant

- B (*en réponse de la question de l'enseignant sur ce qui s'est passé*)  
 T a soufflé dedans et ça a fait des bulles, après elle a enlevé sa bouche et tout a remonté à cause de la pression de l'air  
 Enseignant Quelle pression de l'air ?  
 B le ... euh ... l'intérieur de la bouteille  
 Enseignant C'est bien. Maintenant pourquoi cela ne s'est pas produit plus tôt ?  
 C et T parce qu'il n'y avait pas assez d'air... de pression de l'air  
 Enseignant ... il n'y avait pas assez d'air à l'intérieur, mais quand tu as soufflé, il y avait plus d'air et cela a forcé... Pourquoi s'est-il arrêté ? Il ne se passe rien maintenant... Pourquoi il ne se lance pas maintenant ?

Ce discours implique beaucoup d'interactions entre l'enseignant et les élèves, mais dans ce cas de figure, l'expression de l'autorité de l'enseignant est indubitable. Comparons cela avec le rôle de l'enseignant dans l'encadré 11, où il encourage également l'approfondissement de la réflexion, l'utilisation des preuves et la clarté de la signification, mais sous forme d'échanges partagés plutôt que dominés. L'enseignant n'évalue pas les réponses des élèves, mais les incite à les exprimer et les expliquer.

Alexander a identifié le rôle de l'enseignant dans ces interactions verbales sous forme de « l'enseignement axé sur le dialogue ». Il le décrit comme « une approche pédagogique distincte » qui « exploite le pouvoir de parler afin de stimuler, développer la réflexion des enfants et avancer dans leur apprentissage et leur compréhension. Elle permet également à l'enseignant de diagnostiquer et d'évaluer avec une plus grande précision. »<sup>67</sup> C'est à travers un enseignement axé sur le dialogue que les enseignants peuvent « orienter les discussions en classe en gardant à l'esprit les objectifs éducatifs spécifiques ». <sup>68</sup> En science, cette « orientation » met l'accent sur l'utilisation des preuves et peut conduire à ce qui a été décrit comme « l'argumentation », qui est tout à fait différent de l'argumentation dans la vie quotidienne :

En science, les objectifs de l'argumentation consistent à promouvoir autant que possible la compréhension d'une situation et à convaincre ses collègues de la validité d'une idée spécifique. Plutôt que d'essayer de gagner un débat comme les gens le font souvent dans des contextes non scientifiques, l'argumentation scientifique consiste idéalement en un partage, un traitement et un apprentissage des idées.<sup>69</sup>

L'encadré 11 illustre un exemple de dialogue entre un enseignant et deux filles (de dix ans) qui essaient d'utiliser les preuves pour distinguer les œufs crus des œufs cuits. Après avoir désigné C comme l'œuf cru, D et M sont en désaccord sur l'état des deux autres œufs. M pense que B est l'œuf dur au motif que « les mauvais œufs flottent à la surface » et elle considère que l'œuf B se comporte comme si on lui avait fait quelque chose. Cependant elle ne précise pas les conséquences jusqu'à ce que D tente de lui donner raison. Alors c'est comme si c'était la réponse de D, qu'elle interrompt, qui déclenche sa propre réflexion.

L'enseignant n'a pas beaucoup à faire, à part encourager les filles dans leur tentative de mettre au point leur réponse et leur raisonnement. Il suffit de quelques questions occasionnelles comme « Pourquoi tu penses comme ça ? », de l'expression de l'approbation « je vois », et du renforcement « un peu cru », pour encourager les élèves à utiliser des données à l'appui de leurs arguments. Cela se voit clairement lorsque M a dit : « si on travaille sur le principe de ... ». Elle met en relation sa prédiction basée sur son jugement à l'observation que l'œuf flotte très vite dans l'eau salée. Il est intéressant de noter au passage que l'origine de son idée est basée sur sa connaissance sur la façon de distinguer les « bons » et les « mauvais » œufs.

Dans l'évaluation formative des élèves, où l'objectif est de montrer l'évolution de la compréhension d'un phénomène ou d'un événement des élèves, le discours des élèves constitue une source essentielle d'information. Le type de réflexion et de dialogue exploratoire préconisé par Barnes et Alexander est encouragé dans une classe où les enseignants :

- s'attendent à ce que les élèves proposent des explications ;
- apprécient les idées de leurs élèves, même si elles sont « informes » et hautement conjecturales ;
- évitent de donner l'impression que seule la « bonne » réponse est acceptable et encouragent les élèves à la deviner ;
- jugent le moment d'intervenir ou au contraire de laisser la discussion entre les élèves se poursuivre sans interruption.

67 Alexander, R. (2004) *Towards Dialogic Teaching. Rethinking Classroom Talk*. Cambridge: Dialogos, p1

68 *ibid*: p 27

69 Michaels, S., Shouse, A.W. and Schweingruber, H.A (2008) *Ready, Set, Science! Putting research to work in K-8 Science Classrooms*, Washington: National Academies Press, p89.

*Encadré 11 : Dialogue à propos des œufs* <sup>70</sup>

Deux filles de dix ans, D et M, conduisaient une investigation sur trois œufs entiers de poules, appelés A, B et C, et leur manière de se comporter dans l'eau du robinet et l'eau salée. Elles savaient qu'il y avait parmi eux, un œuf dur, un œuf mollet et un œuf cru. Leur mission était de les retrouver.

Lorsque les œufs ont été placés dans de l'eau salée, l'œuf B a flotté à la surface, l'œuf C est resté en bas et l'œuf A était resté suspendu au milieu. La transcription commence avec l'intervention de l'enseignant après les avoir laissé travailler seules pendant un moment.

- Enseignant Pouvez-vous me dire comment vous avez trouvé?
- D Je pense que C est cru
- M On pense toutes les deux que C est cru
- Enseignant Vraiment ?
- D B est ...
- Enseignant (*à D*) Pourquoi tu penses comme ça ?
- M parce que quand on met des œufs dans l'eau, les mauvais flottent à la surface
- D (*en même temps*) Parce que... on les met tous dans ...
- Enseignant Mauvais ?
- M Oui, je pense, ou c'est peut-être ceux qui sont bons ? ... Bon, je ne sais pas
- Enseignant Oui ?
- M ... ils remontent à la surface, donc...  
(*D met les œufs dans l'eau salée*)
- D ... c'est au fond (*indiquant C*)
- M ... s'il est cru, il devrait rester au fond
- Enseignant Je vois
- D C'est ce qu'on pense, C est cru, B est moyen et A est l'œuf dur
- M ... moi je pense que B est l'œuf dur et elle pense que B est l'œuf moyen
- Enseignant Ah, je vois (*à D*) Est-ce que tu peux expliquer, alors, pourquoi tu le penses ?
- D Si on met...euh...sort C (*elle sort C, le met sur la table et remonte ensuite A et B*) et les remet dedans un par un. Mettre d'abord A, non B en premier. C'est ce que... Allyson pense que c'est un œuf dur, moi je pense qu'il est mollet. Si tu le mets dans...(elle met B dans l'eau salée)
- M ... parce qu'il remonte plus vite
- D Il remonte plus vite. Et si tu le mets dans...  
(*Elle met A dans l'eau salée. Il va au fond et remonte très lentement*)
- M et celui là remonte plus lentement
- D Donc je pense que celui-ci (*montrant A*) est un œuf dur parce qu'il est...et bien...
- M Je ne pense pas. Je pense que si on travaille sur le principe de celui là (*montrant B*). Celui-ci remonte plus vite parce qu'il est, tu sais, pas vraiment cuit. C'est un peu cru quoi.
- Enseignant Un peu cru
- M donc, c'est pour ça qu'il va remonter plus vite
- D Oui mais il n'est pas mauvais
- Enseignant Comment ça va être à l'intérieur ?
- M liquide
- Enseignant ça va être liquide encore, je vois

70 Quoted from Harlen, W. and Qualter, A. (2009) *The Teaching of Science in Primary Schools*. 5<sup>th</sup> edn London: Routledge, p 100-101



## Les questions des enseignants

Les questions jouent un rôle central dans la parole au sein de la classe : questions posées par l'enseignant et questions posées par les élèves, entre eux ou à l'enseignant. Le questionnement prend une grande place dans le propos des enseignants et constitue l'un des facteurs les plus importants pour offrir aux élèves l'occasion de développer leur compréhension grâce à l'investigation. Ce n'est pas la fréquence des questions qui importe, mais leur forme, leur contenu et la manière dont elles sont disposées dans les prises de parole en classe.

En ce qui concerne la *forme*, les distinctions les plus importantes se situent entre des questions « ouvertes » et « fermées », des questions « centrées sur le sujet » et « centrées sur la personne ». Les questions ouvertes permettent aux élèves d'exprimer leur point de vue ou leur observation (« qu'est-ce que tu remarques à propos de ...? »), plutôt que de répondre à un point particulier soulevé par l'enseignant (« ... sont-ils tous de la même taille ? »). Les questions centrées sur le sujet interrogent directement sur le sujet (« pourquoi ceci prend plus de temps que ...? ») ; tandis que les questions centrées sur la personne interrogent sur les idées de l'élève (« pourquoi pensez-vous que ... prend plus de temps que ...? »). Les questions ouvertes et centrées sur la personne sont plus susceptibles de donner des informations aux enseignants sur les observations et les pensées de l'élève ; ceci peut fournir des éléments importants pour trouver le bon moyen de l'aider.

En ce qui concerne *le fond*, les questions doivent être adaptées à l'objectif de la demande. Il ne faut pas poser de question sans raison surtout si la réponse ne présente aucun intérêt. Si la réponse est utile, alors elle doit donner les informations ou stimuler les réponses attendues. Les questions sont utilisées dans toutes les parties du cycle d'évaluation formative des élèves (figure 1 page 21), mais dans le contexte de l'ESFI, elles sont surtout utilisées pour aider les élèves à percevoir les étapes à venir dans le développement de leurs idées et compétences, les encourager à la collaboration, au partage des idées et à la réflexion. Le contenu sera différent selon l'objet étudié à un moment donné. Par exemple, des questions telles que :

*Pour quelle raison penses tu que cela...?*

*Pourquoi cela est arrivé selon toi ?*

visent directement les idées des élèves sur les explications possibles, tandis que :

*Qu'est-ce que tu vois se passer ici ?*

*Qu'est-ce que tu penses qu'il va se passer si...?*

peuvent recevoir une réponse sans qu'une explication soit nécessaire.

Les questions liées aux compétences d'investigation seront rédigées en fonction des compétences pertinentes dans une situation donnée. Une question conçue pour encourager les élèves à faire une prédiction pourrait être : « Selon vous, qu'est-ce qui peut faire pousser cette plante plus vite ? » ; si l'objectif est d'encourager l'interprétation des données, elle pourrait être : « selon vous, quels sont les éléments différents dans la vitesse de croissance de cette plante ? » Ces questions, qui encouragent l'utilisation des compétences en investigation scientifique, doivent être distinguées de celles ciblant les idées développées à partir d'une investigation ou celles visant à encourager l'application d'idées à d'autres situations que celles étudiées.

Quelques exemples de questions visant d'autres objectifs sont illustrés dans les encadrés 12, 13 et 14, dans le contexte d'une investigation sur les ombres.

Pour accéder aux idées des élèves, une bonne façon de commencer est d'utiliser des questions ouvertes. Si la réponse à la question est trop vague, il peut être nécessaire de les faire suivre de questions plus ciblées et centrées sur la personne. Pour aider les élèves à aller aux étapes suivantes,

les questions utiles doivent souvent prolonger les idées qu'ils ont exprimées et doivent être conçues pour les encourager à mettre en relation les différentes observations, afin d'apprécier si leurs idées fonctionnent dans d'autres circonstances. Elles peuvent aussi encourager les élèves à envisager des idées différentes en construisant un « échafaudage » autour de solutions alternatives. Ces questions sont susceptibles d'être un peu moins ouvertes que celles destinées à faire émerger des idées.

Les réponses aux questions destinées à manifester les compétences d'investigation des élèves, indiquées dans l'encadré 13, peuvent être des actions ainsi que des mots qui donnent aux enseignants l'occasion de voir les compétences déjà acquises des élèves. Les questions de découverte offrent également l'occasion de développer des compétences d'investigation. Encore une fois, elles sont ouvertes et formulées pour encourager les élèves à répondre sans chercher forcément la « bonne » réponse. Certaines questions peuvent impliquer la construction d'un échafaudage pour soutenir la réflexion dans un certain sens, comme sur les variables mises en jeu, ou par rapport à ce qu'ils ont donné comme réponse à leur question initiale.

*Encadré 12 : Les questions pour découvrir et encourager le développement des idées*

Découvrir

- Qu'est ce qui fait une ombre selon vous ?
- Pourquoi, d'après vous, ces objets produisent-ils des ombres plus sombres que les autres ?
- Comment expliquez-vous la forme de l'ombre ?

Introduire les étapes suivantes

- Comment expliquer avec vos idées pourquoi la couleur de l'ombre est la même pour tous les objets ?
- Avez-vous une autre idée pour expliquer la forme de l'ombre ?
- Sachant que l'objet intercepte la lumière, pourquoi l'ombre est-elle plus grande si l'objet est plus proche de la lampe-torche ?

*Encadré 13 : Les questions pour découvrir et encourager le développement des compétences d'investigation*

Découvrir

- Que voulez-vous savoir sur les ombres ? (les questions qui se posent)
- Que pensez-vous qu'il va se passer si on déplace cet objet par ici ? (prédiction)
- Que pourriez-vous faire pour trouver ce qui cause une différence dans la taille de l'ombre ? (investigation)
- Avez-vous trouvé s'il existe un lien entre la position de la lampe-torche et la taille de l'ombre ? (interprétation)

Introduire les étapes suivantes

- Que découvrez-vous lorsque vous mesurez la taille de l'ombre avant et après avoir déplacé la lampe-torche ?
- Si l'ombre devient plus grande quand vous déplacez la lampe-torche par ici, que pensez-vous qu'il va se passer si vous la déplacez à un autre endroit ?
- Comment pouvez-vous être sûrs que c'est la position de la torche, et non d'autre chose, qui fait la différence ?
- Qu'avez-vous découvert sur la manière de changer la taille de l'ombre??

*Encadré 14 : Les questions pour encourager la collaboration, le partage des idées, la réflexion et l'évaluation*

## La collaboration et le partage des idées

- Combien d'idées différentes votre groupe a-t-il proposées pour essayer d'expliquer ce que vous avez trouvé ?
- Après avoir expliqué ce que vous avez trouvé à partir d'idées différentes, quelle est celle qui vous semble la meilleure ?
- Parmi toutes ces idées, quelles sont celles que vous pouvez tester ?
- Quel est le moyen de savoir quelles sont les idées qui fonctionnent ?
- Quel pourrait être le rôle de chaque membre de votre groupe dans cette investigation ?

## Réflexion and évaluation

- Qu'avez-vous découvert que vous ne connaissiez pas auparavant ?
- Avez-vous changé d'avis sur...?
- Qu'est-ce qui vous a fait changer d'avis ?
- Qu'est-ce que vous ne comprenez pas dans...?
- Y-a-t-il quelque chose que vous voudriez encore savoir ?
- Si vous deviez recommencer, que changeriez-vous pour apprendre davantage ?

Les questions dans l'encadré 14 sont conçues pour encourager les élèves à travailler en authentique coopération et non en tant qu'individus- même s'ils travaillent en groupes ou par deux- ou en concurrence les uns avec les autres. Les questions demandent une réponse qui reflète une pensée rassemblant plusieurs aspects, par exemple la manière d'expliquer certaines observations ou de planifier une investigation, etc. Les questions portant sur la réflexion et l'évaluation ont pour but de faire retracer dans l'esprit des élèves ce qu'ils ont fait et ainsi de leur faire prendre conscience de comment leurs idées ont été modifiées. Sans cette réflexion, leurs idées peuvent retomber à l'ancien mode de pensée. Ces questions demandent aux enfants de parler de ce qu'ils ont appris et de comment ils ont appris, par conséquent ils « apprennent à apprendre », autant qu'à connaître ce qu'ils ont étudié.

**Le temps de la réponse**

Des questions soigneusement formulées méritent des réponses réfléchies et les élèves doivent avoir le temps de répondre. Les questions proposées dans les encadrés 12, 13 et 14 sont conçues pour susciter la réflexion et une réponse réfléchie. Il est peut-être possible de faire un contrôle rapide de la mémoire ou un quiz, mais ce n'est pas ce qui nous intéresse dans le cadre d'un enseignement fondé sur l'investigation. La pression exercée pour répondre rapidement réduit la valeur du questionnement selon les objectifs déjà mentionnés. Il est donc nécessaire de signaler aux enfants que l'enseignant attend une réponse réfléchie et non pas rapide. Voici plusieurs possibilités :

- La première solution est d'augmenter le « temps d'attente », soit la durée entre le questionnement et la réception de réponse. En général, les enseignants s'attendent à une réponse trop rapide et ainsi dissuadent les élèves de réfléchir. Les fameuses études de Budd-Rowe<sup>71</sup> ont montré que la prolongation du temps d'attente de la réponse a un effet remarquable sur la qualité des réponses des élèves. Elle a constaté que les enseignants attendaient, en

71 Budd-Rowe, M. (1974) Relation of wait-time and rewards to the development of language, logic and fate control: Part II, *Journal of Research in Science Teaching*, 11(4) 291-308

moyenne, moins d'une seconde après avoir posé une question si aucune réponse n'a été proposée, avant de reformuler, de donner un conseil ou de reposer une question plus facile. Lorsque les enseignants ont été invités à augmenter le temps d'attente à huit ou neuf secondes, la qualité des réponses des élèves a augmenté de façon spectaculaire.

- La deuxième solution est d'éviter de reformuler une question si la réponse n'est pas formulée aisément. Poser la question d'une manière différente la rend inévitablement plus fermée et moins utile. Lorsque cela se produit régulièrement, les élèves réalisent qu'en attendant, l'enseignant posera toujours une question plus simple, donnant souvent un indice sur la réponse attendue.
- Certains enseignants trouvent qu'il est préférable de ne pas laisser les élèves lever la main pour répondre à ce type de question de réflexion. Tout le monde doit être en mesure de répondre avec un délai de réflexion. Ainsi, l'enseignant laisse le temps de la réflexion, puis appelle les élèves par leur nom pour participer. La pensée, et non la vitesse, est valorisée et elle encourage tous les élèves à accorder leur attention à la question. Ils s'engagent ainsi davantage à écouter les réponses des autres élèves.
- Une autre stratégie, adaptée à certaines questions et situations, peut être de laisser s'instaurer une discussion entre l'élève proposant une réponse et un partenaire ou un groupe, pendant deux ou trois minutes avant que l'enseignant ne demande les contributions.

## Les commentaires<sup>72</sup> faits aux élèves

Les commentaires ont été décrits comme « l'une des influences les plus puissantes sur l'apprentissage et la réussite » avec un impact qui peut être « soit positif soit négatif. <sup>73</sup> » Ils jouent un rôle clé dans l'évaluation formative des élèves puisque c'est le mécanisme par lequel les possibilités d'apprentissage futures sont affectées par l'acquis antérieur. Les commentaires sont donnés par les enseignants aux élèves oralement ou par écrit. Ils peuvent également être donnés inconsciemment par le geste, l'intonation ou l'action comme lors de l'attribution de tâches aux élèves. La manière d'émettre des commentaires et leur contenu sont influencés par la conception de l'apprentissage qu'ont les enseignants (voir chapitre 4). La vision constructiviste de l'apprentissage conduit à une interaction entre l'enseignant et les élèves, dans laquelle les élèves répondent aux commentaires et aux suggestions de l'enseignant, contrairement à la communication à sens unique d'enseignant à élèves qui correspond à une vision comportementaliste de l'apprentissage.

Deux aspects essentiels des commentaires faits aux élèves méritent d'être examinés : la forme et le contenu.

### La forme des commentaires

Traditionnellement, la plupart des commentaires sur le travail des élèves sont sous forme de note, de pourcentage ou d'une autre forme de jugement pertinent. Une étude de Butler<sup>74</sup> sur les commentaires a fait une comparaison entre deux formes de notes : la première incluant des commentaires sur le travail et les méthodes pour l'améliorer, la seconde étant juste constituée de commentaires et de note. Cette étude ingénieuse et complexe implique différents types de tâches, ainsi que des élèves à différents niveaux. L'un des résultats montre que des commentaires, formulés seuls, ont engendré une meilleure réussite pour tous les élèves et toutes les tâches. Il est intéressant de voir que la combinaison de commentaires et de notes n'est pas plus efficace que les notes seules.

72 *Commentaire* traduit ici *feedback*, c'est à dire le retour que fait l'enseignant vers l'élève.

73 Hattie, J. and Timperley, H. (2007) The power of feedback. *Review of Educational Research*, 77, 81-112.

74 Butler, R. (1988) Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on interest and performance, *British Journal of Educational Psychology* 58, 1-14

Les élèves semblent se préoccuper des seules notes et ignorent alors les commentaires qui les accompagnent. Ils cherchent dans les notes un jugement plutôt qu'une aide pour la poursuite de leur apprentissage. En absence de note, ils cherchent plutôt à faire ce que l'enseignant souhaite. Les commentaires des enseignants ont alors une chance d'améliorer l'apprentissage. Pour cela, ces commentaires doivent être positifs, sans jugement, et si possible en permettant à l'élève d'identifier les étapes suivantes.

Dans leur analyse de l'évaluation formative des élèves, Black et al (2003) ont présenté aux enseignants d'une part les recherches de Butler, d'autre part les résultats d'observations faites de leurs propres classes. Ceci montra que :

- Les élèves lisent rarement les commentaires et préfèrent comparer leurs notes avec celles des autres : c'est d'ailleurs souvent leur première réaction.
- Les enseignants donnent rarement aux élèves le temps de lire les commentaires écrits sur le travail en classe. Seulement quelques élèves (voire aucun) les relisent à la maison.
- Souvent les commentaires sont courts ou non spécifiques, se limitant par exemple à une remarque du type « les détails ? »
- Les mêmes commentaires écrits se reproduisent fréquemment dans le cahier des élèves, par conséquent les élèves n'en prennent pas note et ne donnent pas suite à ces commentaires.<sup>75</sup>

Lorsque les élèves ont été interrogés sur la manière dont leurs cahiers ont été corrigés, ils ont exprimé leur désaccord à l'utilisation de l'encre rouge et souhaité que leurs enseignants écrivent lisiblement avec des commentaires compréhensibles.

### **Le contenu des commentaires**

Les recherches et l'expérience de pratiques efficaces ont toutes montré l'importance de la distinction entre les commentaires qui fournissent des informations et les commentaires qui formulent un jugement. Cette distinction s'applique aussi bien aux commentaires écrits qu'oraux. Les commentaires qui fournissent des informations :

- Sont centrés sur la tâche et non sur la personne ;
- Encouragent les élèves à réfléchir sur le travail et non à donner la « bonne » réponse ;
- Indiquent les étapes suivantes avec des indications de progresser.

Les commentaires qui formulent un jugement :

- Expriment si *l'élève* a bien fait (ce qui inclut aussi bien les éloges que les critiques) plutôt que si le travail a été bien fait ;
- expriment un jugement qui encourage les élèves à se catégoriser ;
- Donnent lieu à un pourcentage ou une note que les élèves utilisent pour se comparer entre eux.

Un autre aspect des commentaires efficaces est de guider les élèves sur des points pertinents pour atteindre les objectifs d'une leçon en particulier. Ainsi, si l'objectif principal de l'investigation est de développer des compétences en recherche scientifique, alors les commentaires doivent être autour de cet aspect plutôt que sur les résultats obtenus.

Au début de leur travail avec les enseignants, Black et al ont remarqué que les commentaires des enseignants sur les travaux des élèves étaient, soit une évaluation systématique générale (« Bien », « Très bien »), soit focalisés sur la présentation ou l'intégralité du travail. L'un des enseignants, conscient des incohérences de ce type de remarques, a écrit dans ses notes :

---

75 Black et al, (2003) op cit p 43

Un commentaire ennuyeux et inutile tel que « Bon travail, Jaspaul, c'est beaucoup plus propre et ça se voit que vous avez fait des efforts » ne produira pas un changement significatif car il ne dit rien sur l'apprentissage de l'individu. Sans cible, l'élève est conscient que l'enseignant est content de lui et on ne saurait lui reprocher de penser que la propreté est tout ce qui compte... Les élèves ont souvent du mal à se rendre compte de ce qu'ils apprennent, parce que nous, les enseignants, ne le leur communiquons pas d'une manière appropriée.<sup>76</sup>

Voici des commentaires qui aident l'apprentissage :

« Vous avez correctement identifié les éléments chimiques et les composés chimiques, maintenant essayez de donner une explication générale sur la différence entre les éléments et les composés.»

« Revoyez vos notes sur la structure cellulaire pour voir si vous êtes toujours d'accord avec ce que vous avez écrit ici.»

« Qu'est-ce qui vous a fait décider que la masse des objets n'a aucune influence sur la rapidité avec laquelle ils tombent ? »

### Les conclusions sur les commentaires aux élèves

D'après ce que révèlent les recherches, si les commentaires ont pour but de faciliter l'apprentissage, alors :

- Ils doivent être sous forme de commentaires sans aucune forme de notation ;
- Oralement ou par écrit, les commentaires sur le travail des élèves doivent identifier ce qui a été bien fait, ce qui pourrait être amélioré et comment s'y prendre pour l'améliorer ;
- Les commentaires doivent aider les élèves à prendre conscience de ce qu'ils ont appris ;
- Les enseignants doivent vérifier que les élèves ont bien compris leurs commentaires ;
- Un temps doit être prévu pour que les élèves lisent et, si possible, répondent aux commentaires.

### Les commentaires dans l'enseignement

L'évaluation formative des élèves concerne autant les commentaires des élèves adressés aux enseignants que les commentaires de l'enseignant adressés aux élèves. Ces deux aspects sont étroitement liés, car la manière dont les élèves répondent aux questions et aux commentaires de l'enseignant et des autres élèves éclaire l'enseignant pour les décisions à prendre pour la suite de l'activité. Puisque chaque apprenant est un individu singulier, il n'existe pas de voie unique à suivre pour assurer l'apprentissage. Les enseignants doivent juger la valeur d'une intervention à partir de l'impact de leurs questions et d'autres type d'actions. Afin de recueillir des données pertinentes pour décider de leurs interventions, les enseignants ont besoin d'être parfaitement clairs avec les élèves sur les objectifs ciblés. Puis, ils peuvent choisir parmi une vaste gamme les données importantes à utiliser. Ce faisant, ils seront en mesure de recueillir des données pendant que les élèves participent aux investigations, en les observant, les interrogeant et en écoutant la manière dont ces élèves utilisent les mots et étudient les documents. Une source importante de feedback en direction des enseignants provient de l'auto-évaluation et de l'évaluation par les pairs entre élèves, car si les élèves ne sont pas clairs sur ce qu'ils doivent réaliser par leur travail, ils risquent alors d'utiliser des critères inappropriés pour juger de leur réussite.

<sup>76</sup> Black et al, (2003) op cit p 45

Ce feedback aide les enseignants à décider de la manière d'intervenir sur les activités des élèves. Le processus est cyclique et chaque décision modifie la situation. Toutes les interventions n'auront pas l'impact positif désiré, car ce qui se passe dans les salles de classe ne se déroule pas toujours, voire rarement, comme prévu. Les commentaires que les enseignants reçoivent en observant les réactions des élèves leur permettent d'essayer autre chose, si nécessaire, afin d'aider les élèves à progresser. Lorsque les élèves sont en difficulté, l'enseignant peut être amené à changer son organisation plutôt que de risquer un sentiment d'échec. Ainsi, les commentaires permettent aux enseignants d'adapter leur enseignement afin d'optimiser l'apprentissage.

## L'auto-évaluation et l'évaluation par les pairs chez les élèves

Un objectif commun à l'évaluation formative des élèves et à l'ESFI est de faire en sorte que les élèves deviennent de plus en plus en mesure de participer aux décisions concernant la qualité de leur travail et de développer leur compréhension de ce qui est à apprendre. Les apprenants sont, de toute façon, responsable de leur apprentissage, mais le fait d'*assumer la responsabilité* dépend de leur participation aux décisions. Cette participation est représentée par les flèches doubles dans la figure 1, page 21.

Les élèves, comme tous les apprenants, dirigent leur effort de manière plus efficace en sachant ce qu'ils visent à atteindre plutôt que simplement ce qu'ils ont à faire. Les enseignants réussissent à communiquer aux élèves ce qu'ils doivent faire (« enveloppez les cubes de glace dans les différents matériaux et observez celui qui prend le plus de temps à fondre »), mais ont en général plus de mal à leur donner un objectif (« voir si certains matériaux sont meilleurs que d'autres pour éviter la fonte de la glace, et essayez d'expliquer ce que vous trouvez »). La condition préalable à un jugement utile est de demander aux élèves de comprendre ce qu'ils essaient de faire : non pas ce qui doit être trouvé, mais quelle est la question ou le problème à résoudre.

Bien sûr, si la question ou le problème est soulevé par les élèves, la nécessité d'une communication ne se pose pas. Mais ce cas ne représente qu'une partie de l'ensemble des pratiques pédagogiques utilisées par les enseignants. Il y aura toujours des situations où, même si les enseignants introduisent la question de façon à ce que les élèves l'adoptent comme leur propre question, il faut quand-même s'assurer que les élèves ont bien compris l'objectif de l'activité. L'énoncé des objectifs au début de la leçon n'est pas le seul moyen, ni forcément le meilleur. La compréhension des objectifs et des raisons pour lesquelles l'élève travaille d'une façon spécifique peut être renforcée par le dialogue et des questions posées pendant l'activité ainsi que par des échanges sur ce qui a été fait et trouvé.

Afin d'évaluer leur travail, les élèves doivent non seulement connaître le but de ce qu'ils font, mais posséder aussi une certaine notion des standards de connaissances à viser, c'est ce que l'on appelle un « bon travail » dans un contexte particulier. Une partie est exprimée implicitement par les commentaires que les enseignants donnent aux élèves. Les enseignants peuvent aussi discuter plus explicitement ce qui rend un travail ou une investigation meilleure qu'une autre, en utilisant des exemples anonymes recueillis à cette fin. Alternativement, les exemples tirés des documents publiés pour aider les enseignants à évaluer les travaux pourraient être partagés avec les élèves (voir plus loin). L'encadré 15 décrit quelques exemples de la façon dont des enseignants du primaire ont abordé la discussion sur ce qui est un "bon" travail.

Avec des élèves plus âgés, des approches plus directes et plus sophistiquées peuvent être utilisées pour encourager l'auto-évaluation. L'encadré 16 décrit comment un professeur de biologie a aidé ses élèves de seize ans à comprendre la manière d'évaluer leur travail en clarifiant leurs idées sur la nutrition des plantes.

*Encadré 15 : Communiquer aux élèves de l'école primaire les critères d'évaluation de leur travail*<sup>77</sup>**Utilisation des exemples**

Une enseignante d'une classe d'élèves de dix ans a passé un certain temps au début de l'année scolaire à discuter avec sa classe de ce qui constitue un « bon » rapport d'investigation scientifique. Elle a donné aux élèves deux exemples anonymés de notes sur une investigation produite par les élèves des années précédentes. Le premier était un compte-rendu clair, bien aménagé afin que le lecteur puisse comprendre ce qui avait été fait, même si l'écriture était inégale avec quelques fautes d'orthographe. Il y avait des diagrammes qui accompagnaient le texte avec des légendes. Les résultats étaient dans un tableau, et l'élève avait expliqué leur signification en admettant que les résultats ne répondaient pas entièrement à la question initiale. Il y avait également un commentaire sur la façon dont les choses auraient pu être améliorées. L'autre compte-rendu était propre, joli à regarder (les diagrammes étaient coloriés et sans légendes) mais le contenu ne contenait aucune des caractéristiques indiquées dans l'autre travail.

L'enseignant a demandé aux élèves de travailler en groupe pour comparer les deux travaux en listant les éléments positifs et négatifs. Ensuite, ils ont été invités à exprimer ce qu'ils pensaient être les éléments les plus importants pour établir un « bon » rapport. L'enseignant a rassemblé toutes les idées en ajoutant quelques idées personnelles avec lesquelles les élèves étaient d'accord. Plus tard, elle a fait des copies pour tous les élèves à conserver dans leurs dossiers de sciences. Mais elle a également continué à explorer avec eux la façon de mener une investigation dans le but d'écrire un bon rapport. Ces points ont été exprimés dans le langage des enfants et imprimés pour eux.

**Brainstorming**

Une variante de ce qui précède est d'organiser un *brainstorming*, par exemple sur la manière de mener une investigation spécifique afin que les enfants soient sûrs du résultat. La liste de ce à quoi il faut être attentif peut être transformée en questions (A-t-on la même chose partout sauf ...? Avons-nous changé...? Avons-nous cherché ...? Avons-nous vérifié les résultats? etc.) Avant de terminer leur investigation, les élèves vérifient à l'aide de leur liste et celle-ci devient un outil d'auto-évaluation pour leur travail.

**Discussion sur le « meilleur travail »**

Cette approche peut être utilisée avec des élèves d'environ huit ans. Les élèves sélectionnent leur « meilleur » travail pour le mettre dans un dossier. Un moment est réservé à une discussion entre l'enseignant et chaque élève, portant sur les raisons pour lesquelles certains travaux ont été sélectionnés. Au cours de cette discussion, la manière dont l'élève juge la qualité de son travail devient claire. Tous les points sont acceptés sans commentaire, qu'ils reflètent ou non ce que l'enseignant pense être un bon travail. Pour clarifier les critères utilisés par les élèves, l'enseignant pourrait demander : « dis-moi ce que tu as apprécié spécialement dans ce travail ? » Peu à peu, il sera possible de proposer des critères sans imposer les choix des élèves, par exemple par les commentaires sur le travail : « C'est un très bon moyen de montrer tes résultats, je vois tout de suite ce qui est meilleur. », « je suis content que tu penses que c'est ta meilleure investigation, parce que même si tu n'as pas obtenu les résultats que tu attendais, tu l'as fait avec beaucoup d'attention et tu as fait en sorte que le résultat soit juste. »

77 Adapted from Harlen, W. (2006) *Teaching, Learning and Assessing Science 5-12*. London: Sage. p171



*Encadré 16: l'utilisation de l'auto-évaluation par des élèves du secondaires, afin d'améliorer leur travail*<sup>78</sup>

Les élèves étaient invités à discuter la question suivante: « Si une personne mal intentionnée rejetait un produit chimique pouvant détruire la chlorophylle, quel effet cela aurait-il sur les plantes? » Chaque groupe de quatre personnes devait écrire entre trois et cinq critères qu'ils jugeaient nécessaires pour donner une bonne réponse écrite à cette question. Ces critères ont été discutés ensuite par l'ensemble de la classe dans le but d'élaborer une liste finale:

- Dire quelle est l'action de la chlorophylle ;
- Expliquer la photosynthèse comme un processus ;
- Ecrire l'équation de la photosynthèse ;
- Décrire l'effet sur les plantes sans chlorophylle ;
- Ajouter des effets secondaires pouvant stopper la photosynthèse.

Les élèves ont ensuite écrit leurs réponses dans des devoirs corrigés par l'enseignant avec des commentaires écrits portant uniquement sur leur travail. Puis, par paire, les élèves lisent les commentaires et le travail de l'autre pour vérifier qu'ils ont bien compris les instructions de l'enseignant pour s'améliorer. Puis ils ont eu du temps pour remanier et améliorer leurs réponses pendant le cours.

### L'évaluation par les pairs

L'évaluation par les pairs entre élèves s'est vue attribuer un rôle central dans l'évaluation formative des élèves après avoir été vigoureusement défendue par Sadler.<sup>79</sup> Black fournit plusieurs arguments pour encourager les élèves à juger mutuellement leur travail. Certains de ces arguments ont pour but d'aider les élèves à mieux comprendre les objectifs et les critères de qualité en examinant le travail d'un autre plutôt que le leur. D'autres sont fondés sur l'aspect pratique : les élèves sont plus motivés à améliorer leur travail s'ils savent qu'il sera lu par un autre élève.<sup>80</sup> Cependant, comme d'autres chercheurs l'ont souligné, ces arguments ne tiennent pas compte des relations de pouvoir associées à une personne qui juge le travail des autres.<sup>81, 82</sup> Des études de cas par Crossouard, portant sur l'évaluation par les pairs chez des élèves âgés de onze et douze ans, fournissent des indications inquiétantes de la manière dont le sexe, la classe sociale et les hiérarchies scolaires sont « impliqués dans des processus auxquels est attribuée une « neutralité » supposée de jugement dans l'évaluation ». <sup>83</sup> Les avantages de l'évaluation par les pairs ont été observés comme inégalement répartis, leur mise en œuvre soutenant certains élèves tout en étant « oppressante » pour d'autres. Ainsi, les enseignants peuvent avoir besoin d'aide dans l'identification des questions d'équité et dans la prise de conscience de l'influence de la classe sociale, du sexe et de la capacité générale, lorsqu'ils mettent en œuvre une évaluation des élèves par leurs pairs.

78 Based on Black et al (2003) op cit p 63

79 Sadler (1989) op cit

80 Black et al (2003) op cit p 50

81 Crossouard (2012) Absent presences: the recognition of social class and gender dimensions within peer assessment interactions, *British Educational Research Journal*, 38 (5) 731-748

82 Pryor, J. and Lubisi, C. (2001) Reconceptualising educational assessment in South Africa – testing times for teachers, *International Journal for Educational Development*, 22 (6), 673-686

83 Crossouard (2012) op cit p 736

## Chapitre 6

# La mise en œuvre de l'évaluation sommative des élèves dans l'ESFI

Après avoir examiné au chapitre 5 les modalités portant sur la mise en pratique d'une évaluation formative des élèves pour renforcer l'apprentissage par l'investigation, nous examinons dans ce chapitre quelques méthodes concernant la mise en œuvre d'une évaluation sommative des élèves. Son objectif consiste, à un moment donné, à fournir des informations sur les élèves en mettant l'accent sur leurs connaissances scientifiques, leur compréhension et leurs compétences en investigation scientifique. Contrairement aux exigences de l'évaluation formative, la fiabilité est importante dans cette évaluation sommative, car ses résultats peuvent être utilisés pour comparer ou sélectionner des élèves mais aussi comme des indicateurs de l'efficacité des enseignants et des écoles. Toutefois, la validité est également importante. Compte tenu de l'interaction entre la validité et la fiabilité, mentionnée au chapitre 1 (encadré 2), l'un des facteurs fondamentaux du choix des méthodes d'évaluation des élèves est de s'assurer que la validité n'est pas compromise dans la recherche d'une apparente précision. Ici, nous nous concentrons sur l'évaluation sommative des élèves, en évaluant leur degré de compréhension en science et leurs compétences en investigation, lesquelles constituent le plus grand défi : négliger ces dernières dans l'évaluation des élèves représente la plus grande menace pour la pratique de l'ESFI.

Ce point de vue ne signifie pas que d'autres résultats, tels que la connaissance des faits, le vocabulaire scientifique ou l'utilisation de conventions, ne doivent pas être évalués. Cependant il existe déjà de nombreuses façons bien connues de le faire, par exemple par le biais des tests et des quiz rapides en classe, élaborés par l'enseignant et mis en œuvre à des moments appropriés. Il est plus difficile d'évaluer les progrès accomplis par des élèves en terme de compréhension et de compétences qui sont les véritables objectifs de l'ESFI. S'ils ne sont pas intégrés dans l'évaluation des élèves, il y a de grande chance pour qu'ils soient négligés dans l'enseignement.

## L'évaluation de la compréhension et des compétences en investigation

### La compréhension et les notions-clés

Dans un premier temps, nous devons définir la compréhension et son rapport aux connaissances, tant générales que spécifiques. La compréhension n'est pas facile à établir, car selon White:

La compréhension est une fonction continue des connaissances d'une personne, sans représenter une dichotomie ni un contenu linéaire. Dire si quelqu'un comprend est un jugement subjectif, qui varie en fonction du juge et de la situation de la personne jugée. Une connaissance varie en fonction de sa pertinence pour la compréhension, mais cette pertinence relève également d'un jugement subjectif.<sup>84</sup>

La compréhension existe à différents niveaux et de différentes manières pour expliquer les phénomènes. La compréhension d'un même phénomène par un élève de l'école primaire et par un étudiant scientifique de l'enseignement supérieur sera différente, mais les deux peuvent répondre à

84 White, R.T. (1988) *Learning Science*. Oxford: Blackwell. p 52

l'attente du niveau d'étude concerné. Par exemple, un jeune élève pourra expliquer le phénomène de dissolution en référence à ce qu'il observe quand certains solides sont ajoutés à certains liquides. Les élèves plus âgés l'expliqueront en termes de comportement moléculaire, s'appliquant à des solutions de gaz ou de liquides dans des liquides. Le point commun entre ces deux cas se trouve dans les idées auxquelles les élèves ont trouvé un sens en adaptant les expériences qu'ils essaient de comprendre à un moment donné. Comme les expériences des jeunes élèves augmentent et que leurs idées ne permettent plus de donner d'explication satisfaisante, les phénomènes devront être compris d'une façon différente. La compréhension progressive qui accompagne des expériences élargies peut aussi être considérée comme le développement d'idées ou notions successivement « plus amples », rassemblant plus de phénomènes autour d'explications plus puissantes. Les notions-clés qui se construisent sont celles qui peuvent être appliquées dans des contextes différents : elles permettent aux apprenants de donner sens à un large éventail de phénomènes en identifiant les liens essentiels ou des caractéristiques significatives entre des événements, des objets ou des phénomènes différents, sans être détournés par des différences portant sur des caractéristiques superficielles. Ces notions-clés peuvent être des idées *de science* (portant sur les forces et le mouvement) ou porter sur les *procédures de la science* (par exemple sur la manipulation de variables). La mémorisation des faits ou une série de procédures ne permettent pas d'utiliser ces savoirs à des contextes situés au-delà de celui dans lequel il a été acquis. Les connaissances comprises deviennent alors des connaissances utiles qui peuvent être utilisées dans la résolution de problèmes et la prise de décision.

Ainsi, bien que les progrès dans la compréhension n'indiquent pas la *quantité* de connaissances disponibles, ils indiquent si elles sont organisées dans l'esprit sous forme de notions-clés, de sorte qu'elles puissent être facilement accessibles et appliquées à de nouvelles expériences. Cette vision de la compréhension est un aspect essentiel qui distingue la performance d'un « novice » de celle d'un « expert ».

Les experts ont acquis de vastes connaissances et de compétences dans un domaine particulier. Mais surtout, leur esprit a organisé ces connaissances de manière à les rendre plus accessibles et utiles.

... Ces méthodes de codage et d'organisation aident les experts à interpréter de nouvelles informations et à remarquer des caractéristiques et des formes d'informations utiles qui risquent d'être négligées par les apprenants moins compétents. Ces schémas permettent également aux experts, lorsqu'ils sont confrontés à un problème, de faire appel aux aspects pertinents dans leur stock de connaissances.<sup>85</sup>

Voici donc quelques indications portant sur l'évaluation des élèves qui permettent de tirer des conclusions sur leur degré de compréhension :

- La tâche proposée aux élèves pour leur évaluation doit impliquer l'application de connaissances et non pas seulement un rappel.
- Cela signifie que la tâche doit être nouvelle, mais en même temps pas trop éloignée de l'expérience des élèves car sinon elle n'aura pas de sens pour eux.
- Sachant que la science existe dans le monde réel et non dans un monde imaginaire, la tâche doit être authentique avec des éléments et des données réelles.

Il y a aussi d'autres exigences : la tâche d'évaluation doit avoir une durée raisonnable ; elle ne doit pas dépendre de trop de compétences (telles que la capacité de lecture et d'écriture) et s'en tenir à la compréhension qui est l'objet de l'évaluation.

---

85 Pellegrino et al op cit p 72/73

## Les compétences en investigation scientifique

Dans toutes les tâches d'évaluation qui, dans un contexte donné, traitent de sujets faisant sens pour les élèves, leurs performances sont affectées par leur familiarité avec ce contexte et ces sujets. Toutes les compétences sont utilisées dans un certain contexte et par rapport à un certain contenu. Bien que le contexte ait également une influence sur l'application des concepts scientifiques, il présente des problèmes particuliers lorsque l'objectif de la question est d'obtenir des données sur l'utilisation des compétences en investigation.

Des questions et des tâches doivent porter sur *quelque chose* ; les observations sont faites sur des objets et des événements spécifiques ; les investigations sont planifiées pour répondre aux questions sur des phénomènes spécifiques ; les compétences doivent s'appliquer à un sujet défini. La nature du sujet fait une différence quant aux compétences utilisées. Un élève peut être en mesure de planifier une investigation appropriée à une situation où il possède une certaine connaissance de ce que peuvent être les variables à contrôler, mais il peut tout à fait échouer si le sujet lui est inconnu. Ceci a des conséquences importantes pour l'évaluation des élèves. Le sujet d'une tâche ou d'une épreuve spécifique est simplement l'un des nombreux sujets possibles qui pourraient être choisis. En théorie, le résultat d'un élève serait différent si un autre sujet avait été choisi, il y a donc une variation ou une erreur dans les résultats associés au choix. C'est ce que l'on appelle « l'erreur d'échantillonnage ». Il s'agit d'une erreur inévitable car il n'existe pas deux tâches, se référant à des sujets ou des contextes différents, qui puissent être exactement équivalentes.

Il en résulte quelques points concernant l'évaluation des compétences des élèves en investigation scientifique:

- Tout d'abord, il est évident que les élèves doivent être impliqués dans l'utilisation des compétences d'investigation afin d'évaluer ce qu'ils peuvent faire ;
- Deuxièmement, puisque le contexte et le thème de la situation dans laquelle les compétences doivent être utilisées affectent la capacité d'utiliser les compétences, la tâche doit être élaborée si possible dans un contexte connu ou dans plusieurs contextes pour réduire l'erreur d'échantillonnage ;
- Troisièmement, comme dans le cas de l'évaluation de la compréhension des élèves, les tâches doivent être authentiques et engageantes pour les élèves.

La tâche qu'ils entreprennent peut impliquer :

- La réalisation, dans la pratique, d'une investigation complète pour répondre à une question ou un problème avec des occasions de collecter les preuves dans le cadre de l'application d'un ensemble de compétences ;
- La production d'un schéma sur papier, décrivant une investigation complète pour répondre à une question ou à un problème donné ;
- L'analyse d'une partie spécifique d'une investigation, telles que la manipulation ou le contrôle des variables, la collecte de preuves, ou l'interprétation de certaines données fournies.

## Les méthodes d'une évaluation sommative des objectifs de l'ESFI

Cette brève analyse des aspects nécessaires pour évaluer la compréhension et les compétences d'investigation scientifique des élèves montre que, dans les deux cas, les élèves ont besoin de travailler sur des tâches qui soient en rapport avec certains aspects de l'investigation. Il faut également que la tâche contienne des éléments nouveaux, afin que les élèves se servent de leurs connaissances et compétences sans se limiter à un rappel d'informations, d'explications ou de procédures mémorisées. Par exemple, la tâche consistant à trouver comment l'épaisseur d'un fil conducteur affecte sa résistance n'évaluera pas la capacité de planifier et de mener une investigation, s'il s'agit d'une expérience que les

élèves ont déjà faite plusieurs fois. De la même façon, une question portant sur l'estimation de l'âge d'un arbre à partir de ses anneaux de croissance ne conduira pas à la compréhension des idées sur la croissance des arbres si l'explication a été mémorisée pour elle-même et sans lien avec celles-ci.

Pour une meilleure compréhension, la tâche doit exiger soit une explication d'un événement, soit une interprétation des données, soit une prédiction concernant l'application de certains concepts. Au niveau des compétences, la tâche doit être accomplie en utilisant une ou plusieurs compétences d'investigation telles que la prédiction, la planification, la réalisation d'une investigation ou l'interprétation des données fournies. Toutefois, comme déjà indiqué à la page 14, il n'est pas possible d'évaluer les compétences des élèves sans impliquer une certaine connaissance du sujet en question. En même temps, les tâches utilisées pour l'évaluation de la compréhension des élèves nécessitent l'utilisation de compétences (explication, interprétation, prédiction). Ainsi, dans *toutes* les tâches seront requis certains aspects de la compréhension et des compétences. Ce qui détermine si la tâche évalue compréhension et compétences, ce sera essentiellement le degré de demande pour l'une ou l'autre et la valeur accordée aux différents types de réponses dans la notation.

En ce qui concerne les méthodes d'évaluation des élèves dans cette optique, nous allons d'abord examiner le rôle des tests, puisqu'il s'agit de la modalité la plus répandue de l'évaluation sommative des élèves. Ensuite, nous verrons des alternatives aux tests.

## Le potentiel et les limites des tests pour évaluer les objectifs de l'ESFI

L'évaluation sommative des élèves doit être aussi fiable (sans erreur ni préjugé) que possible. On a tendance à utiliser des tâches spéciales ou des tests qui peuvent être contrôlés et présentés à tous les élèves de la même manière, ce qui semble donner les mêmes chances aux élèves pour montrer ce qu'ils peuvent faire. Le test, comme mentionné au chapitre 1, est une méthode d'évaluation des élèves dans laquelle les procédures, comme la tâche à accomplir, les conditions et le temps de réponse, sont précisées. Les tests sont généralement corrigés (notés) avec un schéma prescrit (rubrique), soit par l'enseignant des élèves, soit par des correcteurs externes (souvent des enseignants d'autres écoles), soit par des machines. La raison de l'uniformisation des procédures est de permettre de comparer les résultats des élèves qui subissent les tests à différents endroits. Il existe différentes formes de tests selon la nature de la tâche et la forme de la réponse :

- Les tests de performance (parfois appelé « activité pratique ») ;
- Les tests intégrés (dans le contexte d'un travail régulier) ;
- Les tests à choix multiples (où les réponses alternatives sont fournies) ;
- Les tests comportant des réponses ouvertes (où les élèves écrivent des réponses courtes ou longues avec leurs propres mots) ;
- Les tests à livre ouvert (où les élèves ont accès à un nombre contrôlé de sources) ;

... et bien d'autres. La plupart implique de l'écrit, sauf dans certains tests de performance et de lecture pour les jeunes enfants. Les tests plus formels, menant à un certificat ou un diplôme, sont souvent décrits comme des examens.

On comprend mieux ce que l'ensemble des tests permettant d'évaluer les connaissances scientifiques et les compétences en investigation scientifique des élèves en examinant quelques exemples de tests écrits et de performance.

### Les épreuves écrites

Les caractéristiques des questions orales de qualité, mentionnées dans le chapitre 5, s'appliquent également aux questions écrites. Afin d'accéder à la pensée des élèves, il est utile de formuler des questions en termes de ce que pensent les élèves. Des questions « ouvertes » qui demandent aux

élèves de répondre avec leurs propres mots sont aussi plus susceptibles d'obtenir des informations sur ce qu'ils savent et comprennent que les questions à choix multiples. Elles sont ouvertes aux prédictions et aux indices donnés par les mots employés dans les options. Toutefois, des questions à réponse ouverte ont aussi des inconvénients, en particulier pour les jeunes élèves qui éprouvent plus de difficultés à être précis à l'écrit qu'à l'oral. L'interprétation des réponses écrites des élèves peut également poser un problème, bien qu'il existe maintenant des programmes informatiques qui peuvent faire cela avec une grande fiabilité.<sup>86</sup> En outre, il est nécessaire que la question fasse comprendre aux élèves le type et l'ampleur de la réponse requise. Par exemple, une question comme : « Quelles sont les principales différences entre X et Y? » peut conduire à une longue liste de différences, certaines pertinentes, certaines moins, et d'autres inexacts. Il est donc préférable de préciser les critères des réponses, comme par exemple : « Écrivez, selon vous, les trois différences les plus importantes entre X et Y qui leur permettent de survivre dans des habitats différents. » Ensuite, il faut fournir des espaces de réponse qui indiquent les limites de la longueur de la réponse. Les avantages et les inconvénients des différents formats de questions constituent en général un mélange : l'équilibre de chaque type varie selon les objectifs du test, l'objet particulier de la question et l'âge des élèves. D'autres points émergeront de l'analyse de quelques exemples.

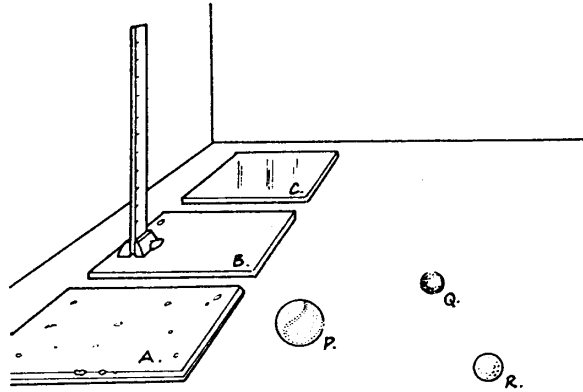
---

86 Streeter, L. et al (2011) *Pearson's Automated Scoring of Writing, Speaking, and Mathematics*. Pearson

Exemple 1: *Rebond de balles* (Extrait du rapport APU sur la performance à l'âge de onze ans)<sup>87</sup>

Emma et Anita cherchent à savoir si la nature de la surface sur laquelle la balle rebondit a un effet sur la hauteur du rebond

Elles ont trois types différents de surfaces appelés A, B et C et trois balles différentes appelées P, Q et R.



Afin de réaliser un test équitable, que doivent-elles changer dans l'expérience et que doivent-elles conserver?

Cochez changer ou conserver pour chaque élément ci-dessous.

	Changer	Conserver
La balle	<input type="checkbox"/>	<input type="checkbox"/>
La surface	<input type="checkbox"/>	<input type="checkbox"/>
La hauteur de la balle lâchée au dessus du sol	<input type="checkbox"/>	<input type="checkbox"/>

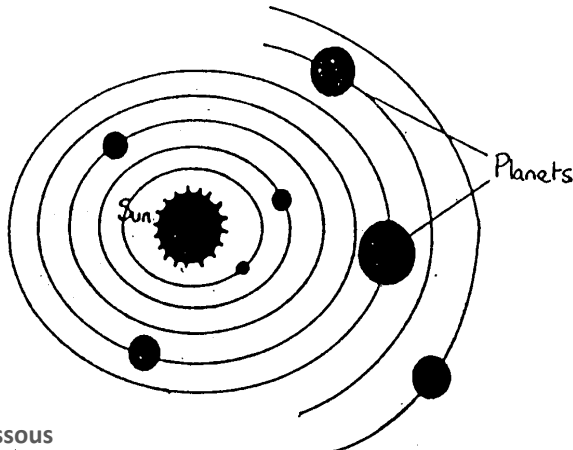
### Commentaire

L'exemple 1 (rebond de balles) demande aux élèves d'identifier la situation décrite et la question posée. Le sujet est probablement connu par tous les élèves, donc le niveau de connaissance requis est faible et la difficulté de la tâche est de procéder à un test juste. Le format de réponse et l'exigence du schéma de notation pour lequel les réponses doivent être correctes dans chaque case rendent très faible la chance de deviner la bonne réponse. C'est une façon de transformer une question à choix multiples pour réduire la chance de deviner les réponses. Cela signifie également que les élèves doivent lire et comprendre les instructions pour donner leur réponse, sinon il y a un risque d'échec pour des raisons autres que l'absence des compétences nécessaires à la réponse.

87 DES, DENI and WO (1985) *APU Science in Schools Age 11 Report no 4*. London: HMSO

Exemple 2: *Les planètes* (De APU, âge de onze ans Rapport No 1)<sup>88</sup>

Les planètes tournent autour du Soleil



Regarder le tableau ci-dessous

Planète	Distance du Soleil à la planète	Durée pour faire le tour du Soleil
Mercuré	58 millions de kilomètres	88 jours
Vénus	108 millions de kilomètres	225 jours
Terre	150 millions de kilomètres	1 an
Jupiter	780 millions de kilomètres	12 ans
Uranus	2870 millions de kilomètres	84 ans
Neptune	4500 millions de kilomètres	165 ans

a) Il y a une planète qui ne figure pas dans ce tableau. Elle se situe à environ 1430 millions de kilomètres du soleil. Selon vous, environ combien de temps faudra-t-il pour que cette planète fasse un tour complet autour du soleil?

100 ans

100 jours

30 ans

300 ans

b) Pourquoi pensez-vous qu'il faille ce temps ?

Parce que.....

.....

*Commentaire*

Dans l'exemple 2 (les planètes), toutes les informations nécessaires pour répondre à la question sont données et les élèves n'ont pas besoin de connaître quoi que ce soit sur les planètes. L'objectif prévu de la tâche consiste à reconnaître la relation entre les deux séries de chiffres donnés. Cependant, présenté « à froid » à un élève (c'est-à-dire en dehors d'une étude sur le système solaire), ce test peut donner l'impression d'exiger plus de connaissances qu'il n'en demande en réalité. Certains élèves peuvent ne pas répondre si leur première impression est qu'ils ont besoin d'avoir des connaissances sur les planètes. La notation peut être fixée pour que certains points soient accordés pour la partie a) même si il y a une chance sur cinq de deviner la bonne réponse, tout en donnant plus de points pour la partie b) selon qu'elle permet ou non de conclure que la relation est comprise explicitement. La nature ouverte de la partie b) est plus exigeante que la sélection parmi les choix offerts en a). Elle fournit davantage d'informations sur la façon dont les élèves interprètent les données, mais sa notation est moins facile.

88 DES, DENI and WO (1981) *Science in Schools Age 11 Report No 1*. London: HMSO



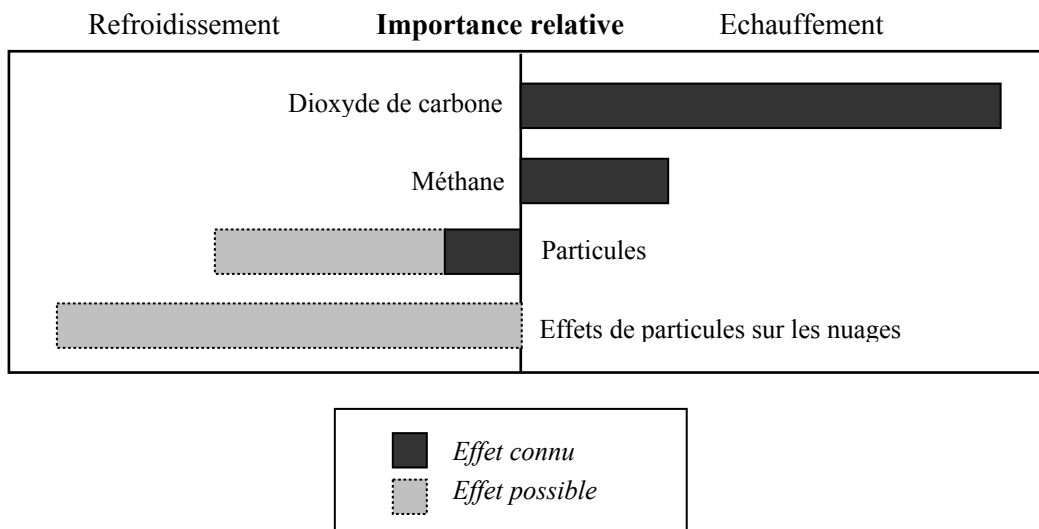
Exemple 3: *Le changement climatique* (extrait de l'évaluation de PISA sciences 2000)<sup>89</sup>

Lisez les informations ci-dessous et répondez aux questions suivantes

### QUELLES SONT LES ACTIVITÉS HUMAINES QUI CONTRIBUENT AU CHANGEMENT CLIMATIQUE?

La combustion du charbon, du pétrole et du gaz naturel, la déforestation, ainsi que les diverses pratiques agricoles et industrielles modifient la composition de l'atmosphère et contribuent au changement climatique. Ces activités humaines ont créé une augmentation des concentrations de particules et de gaz à effet de serre dans l'atmosphère.

L'importance relative des principaux contributeurs au changement de température est illustrée sur la figure 1.



**Figure 1: L'importance relative des principaux contributeurs aux changements de température de la Terre.** Source: adapté de <http://www.gcric.org/ipcc/qa/04.html>

Les barres qui s'étendent vers la droite de la ligne centrale indiquent un effet d'échauffement. Les barres qui s'étendent vers la gauche de la ligne centrale indiquent un effet de refroidissement. L'effet relatif des « particules » et des « effets de particules sur les nuages » sont assez incertains : dans chaque cas, l'effet possible se situe dans la plage indiquée par la barre gris clair.

La figure 1 montre que les augmentations des concentrations de dioxyde de carbone et de méthane ont un effet sur le réchauffement. L'augmentation des concentrations de particules a un effet de refroidissement sous deux formes : les « particules » et les « effets de particules sur les nuages ».

89 OECD (2000) op cit p 88

**Question 1:**

Utilisez les informations de la figure 1 pour soutenir le point de vue selon lequel la priorité doit être accordée à la réduction des émissions de dioxyde de carbone des activités humaines mentionnées.

---

---

---

**Question 2:**

Utilisez les informations de la figure 1 pour soutenir le point de vue selon lequel les effets de l'activité humaine ne constituent pas un réel problème.

---

---

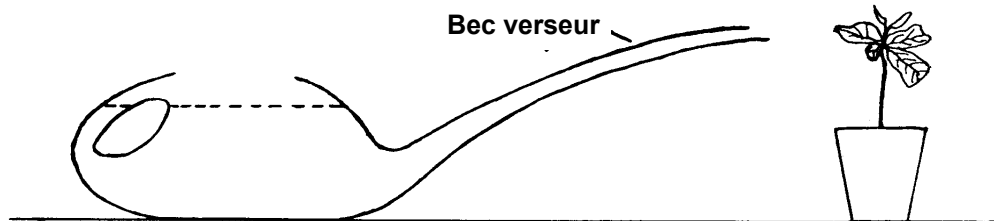
---

*Commentaire*

L'exemple 3 (le changement climatique) est un autre type d'exercice où les informations sont données volontairement aux élèves, pour venir ici à l'appui de conclusions alternatives par rapport aux actions suggérées par les données. L'information est authentique et présente le type de problème que des élèves, ayant des notions scientifiques, devraient être capables d'aborder. Les deux parties de la tâche illustrent l'incertitude portant dans certains cas sur l'interprétation de l'information scientifique. En théorie, toutes les informations sont fournies et les élèves ont appris à interpréter un graphique. Ils n'ont pas besoin de savoir comment le dioxyde de carbone, le méthane, les particules et leurs effets sur les nuages provoquent un échauffement ou un refroidissement. Cependant, sans aucune connaissance de ces sujets, la question risque d'être dénué de sens et les élèves risquent de ne pas s'engager dans le problème posé.

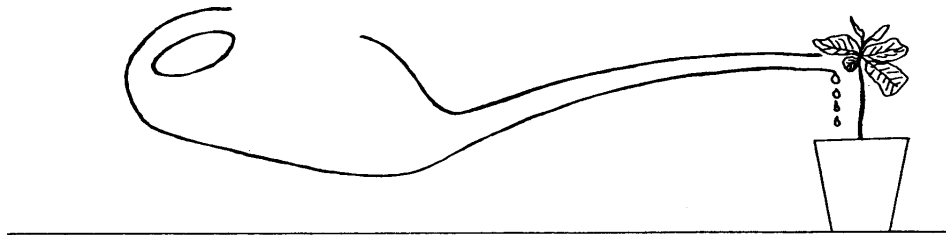
Exemple 4: L'arrosoir (de APU, âge de onze ans Rapport No 1)<sup>90</sup>

**a) La ligne pointillée indique l'emplacement de la surface de l'eau dans cet arrosoir**



**Tracez une ligne pour montrer l'emplacement de la surface de l'eau dans le bec**

**b) L'arrosoir est incliné pour que l'eau s'écoule doucement à travers le bec verseur**



**Tracez une ligne pour indiquer la surface de l'eau maintenant**

#### Commentaire

L'exemple 4 (l'arrosoir) exige l'application de connaissances portant sur l'écoulement de l'eau jusqu'à atteindre l'équilibre. Le contexte est censé être connu et la réponse sous forme de dessin dépend moins de l'écriture et du vocabulaire que les autres exercices. Cependant, même si le système de notation peut tolérer une certaine inexactitude dans le dessin, la réalisation correcte de la tâche exige toujours des élèves une lecture correcte de la façon dont ils sont censés répondre.

<sup>90</sup> DES, DENI and WO (1981) op cit p 98

## Observations générales sur les tests écrits

Plusieurs aspects émergent de ces exemples. Le plus évident est celui d'une exigence incontournable portant sur la lecture et la compréhension de la question, ainsi que sur la capacité d'écriture selon la forme de réponse. Ensuite, vouloir placer la tâche dans un contexte qui puisse sembler réel à l'élève exige la présence d'un scénario dans la tâche. Les élèves doivent lire et s'imprégner du contexte afin de répondre à la question. Il est donc intéressant de se demander si la nature de ce contexte fait toute la différence quant à la façon dont l'élève répond. Par exemple, est-ce que l'identification des variables dans l'exemple 1 peut paraître plus difficile pour certains élèves s'il s'agissait de comparer les effets de modifier des ingrédients dans la préparation un gâteau, ou de la vitesse de voitures jouets sur une pente? Est-ce que la capacité d'identifier une relation dans les données serait affectée par une présentation graphique plutôt que numérique? La forme de la réponse affecte-elle les performances des élèves? D'après les études faites, les réponses à toutes ces questions montrent que ces caractéristiques ont leur importance et que le contexte de la question est particulièrement important pour la capacité et la volonté des élèves de s'engager et de montrer ce qu'ils peuvent faire. D'autres circonstances qui soulèvent des questions d'équité, en particulier pour les élèves de pays ayant une grande diversité de populations, sont indiquées dans l'encadré 17.

### Encadré 17: L'inégalité des chances dans les tests

Les études montrent que la familiarité avec le contexte dans lequel les exercices de test sont mis en place est un facteur influant sur la performance des élèves. Lorsque le contenu est connu de manière inégale parmi la population, ceci soulève des questions quant à l'équité du test. Une expérience pertinente a montré qu'un exercice de test peut exclure certains élèves, par exemple les filles ou ceux qui sont issus de milieux défavorisés. Certains élèves peuvent avoir du mal avec la langue<sup>91</sup> du test lorsqu'il s'agit de leur deuxième ou troisième langue. Ces élèves et les personnes ayant des besoins physiques ou des difficultés d'apprentissage peuvent ne pas avoir l'occasion de montrer ce qu'ils peuvent faire, et par conséquent peuvent se voir refuser l'accès à des étapes plus avancées d'éducation ou de formation.

Pour minimiser les conséquences d'un contexte ou d'un format particulier, une grande variété de contextes et de formats doit être inclus dans un test. Comme la durée du test qu'un élève peut entreprendre est limitée, il y a une tendance à privilégier plus de questions courtes qui peuvent être réparties sur plusieurs contextes, et plus de questions fermées à choix multiples plutôt qu'une réponse formulée dans leurs propres mots. Ces choix ont l'avantage supplémentaire de pouvoir être corrigés par des machines.

Il est important de noter que la restriction sur la durée et donc sur la quantité de contextes est plus sévère lorsque tous les élèves reçoivent les mêmes questions. Le problème n'est pas aussi grave dans le cas d'une enquête de population où un grand nombre de questions peuvent être utilisées et divisées en un certain nombre de sous-tests pour aboutir à un échantillon aléatoire de la population. C'est la conception utilisée dans les enquêtes nationales telles que le NAEP, l'APU et les enquêtes internationales comme le TIMSS et PISA. L'objectif de ces enquêtes est de

déterminer les performances au niveau national ou régional, qui sont dérivées de la combinaison des résultats des sous-tests. Les résultats des élèves pris individuellement, ou même de toute une école, ont peu de valeur en termes d'objectifs globaux à évaluer, mais lorsqu'ils sont combinés avec l'ensemble des échantillons, le résultat montre une bien meilleure image de la performance nationale que l'attribution des mêmes questions à tous les élèves. Lorsque les différentes questions sont données à des échantillons d'élèves différents, il n'est plus nécessaire d'accumuler dans un petit test des questions courtes et donc plus de temps peut être accordé aux élèves pour les laisser s'engager dans un contexte particulier. Ce n'est pas un hasard si les exemples cités ci-contre ont été créés et utilisés dans les programmes de l'APU et du PISA.

91 Noble et al (2012) "I never thought of it as freezing": How Students Answer Questions on large-scale science tests and what they know about science, *Journal of Research in Science Teaching*, 49 (6) 778–803.

L'encadré 18 énumère un certain nombre de sources de questions de performance écrites et publiées qui ont été utilisées dans des tests scientifiques et des enquêtes. Ils incluent, mais ne sont pas limités aux questions qui évaluent l'application des connaissances, la compréhension et la mise en œuvre de compétences d'investigation, pertinentes pour évaluer les résultats de l'ESFI.

#### *Encadré 18: Les sources de questions et de tâches utilisées dans des tests d'évaluation*

##### **SEAR (Science Education Assessment Resources)<sup>92</sup>**

Il comprend une banque de ressources d'évaluation des élèves qui couvre l'ensemble des âges de la scolarité de l'année 1 à l'année 12. Les tâches sont classées et il est possible d'y accéder selon le niveau de connaissance scientifique, l'objectif d'évaluation, le type de tâche et l'objet du résultat d'apprentissage. Six niveaux de connaissance scientifique sont définis par une « carte de progression des connaissances scientifiques » associée à PISA. La banque inclut des tâches ciblant les objectifs diagnostiques, formatifs et sommatifs : elles sont dans des formats à choix multiples, ouverts et pratiques. Il existe des systèmes de notation complets pour accompagner chaque question.

##### **Les questions scientifiques publiées par PISA<sup>93</sup> (Programme for International Student Assessment)**

Il s'agit d'une collection de questions qui ont été utilisées dans les enquêtes d'élèves âgés de 15 ans et publiées dans divers rapports. Chaque question comprend des supports stimulants et un système de notation.

##### **Les questions scientifiques d'APU (Assessment of Performance Unit)**

Des exemples de questions et de tâches écrites et pratiques, utilisées dans l'enquête des élèves de 11, 13, et 15 ans en Angleterre et au Pays de Galles dans les années 1980, sont disponibles dans les rapports écrits pour les enseignants.<sup>94</sup> Le système de notation et une description de réponses des élèves sont fournis pour chaque question.

##### **NAEP (National Assessment of Educational Progress – the Nation's Report Card) les questions scientifiques de l'échantillon 2011<sup>95</sup>**

Il fournit toutes les questions publiées dans l'évaluation scientifique des élèves en 2011 pour les grades 4, 8 et 12. Les questions incluent les réponses à choix multiple, les réponses courtes et longues. Pour chaque question, il y a les aides à notation, les réponses des élèves et les données de performance.

## **Questions de performance**

Les deux points faibles des tests écrits tiennent à ce qu'ils dépendent de la lecture et de l'écriture d'une part, de la nécessité de s'engager dans le problème posé en se limitant au contexte unique sur la feuille d'autre part. Ces lacunes peuvent être évitées si l'élève est capable de procéder à une investigation avec des objets et des équipements réels. Il existe bien sûr d'autres restrictions. La question doit être présentée à l'élève de façon à ce que ce dernier la considère ensuite comme sa propre question. De plus, la situation du test est loin d'être celle que l'on trouve dans une classe normale, puisqu'il se peut que les élèves travaillent seuls (parfois par deux) avec un administrateur présent pour observer leurs actions. Cependant, le test donne l'occasion aux élèves d'explorer, d'essayer et de recommencer si nécessaire. L'exemple 5 montre une investigation sur les serviettes de papier qui a été utilisée dans les enquêtes d'APU en Angleterre auprès d'élèves âgés de 11, 13 et 15 ans, dans les années 1980.

92 Australian Government Department Education, Employment and Workplace Relations. *Science Education Assessment Resources (SEAR)* <http://cms.curriculum.edu.au/SEAR>

93 OECD (2006) PISA released items: Science. Paris: OECD <http://www.oecd.org/pisa/38709385.pdf>

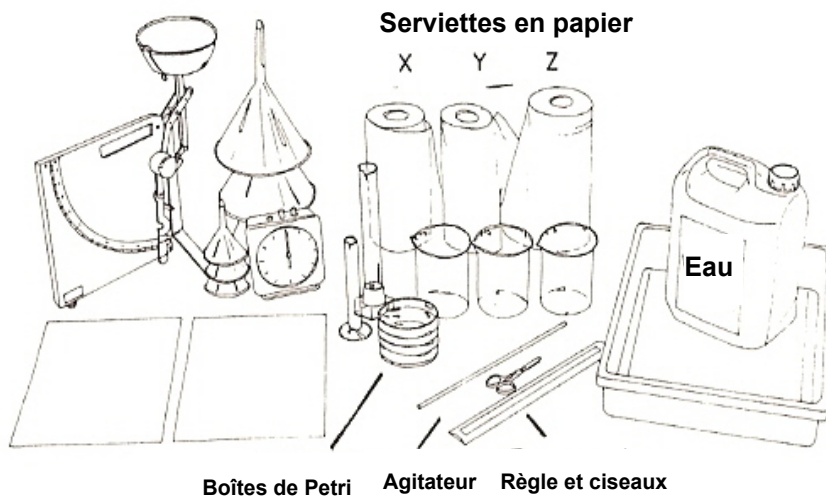
94 <http://www.nationalstemcentre.org.uk/elibrary/collection/727/assessment-of-performance-unit-science-reports-for-teachers>

95 US Department of Education [http://nationsreportcard.gov/science\\_2011/sample\\_quest.asp](http://nationsreportcard.gov/science_2011/sample_quest.asp)

*Exemple 5 : Les serviettes en papier (APU, rapport pour les enseignants en science no 6, 1985)<sup>96</sup>*  
*Commentaire*

Vous avez en face de vous trois types de serviettes en papier appelés X, Y et Z.  
 Voici ce que vous devez trouver :

Quel type de papier peut  
 contenir le plus d'eau ?



Vous pouvez utiliser l'un des objets situé devant vous.

Choisissez ce dont vous avez besoin pour répondre à la question.

Indiquer clairement vos résultats afin qu'une autre personne puisse comprendre ce que vous avez découvert.

Cette tâche gérée par un examinateur qualifié s'adresse à chaque élève pris individuellement. L'équipement est fourni avec le test et mis sur une table avant que l'élève entre dans la pièce. L'élève est invité à aller examiner l'équipement. L'examinateur lui remet une feuille comportant la question, des espaces pour écrire et pour noter des résultats. L'examinateur doit également vérifier que l'élève a bien compris la question. Pendant l'investigation, l'examinateur effectue une check-list détaillée et n'intervient que si la sécurité est menacée. Lorsque l'élève a terminé la tâche pratique, l'examinateur doit vérifier que le contenu écrit est compréhensible et que la réponse comporte des éléments de preuves. Le test se termine par une question adressée à l'élève : quels changements, le cas échéant, feriez-vous si vous pouviez recommencer l'exercice ?

96 Welford, G., Harlen, W. and Schofield, B. (1985) *Practical Testing at ages 11, 13, and 15*. London: DES, WO and DENI

Réduire l'erreur due à certains facteurs peut augmenter d'autres effets, en particulier celui du contexte. Lors d'un test d'une durée raisonnable, chaque élève ne peut entreprendre qu'un petit nombre (deux ou trois au plus) d'investigations. Même dans une enquête nationale, il y a une limite au nombre utilisable et l'impact du contexte met en échec toute tentative d'obtenir un score sur la « performance d'investigation » issus des performances de plusieurs investigations. Les enquêtes fournissent des preuves montrant que les élèves qui obtiennent de bons résultats dans une investigation n'auront pas nécessairement le même résultat dans un autre test sur les mêmes compétences mais dans un contexte différent.

Cela a été confirmé dans une étude mûrement réfléchie aux Etats-Unis. Pine et al (2006) ont évalué des élèves de cinquième grade (10-11 ans) avec plusieurs tâches pratiques, y compris celle sur les « serviettes en papier » et une autre sur la longueur d'un ressort avec des poids différents suspendus, appelé le « ressort ». Ils n'ont constaté « aucune corrélation dans les scores de chaque élève ». Les élèves, ayant soit un 9 ou un 1 au test sur « le ressort », ont eu des scores allant de 1 à 9 au test sur les serviettes en papier.<sup>97</sup> Puisqu'une tâche particulière est l'une parmi la vaste gamme de tâches possibles, « l'erreur d'échantillonnage » est importante, et cela signifie que pour obtenir un score fiable d'un élève travaillant seul, ce dernier devrait entreprendre un nombre de tâches totalement inacceptable. La seule solution significative pour exprimer la performance est de décrire chaque investigation et l'ensemble des différentes manières dont chaque élève a répondu.

## Les alternatives aux tests

Il ressort de la discussion précédente que les tentatives faites pour fournir une évaluation fiable, portant sur la compréhension et les compétences d'investigation, à l'aide de tests se heurtent à des obstacles impossibles à contourner. Nous avons déjà mentionné quelques-unes des exigences fondamentales pour faire une évaluation et des connaissances scientifiques et des compétences en investigation :

- Les élèves sont réellement impliqués dans la réalisation de l'investigation ;
- Ils sont engagés par une question qui est nouvelle pour eux;
- La connaissance, à coup sûr nécessaire, n'est pas l'objet de l'évaluation, mais est à leur disposition.

Lorsqu'une tâche est présentée aux élèves dans un contexte défini et limité par un format spécifique comme dans un test, il y a nécessairement des incertitudes pour répondre aux exigences ci-dessus. Une véritable investigation a lieu lorsque les élèves cherchent à répondre à une nouvelle question dont ils ne connaissent pas la réponse. Mais qui peut juger ce qui est « nouveau » pour un élève en particulier ? La réponse d'un élève, formulée en dehors du contexte normal de l'apprentissage et de l'interaction avec les autres, peut-elle vraiment refléter leur capacité ?

Les recherches menées au Danemark<sup>98</sup> sur des questions de tests de PISA-sciences apportent une réponse claire à cette dernière question. Il a été demandé aux élèves de répondre à quelques questions de PISA à l'oral lors d'un entretien, puis d'effectuer par paire une investigation décrite dans cette question de PISA. La conclusion a montré que « par comparaison directe en suivant les critères de notation du PISA, la performance des élèves a augmenté de 25% lorsqu'il leur a été demandé de mettre en œuvre leurs connaissances dans un format de test différent, orienté dans un contexte socio-culturel qui leur soit familier ».

97 Pine, J., Aschbacher, P., Rother, E., Jones, M., McPhee, C., Martin, C., Phelps, S., Kyle, T. and Foley, B. (2006) Fifth graders' science inquiry abilities: a comparative study of students in hands-on and textbook curricula, *Journal of Research in Science Teaching* 43 (5): 467-484. P 480

98 Dolin, J., & Krogh, L. B. (2010): The Relevance and Consequences of Pisa Science in a Danish Context. *International Journal of Science and Mathematics Education*, 8, 565-592.

Outre les interrogations concernant les tests utilisés pour l'obtention de décisions et jugements importants (par exemple passage dans la classe supérieure), telles que mentionnées au chapitre 4, bien d'autres problèmes soulignent les déficiences de l'évaluation par tests.

Existe-t-il alors des alternatives à l'emploi de tests pour faire une évaluation sommative des élèves ? Oui, heureusement : les expériences dont les élèves ont besoin pour développer les compétences souhaitées, la compréhension et les attitudes permettent également d'évaluer leurs progrès. Le facteur fondamental est le jugement de l'enseignant. L'évaluation des élèves réalisée par les enseignants peut utiliser des preuves provenant d'activités régulières, complétées si nécessaire par des tâches spécialement conçues pour donner aux élèves l'occasion d'utiliser les compétences et la compréhension en question.

Pendant une période définie, un semestre, où la performance est prise en compte, les élèves ont la possibilité de s'engager dans un certain nombre d'activités dans lesquelles un ensemble de compétences peuvent être développées. Ces activités offrent aux enseignants des possibilités d'évaluer ce développement. En d'autres termes, les limites de ce que prouve un test ne s'appliquent pas lorsque l'évaluation des élèves est le fait de l'enseignant.

D'autres avantages vont au-delà de cette évaluation, plus valide, de la compréhension et des compétences d'investigation des élèves, car un plus grand éventail de compétences peut alors être inclus. L'observation des élèves pendant leur travail régulier permet de recueillir les informations sur le processus d'apprentissage plutôt qu'uniquement sur ce qu'il produit. Même les différents types de questions de l'exemple 5 ne peuvent pas évaluer les qualités comme la réflexion lors du processus d'apprentissage. Or ces informations sont utiles pour la sélection des élèves, soit dans les cours avancés de formation professionnelle, soit dans les études académiques, car le fait que les élèves aient appris à apprendre et qu'ils s'engagent vers un approfondissement de ce qu'ils savent est aussi important que ce qu'ils ont déjà appris.

## La mise en œuvre d'une évaluation par les enseignants

L'évaluation sommative des élèves par les enseignants consiste à décider, collecter, interpréter et communiquer des preuves afin d'apporter une synthèse des résultats des élèves. Tout comme il existe différents types de tâches et de moyens d'y répondre dans le cas des tests, il y a donc différentes méthodes d'évaluation des élèves par les enseignants en fonction de comment, quand et où les preuves sont collectées, interprétées et utilisées. Avant d'entrer dans les détails, il convient de clarifier ceci : l'évaluation des élèves par les enseignants n'est pas simplement l'utilisation de leurs jugements personnels sur le choix et l'interprétation des éléments de preuve à utiliser. L'évaluation sommative des élèves par les enseignants utilise des procédures convenues et est soumise à des mesures de contrôle de la qualité, appropriées pour l'utilisation des résultats, c'est-à-dire un contrôle plus strict pour des enjeux plus importants.

Le processus nécessite certains *résultats ou données*, des *critères* permettant de juger et des *procédures* pour parvenir à un jugement.

Les *preuves* ou les *données* relatives à la compréhension et aux compétences d'investigation scientifiques peuvent provenir de :

- l'observation des élèves impliqués dans l'investigation ;
- un portfolio de travaux collectés pendant une période définie, comprenant des explications, des réflexions, des photographies et d'autres produits de l'investigation ;
- les cahiers ou les présentations sous forme numérique des élèves ;
- les exposés faits par les élèves individuellement ou en groupe.



Les critères varient selon le degré de précision. Par exemple, un critère hautement précis serait « *l'élève sait que les aimants attirent certains matériaux et pas d'autres, et qu'ils peuvent se repousser les uns et les autres* ». Cela ne peut concerner que des travaux sur les aimants. Alors que : « *l'élève peut grouper les matériaux en fonction de leurs propriétés physiques* » peut concerner des informations issues de diverses tâches d'apprentissage».

Pour atteindre ce niveau de précision des critères les enseignants sont tenus, pour élaborer l'évaluation de leurs élèves de tenir compte de certains documents. Cela se produit dans certaines formes d'évaluation du portfolio des élèves, où les tâches contenues dans celui-ci sont directement prescrites avec les critères correspondants. Cependant, plus les tâches sont précises, moins il y a de chance de collecter des résultats à partir d'un large éventail d'activités, ce qui est pourtant l'objectif principal de l'évaluation des élèves par les enseignants. Une analyse de la recherche sur la validité et la fiabilité de l'évaluation des élèves par les enseignants<sup>99</sup> a conclu que les modalités les plus fiables étaient celles où les critères étaient détaillés mais génériques et applicables à une série d'activités en classe. Les études indiquent que les critères bien spécifiques aident les enseignants à formuler des jugements fiables. Les critères les plus efficaces guident le choix des épreuves sans les prescrire en détail.

Les *procédures* pour émettre des jugements peuvent partir d'un choix de critères, suivi de la recherche d'épreuves ou bien d'un choix d'épreuves, suivi de leur mise en relation avec des critères qui confirment la qualité du travail. En général, il y a toujours une combinaison de ces modalités avec un va-et-vient entre les données et les critères afin de trouver le « meilleur ajustement ». Plutôt que d'essayer de faire correspondre une tâche donnée à un critère spécifique, les enseignants recueillent les résultats à partir de plusieurs activités pertinentes pour aboutir à un jugement « équilibré » par rapport aux critères qui correspondent le mieux aux épreuves. Les critères servent également à attirer l'attention sur les résultats de certains types de tâches, pour que les enseignants, attentifs à la recherche de certains comportements, aient moins de chance de les manquer.

Il est fréquent d'assigner les critères à des « niveaux » différents, comme indiqués dans l'encadré 19 (page 75), afin que le résultat de l'évaluation des élèves puisse être exprimé par rapport au niveau présent de l'élève. Dans un domaine donné de l'apprentissage, les niveaux sont déterminés en mesurant la progression des élèves, en s'appuyant sur les résultats de recherches existantes, sur les informations provenant de l'expérience des enseignants ainsi que de quelques essais et erreurs. Il existe des exemples comme les *Developmental Assessment Materials*, élaborés en Australie<sup>100</sup>, et les descriptions de la performance à différents niveaux sous forme d'un certain nombre de normes et objectifs de l'évaluation nationale des élèves, tels que le National Curriculum de l'Angleterre (Encadré 19). Cependant, l'expérience de l'utilisation (ou, plus exactement, d'une mauvaise utilisation) des « niveaux » a provoqué des réactions négatives quant à leur application dans la mesure de la performance des élèves. Nous allons examiner plus loin les avantages et les inconvénients de l'utilisation de niveaux (page 80), car les arguments s'appliquent également à l'utilisation des tests. Concentrons-nous maintenant sur les solutions possibles, afin d'améliorer la fiabilité de l'évaluation des élèves faite par les enseignants.

99 Harlen, W. (2004) Trusting teachers' judgements: research evidence of the reliability and validity of teachers' assessment for summative purposes, *Research Papers in Education*, 20(3); 245-270.

100 Masters, G. and Forster, M. (1996) *Progress Maps*. Camberwell, Victoria, Australia: ACER

## Amélioration de la fiabilité d'une évaluation réalisée par les enseignants

La critique d'une telle évaluation la plus souvent citée porte sur la fiabilité des résultats. Si aucune mesure n'est prise pour assurer la qualité, alors les jugements des enseignants sont susceptibles de contenir un certain nombre d'erreurs. Les études ont montré des preuves de partialité, des sources d'erreur non aléatoires et dues à la prise en compte d'informations portant sur les aspects non pertinents du comportement des élèves, ou encore des influences du genre, des besoins éducatifs spéciaux, ou de la capacité générale ou verbale d'un élève lorsqu'on juge la performance dans une tâche spécifique. Par exemple, le comportement général de l'élève influe sur le jugement porté par l'enseignant, concernant en particulier les résultats de jeunes élèves. Comme le comportement en classe est souvent associé au genre dans les premières années de l'école (les petites filles ont tendance à mieux se comporter que les petits garçons), le résultat est parfois influencé par des préjugés sexistes. L'effet est beaucoup moins important pour les élèves plus âgés. Les autres causes du manque de fiabilité découlent de la prise en compte d'informations non pertinentes ici (comme la propreté ou l'orthographe lorsque ce ne sont pas les objectifs ciblés de la tâche), de variations dans l'interprétation des critères, enfin de l'association de performances dans des contextes spécifiques à des critères beaucoup plus généraux. Cependant, il existe plusieurs moyens efficaces pour rendre la fiabilité égale, voire supérieure à celle des tests. Les principales solutions, que nous examinons maintenant, font appel à une modération en groupe, à l'utilisation d'exemples, d'un test ou d'une tâche de temps à autre pour vérification.

### Modération en groupe

Cette modération suppose une réunion entre les enseignants pour examiner des échantillons du travail des élèves. L'objectif n'est pas de vérifier des décisions prises sur un travail en particulier, mais plutôt de parvenir à une compréhension commune des critères et de leur application. L'intention est d'influencer le processus d'évaluation des élèves et ainsi d'assurer une plus grande fiabilité des résultats pour tous les élèves. Les enseignants apportent à la réunion plusieurs exemples de travaux d'un élève, puisque les jugements ne doivent pas être fondés sur un seul travail. Ils décrivent le contexte et discutent le fondement des jugements. La modération en groupe présente des avantages au-delà de l'amélioration de la qualité des résultats de l'évaluation des élèves. Elle exerce également et pleinement sa fonction de développement professionnel. La participation à la réunion pour discuter des conclusions tirées de l'observation des élèves et de l'étude de leurs travaux apporte aux enseignants des indications sur le processus d'évaluation des élèves et améliore non seulement l'évaluation sommative des élèves, mais aussi la pratique de l'évaluation formative.

### L'utilisation d'exemples

La présentation d'exemples de travaux d'élèves et une analyse de certains aspects des critères d'évaluation des élèves contribue à éclaircir l'usage pratique de ces critères. Les exemples pertinents indiquent aussi les situations qui permettent aux élèves de montrer leurs compétences et leur compréhension. Bien que le résultat de l'évaluation soit identifié dans ces exemples, l'accent est mis sur le processus permettant d'arriver à ce résultat. De nombreuses sources d'exemples de travaux d'élèves ou des descriptions de leurs actions sont annotées pour mettre en évidence les caractéristiques essentielles d'un jugement. Certains matériaux scolaires contiennent des exemples d'actions, de mots, de discours, d'écriture ou de dessins d'élèves et discutent des aspects qui permettent de décider si certains critères fondamentaux ont été respectés. Par exemple, les guides pour les enseignants *Nuffield Primary Science* (1995) comprennent chacun un chapitre sur l'évaluation des élèves, où les travaux des élèves sont reproduits avec un commentaire sur les aspects qui ont conduit à un jugement du niveau atteint. L'exemple 6 est tiré du Guide des enseignants sur les *Matériaux* pour les élèves âgés de 8 à 11 ans, où un élève décrit l'investigation faite par son groupe, portant sur la dureté des différents métaux. (Ce travail est utilisé ici à titre d'exemple car dans la pratique, le niveau de performance ne doit pas être évalué à partir d'un seul travail).

Exemple 6: Rapport de Matthew, décrivant l'investigation faite par son groupe (enfants de 10 ans) et portant sur la force des métaux.

Experiments  
Bending Metal

We are trying finding out how strong some metals are

What we did

We did the experiment with some different metals which are silver-Nickel, aluminium, zinc, brass, mild steel, stainless steel,  
We bent them with a pair of pliers untill they snapped and here are the results.

Results

Nickle	0 too hard to snap
aluminium	29 and snapped
zinc	13 and snapped
Brass	57 and snapped
Mild steel	60 and snapped
steel	56 and snapped

Why we think this happened

We found out that mild steel was the strongest, and the easiest to snap was zinc, next was aluminium. We couldn't snap the Nickel because it was too small, (a disc)

Les critères utilisés dans l'évaluation des élèves (encadré 19) viennent des objectifs ciblés de l'évaluation nationale des élèves en Angleterre, portant sur l'investigation en science.

Voici le commentaire de l'exemple 6 :

L'investigation du groupe de Matthew découle de l'affirmation selon laquelle les métaux sont solides. L'enseignant les a encouragés à aller plus loin en cherchant un moyen de voir si les différents métaux étaient aussi solides les uns que les autres. Leur investigation consistait à comparer la facilité avec laquelle on peut casser certains métaux en les pliant plusieurs fois vers l'avant et l'arrière à l'aide d'une pince. Ils ont quantifié leurs observations en comptant (en dénombrant), mais n'ont pas du tout mentionné la nécessité de contrôler le degré de flexion dans chaque cas, ni la taille de chaque morceau de métal. Cependant, ils ont fait leurs observations de manière systématique à l'aide d'un tableau et ils ont interprété leur résultat : ce qui les rapproche du niveau 3. Il faudra faire plus attention à la justesse de leurs expériences pour en déduire une réussite au niveau 3.<sup>101</sup>

101 Nuffield Primary Science Teachers' Guide *Materials*. (1995) London: Collins Educational. p 86

**Encadré 19: Exemple de critères d'évaluation des élèves en investigation scientifique**

(Note : un niveau s'étend sur environ deux ans, le niveau 2 correspond au niveau attendu pour les élèves âgés de 7 ans)

**Niveau 2:** Faire des suggestions et répondre aux suggestions des autres sur les façons de trouver des réponses ou de comparer des matériaux. Utiliser du matériel pour faire des observations. Noter ce qu'ils trouvent et comparer leurs observations avec le résultat attendu.

**Niveau 3:** Raconter le résultat attendu et suggérer des moyens de recueillir des informations pour tester leur prédiction. Réaliser des tests pertinents en sachant pourquoi ils sont pertinents. Noter ce qu'ils ont trouvé de différentes façons. Remarquer les régularités qui se dégagent des résultats.

**Niveau 4:** Faire des prédictions qui guident la planification de l'investigation. Utiliser un équipement approprié et faire des observations adéquates et pertinentes. Interpréter, tirer des conclusions et tenter de relier les résultats aux connaissances scientifiques.

**Niveau 5:** Planifier des investigations de prédictions contrôlées et fondées sur des connaissances scientifiques. Utiliser des graphiques, des schémas ou des tableaux pour noter et soutenir l'interprétation. Examiner les constatations en fonction des connaissances scientifiques.

Les enseignants devant fonder leurs décisions sur une série de travaux de chaque élève afin de ne pas juger uniquement d'un seul travail, il est donc très utile d'avoir un modèle sous forme d'un portfolio d'élève plutôt que des morceaux de travail de plusieurs élèves. Cela permet aux enseignants d'appliquer les critères de manière plus globale. Par exemple, un coup d'œil sur les descriptions des niveaux de compétences d'investigation dans l'encadré 19 montre qu'un travail individuel peut ne pas correspondre aux descriptions et le portfolio peut ne pas remplir tous les critères d'un niveau. L'approche préconisée est celle du « meilleur ajustement », en comparant les preuves à celles correspondant au niveau le plus probable, à celui juste en-dessous et à celui juste au-dessus.

Des exemples peuvent être utilisés lors de la modération en groupe, mais ils sont particulièrement utiles pour les enseignants isolés qui ne peuvent pas participer à des réunions de modération en groupe.

**Utiliser un test ou une tâche**

Dans cette modalité, un test court ou une tâche particulière est utilisé comme un moyen de modération de vérification des jugements des enseignants, mais non comme une mesure distincte de la performance. Les tests sont utilisés de cette manière en Ecosse pour évaluer l'anglais et les mathématiques. Les enseignants décident sur la base d'une série d'épreuves provenant des activités quotidiennes menées au fil du temps pour savoir si un élève a rempli les critères d'un sujet à un niveau particulier. Une option pour les enseignants afin de modérer ce jugement est d'utiliser un test court à un niveau indiqué par les propres évaluations des enseignants. Le test, élaboré et noté par les enseignants, est extrait d'une banque de questions conçues en externe. L'enseignant compare les résultats du test avec ceux de leur propre évaluation des élèves en classe. Le test peut être effectué chaque fois que l'enseignant le juge nécessaire et non à une période fixée. Il se fait individuellement et informellement pour éviter l'anxiété chez les élèves. Une fois confirmé par la modération, le niveau atteint selon le jugement de l'enseignant est enregistré puis communiqué au moment opportun.

## Quelques exemples d'évaluation des élèves reposant sur les enseignants

L'évaluation des élèves réalisée par les enseignants n'est bien sûr pas une nouveauté, il s'agit d'une pratique de routine dans la notation scolaire, où les résultats de cette évaluation n'ont pas de conséquences importantes; Dans ce cas, la modération requise est moins rigoureuse que lorsque des décisions importantes peuvent être prises sur la base des résultats. Un exemple d'évaluation des élèves par les enseignants dans le cas d'enjeux importants a déjà été mentionné au chapitre 3. Voici quelques points pertinents.

### *L'évaluation des élèves par les enseignants en fin du cycle secondaire*

Le fonctionnement du système d'évaluation des élèves fondé sur les enseignants, dans le cas du Senior Certificate au Queensland (Australie), est un exemple bien documenté de fonctionnement d'un tel système. Il existe depuis 1972 après l'abolition des examens externes au Queensland. Les explications en faveur de l'évaluation des élèves par les enseignants reflètent les arguments qu'une telle évaluation est en mesure d'inclure une série de résultats d'apprentissage à la fois académiques et professionnels pour soutenir l'acquisition plutôt que pour contrôler le programme scolaire. De plus, selon Maxwell:

Un principe important de l'évaluation des élèves fondée sur l'école est qu'elle doit être progressive et continue. L'un des objectifs ... est d'alléger la pression maximale résultant d'un seul examen final : le test unique dont tout dépend. Cela exige non seulement que l'évaluation des élèves soit adaptée à la mise en œuvre du programme scolaire, mais aussi qu'elle soit effectuée progressivement tout au long des études. En d'autres termes, la validité de l'évaluation des élèves est améliorée par l'assemblage d'un portfolio à partir de différents types d'évaluation des élèves et de contextes. Par conséquent, la fiabilité est aussi améliorée grâce à de nombreuses possibilités pour l'élève de manifester ses connaissances et ses capacités, puisque les informations ont été collectées à des occasions différentes.<sup>102</sup>

Le processus basé sur le portfolio permet une variation de contenu pour que les programmes puissent être mis en œuvre avec souplesse en répondant aux besoins locaux. L'élément commun est le système des critères progressifs en fonction desquels chaque portfolio est jugé. Il existe aussi un système solide de modération pour les sujets qui sont décisifs pour l'entrée à l'université. Cela assure avec succès la confiance de toutes les parties concernées par le résultat de l'évaluation.

Le portfolio est construit au cours de deux années d'études durant lesquelles son contenu va changer non seulement par l'ajout de nouveaux documents, mais aussi par le remplacement des travaux anciens par des résultats plus récents. Seulement la conclusion finale est prise en compte même si certains éléments ont été recueillis plus tôt que les autres. Ainsi, le dernier portfolio, le plus complet, est constitué de tâches d'évaluation des élèves qui représentent les informations *les plus complètes* et *les plus récentes* des connaissances et des compétences de l'élève. Selon Maxwell :

Les informations *les plus complètes* signifient que les données d'évaluation des élèves doivent être disponibles sur tous les aspects obligatoires du programme scolaire. Les critères importants ne peuvent pas être omis ; les preuves d'évaluation du portfolio des élèves doivent couvrir tous les points requis du programme ... Les informations *récentes* signifient que les évaluations antérieures des élèves qui ne sont plus pertinentes peuvent être rejetées et remplacées par des données plus récentes ... L'objectif ultime est de représenter l'état des connaissances et des capacités que l'élève a acquis généralement à la fin du cursus.<sup>103</sup>

102 Maxwell, G. (2004) op cit. p 2

103 Maxwell, G. (2004) op cit. p 4-5

Ces critères d'évaluation des élèves, publiés en vue d'informer les élèves, les parents ainsi que les enseignants, décrivent de façon qualitative ce que les élèves font. Par exemple, l'une des sous-catégories de « travailler de manière scientifique » est associée à la planification et ses critères sont énoncés en cinq niveaux ou normes de A à E dans l'encadré 20.

*Encadré 20: Les critères de notation des portfolios liés à la planification (Queensland Senior Certificate)*

**Norme A:** planifier une série d'investigations, portant sur des problèmes contenant des éléments nouveaux et/ou complexes ;

**Norme B:** planifier une série d'investigations, portant sur des problèmes contenant des éléments nouveaux et/ou complexes ;

**Norme C:** Planifier une série d'investigations portant sur des problèmes simples ;

**Norme D:** participer à la planification d'investigations portant sur des problèmes simples ;

**Norme E:** participer à certains aspects de la planification d'investigations portant sur des problèmes simples

Dans l'encadré 20, les critères des normes A et B sont identiques, mais le jugement sur la « planification » est seulement l'un des multiples aspects de « travailler de manière scientifique » qui doivent être jugés dans leur ensemble. La comparaison des résultats avec des critères implique des jugements et non pas d'une accumulation: il s'agit de formuler le jugement global le mieux adapté.

La modération comporte plusieurs étapes, commençant par l'approbation du plan de travail de l'école : les détails de la manière dont l'école a l'intention d'offrir des possibilités à l'élève de répondre aux critères définitifs de l'évaluation de ses élèves sur un sujet. La modération de ces sujets pour la sélection universitaire implique des comités externes qui examinent les échantillons de portfolio des écoles et considèrent les preuves proposées pour soutenir ou infirmer les jugements. Réciproquement, un jury d'État est présent pour examiner des échantillons des districts et arbitre les cas difficiles.

Ce processus ouvert et continu de création du portfolio signifie qu'à la fin du cours, il ne devrait y avoir aucune surprise pour les enseignants ni pour les élèves. De plus, la « mise à jour sélective » et la collecte de preuves « complètes et récentes » permettent de ne pas tenir compte d'un départ faible, de performances atypiques ou de confusions antérieures et temporaires (pour une raison quelconque). De manière importante, ces processus facilitent l'utilisation de l'évaluation pour soutenir l'apprentissage, car les élèves profitent des commentaires reçus sur leurs évaluations antérieures. Ils ont également la possibilité de s'auto-évaluer pour décider du moment où ils remplacent un travail antérieur dans leur portfolio.

L'expérience du Queensland plaide ainsi en faveur d'une modération en collaboration, non seulement pour assurer une évaluation fiable des élèves mais aussi pour exercer un perfectionnement professionnel des enseignants.

Le moyen le plus puissant de développer des compétences professionnelles en matière d'évaluation des élèves est la mise en place de conversations professionnelles régulières chez les enseignants concernant les résultats des élèves (conversations de modération). Le mieux est d'axer sur des exemples concrets de portfolios des élèves. L'examen des exemples réels et la discussion des conclusions tirées sur la performance des élèves, en référence aux normes explicites, améliorent le jugement des enseignants et renforcent leurs connaissances et leur expertise sur l'évaluation des élèves avec plus de succès que n'importe quel autre processus.<sup>104</sup>

104 Maxwell, G. (2004) op cit. p 7

*L'évaluation des élèves fondée sur les enseignants, en fin de primaire, en Angleterre*

Un exemple d'évaluation des jeunes élèves par les enseignants est fourni par l'introduction en Angleterre en 1989, dans le cadre de l'évaluation nationale des élèves de 7, 11, 14 et 16 ans, en anglais, en mathématiques et en sciences. En plus de leur propre évaluation sur les niveaux de performance de leurs élèves, les enseignants ont été tenus de faire passer des tests conçus en externe. Après l'introduction de ce système<sup>105</sup>, il a été reconnu que les tests externes ne couvraient qu'une partie du programme scolaire ; l'évaluation des élèves par les enseignants devait donc être le composant principal (cette reconnaissance n'a pas duré, pour des raisons décrites au chapitre 4). L'établissement de rapports officiels d'évaluation par rapport aux niveaux du programme national était nouveau et a attiré l'attention des chercheurs sur la situation des enseignants du primaire. Gipps et al<sup>106</sup> ont observé et interrogé les enseignants d'élèves âgés de 11 ans. Ils ont mis en œuvre plusieurs méthodes différentes, mais le point commun consistait en la collecte quotidienne de données à l'aide de techniques telles que la prise de recul, l'écoute, les questions ouvertes, l'observation, la prise de notes ou la mémorisation. La plupart des enseignants ne voulaient pas perturber leur enseignement habituel. Certains ont conservé des travaux significatifs de leurs élèves, et d'autres un enregistrement des notes attribuées aux travaux des élèves. Si les enseignants ont constaté qu'ils n'avaient pas suffisamment de données pour certains élèves par rapport aux objectifs, ils ciblaient alors ces élèves pour leur porter une attention particulière. À la fin de l'année, au moment d'attribuer un niveau à chaque élève, les enseignants utilisaient les travaux collectés et/ou leurs notes ainsi que des souvenirs pour décider du niveau de performance en faisant le « meilleur ajustement » par rapport aux critères de niveaux comme ceux de l'encadré 19.

*L'évaluation des compétences d'investigation des élèves par les enseignants dans les écoles primaires en France*

Un changement important dans le programme scolaire a été introduit en France en 2006 sous forme du *Socle Commun de connaissances et de compétences*.<sup>107</sup> Il prescrit les connaissances et les compétences que les élèves doivent acquérir en fin de scolarité obligatoire (16 ans) en demandant aux écoles de prendre la responsabilité d'y parvenir. Cela fut et reste un très grand changement vis à vis des responsabilités des écoles. Pour mettre en œuvre cette partie du nouveau programme de sciences, les enseignants des écoles primaires et secondaires ont été confrontés au grand défi de l'enseignement et de l'évaluation des compétences en investigation scientifique des élèves. Le ministère a fourni un livret pour chaque élève dans lequel les enseignants doivent noter, à la fin de chaque année, si l'élève a atteint les compétences et les connaissances requises sous forme d'un simple « oui » ou « non ». Pour soutenir les enseignants dans l'établissement des synthèses et l'évaluation des compétences en investigation des élèves, l'équipe du projet *La main à la pâte* (Lamap) a aidé les groupes locaux dans l'élaboration de tâches qui permettaient aux enseignants d'évaluer les compétences et les connaissances en investigation des élèves. Les tâches comportaient à la fois des questions fermées et ouvertes et une activité pratique, comme illustré dans l'exemple 7.

En outre, le ministère a fourni un guide d'observation et de questions aux enseignants pour les aider à effectuer l'évaluation des élèves à la fois de manière formative et sommative.<sup>108</sup> Quatre sources de données d'évaluation des élèves sont proposées : l'étude des cahiers des élèves ; l'observation pendant les activités de classe ; les explications écrites ou orales des élèves d'une investigation pratique comportant l'expérience, le résultat et des diagrammes ; un test standardisé. Des indications sont également données par rapport aux points importants lors de la correction des travaux écrits des élèves, de l'observation en classe et de l'écoute de leurs exposés.

105 DES/WO (1988) *National curriculum Task Group on Assessment and Testing: A Report*. London: HMSO

106 Gipps, C., McCallum, B. and Brown, M. (1996). Models of teacher assessment among primary school teachers in England, *The Curriculum Journal*, 7 (2) 167-183

107 [http://cache.media.eduscol.education.fr/file/socle\\_commun/00/0/socle-commun-decret\\_162000.pdf](http://cache.media.eduscol.education.fr/file/socle_commun/00/0/socle-commun-decret_162000.pdf)

108 [http://cache.media.eduscol.education.fr/file/socle\\_commun/99/7/Socle-Grilles-de-reference-palier2\\_166997.pdf](http://cache.media.eduscol.education.fr/file/socle_commun/99/7/Socle-Grilles-de-reference-palier2_166997.pdf), page 36

*Exemple 7: Une tâche de type La main à la pâte pour des élèves, proposée par l'académie de Toulouse<sup>109</sup> en fin de l'école primaire en France*

*Dans cette longue tâche, les élèves commencent par lire une page d'informations sur la couverture de glace au pôle Nord et pôle Sud et leur changement de l'hiver à l'été dans les deux régions. Ils répondent à des questions sur les informations fournies dont l'effet de la hausse de la température sur la glace aux pôles. Une enquête est alors proposée et illustrée par une photo illustrant deux verres d'eau. Un verre présente un glaçon flottant dans l'eau, et l'autre montre un cube de glace maintenu au-dessus de la surface de l'eau, grâce à un film plastique se trouvant au dessus du verre. Le niveau de l'eau des deux verres est identique. On demande aux élèves de prédire avec des justifications ce qui va se passer sur le glaçon et le niveau de l'eau des deux verres après un certain temps. Les élèves doivent mener cette enquête de façon concrète, mais au cas où cela n'est pas réalisable il y a aussi une autre photo qui montre le résultat après la fonte des glaçons. Les élèves doivent ensuite répondre à une série de questions pour évaluer la cohérence de leur prédiction et les preuves qu'ils fournissent dans leurs explications des observations. À la fin, une dernière question demande aux élèves de commenter une affirmation répandue, portant sur le réchauffement climatique : « la fonte des glaces polaires provoquera l'augmentation du niveau de la mer ». Une réponse écrite plus développée est demandée avec l'utilisation d'un vocabulaire approprié.*

### **Le double utilisation des données pour une évaluation formative et une évaluation sommative des élèves**

Dans les trois exemples donnés plus haut, on voit qu'il y a des possibilités pour que l'utilisation formative des données soit recueillie pour faire également une évaluation sommative. Cette double utilisation est plus consciente et délibérée dans l'exemple suivant (certes idéalisé) d'une série de leçons de sciences dans lesquelles les élèves de la huitième année (âgés de 13 ans) ont étudié le transfert d'énergie thermique à travers différents matériaux.

Les élèves font l'étude des propriétés isolantes de différents matériaux qui pourraient être utilisés pour fabriquer des manteaux. Ils disposent de récipients métalliques dans lesquels ils peuvent mettre de l'eau pour représenter un « corps » et des morceaux de tissu pour envelopper l'extérieur des récipients. Les thermomètres ont été placés dans l'eau pour mesurer les changements thermiques dans le « corps ». Mais il reste plusieurs choix à faire dans la mise en place de l'enquête. Quelle doit être la température de l'eau à l'intérieur du récipient? Y aura-t-il un résultat utile si l'expérience est réalisée à l'intérieur de la classe car il y fait chaud au lieu de la faire à l'extérieur? Comment faire en sorte que la comparaison soit juste? Certaines décisions exigent l'application de ce que les élèves connaissent de la conduction et d'autres moyens de transférer l'énergie thermique, alors que d'autres nécessitent la compréhension de ce que signifie "réaliser des comparaisons justes".

Pendant que les élèves organisaient l'enquête, l'enseignant a attiré leur attention sur la liste des éléments importants lors de la planification d'une enquête, liste qui avait été élaborée lors d'une discussion antérieure et rangée dans leurs dossiers. L'enseignant observait leurs actions, écoutait leur conversation pendant l'organisation et le lancement de leurs enquêtes, et demandait aux groupes de temps en temps des explications sur les raisons de certains procédés. Au moment de communiquer les conclusions, les enseignants ont encore d'autres occasions de recueillir des preuves intéressantes, afin de favoriser le développement de la compréhension et les compétences en investigation des élèves sur le transfert de l'énergie thermique.

<sup>109</sup> à cette adresse, vous trouverez tous le détail de ces activités : <http://www.fondation-lamap.org/fr/page/14209/des-exemples-d-valuation-adapt-s-des-s-quences-d-enseignement>



Ainsi, au cours de la leçon, l'enseignante répondait à ce qu'elle voyait et écoutait les questions pour provoquer la réflexion sur les décisions ; elle demandait les justifications des affirmations ou des actions, et comment certains procédés d'investigation pouvaient aider à atteindre les objectifs. Autrement dit, l'enseignante a utilisé des preuves d'une manière formative pour faciliter l'apprentissage et a également noté les besoins de certains élèves pour les suivre dans les futures leçons.

Ensuite, à la fin de l'année, lors de l'établissement des rapports de progrès sur les niveaux atteints, l'enseignante réexaminait les résultats provenant de tous les cours de sciences. Pour les objectifs conceptuels et les compétences en investigation, les résultats de différentes activités spécifiques devaient être réunies pour former un jugement global sur le résultat de chaque élève. En prévision de cette étape, l'enseignante a accordé plus de temps aux élèves pour examiner leurs dossiers et comparer leurs explications d'investigations récentes aux anciennes, en examinant la collecte et l'utilisation des données afin de sélectionner les meilleurs exemples de travail comportant ces caractéristiques. Au cours des heures de cours réservées à cet objectif, l'enseignante pouvait parler avec des élèves en particulier afin de s'assurer que les critères de sélection ont été compris et appliqués de manière appropriée pour examiner les preuves en fonction des critères de niveaux. Dans ce cas précis, il s'agissait des descriptions pour les niveaux quatre, cinq et six de l'« investigation scientifique » et des « processus physiques » dans le *English national curriculum*. Les résultats ont été examinés par le directeur du département et les résultats d'un échantillon de trois élèves ont été discutés lors d'une réunion de modération départementale.<sup>110</sup>

## Les rapports issus de l'évaluation sommative des élèves : le pour et le contre de l'utilisation de niveaux

Les niveaux, comme indiqué précédemment, mettent en place les normes (nationales) pour juger de la performance des élèves. Les normes incarnent une vision de progrès dans l'apprentissage des élèves en termes descriptifs, mais elles peuvent également être converties en une échelle dans laquelle les niveaux sont définis sous forme de points dans le progrès global. Les niveaux représentent donc un paramètre condensé permettant de signaler les acquis et ont l'avantage de permettre l'analyse des résultats de groupes d'élèves. Les scores des tests peuvent également être convertis en niveaux. Ainsi en France, la Délégation de l'évaluation, de la prospective et de la performance utilise elle des échelles pour ses évaluations bilan<sup>111</sup> (N.d.T.).

### Les problèmes des niveaux

Cependant, l'utilisation de niveaux a été contestée au cours de la révision du programme scolaire et de l'évaluation des élèves en Angleterre, pour des raisons qui sont sans aucun doute pertinentes dans d'autres systèmes où le nombre d'élèves atteignant certains niveaux est devenue un enjeu élevé dans le jugement porté sur les enseignants et les écoles. Un rapport d'un groupe d'experts influents, en 2012,<sup>112</sup> a formulé de vives critiques sur l'utilisation de niveaux pour rendre compte des progrès des élèves, en particulier à l'école primaire et au collège, en considérant que les élèves sont inévitablement différenciés dans les dernières années de l'enseignement secondaire. Voici les principaux points qui peuvent entrer en résonance avec d'autres systèmes d'évaluation des élèves:

110 Harlen, W. (2007) Op cit p129

111 <http://www.education.gouv.fr/cid265/l-evaluation-globale-du-systeme-educatif.html#evaluations-bilans-nationales>

112 Department for Education, (2011). *The Framework for the National Curriculum. A report by the Expert Panel for the National Curriculum review*. (London: Department for Education).

- L'attribution de « niveaux » encourage un processus de différenciation des apprenants dans la mesure où les élèves vont s'approprier ces niveaux
- Certains élèves se préoccupent davantage de « leur niveau actuel » que de la substance de ce qu'ils savent, peuvent faire et comprennent.
- L'attribution de niveaux augmente la différenciation sociale au lieu de chercher à obtenir un apprentissage solide pour tous les élèves.
- La description d'un élève comme ayant atteint un certain niveau n'exprime rien sur ce que l'élève peut faire, ni ce qui est nécessaire pour progresser.
- Lorsque les niveaux sont considérés comme importants, les enseignants, les parents et les élèves les utilisent à mauvais escient pour classer les élèves.
- Les élèves qui sont considérés comme incapables d'atteindre les niveaux cibles ont souvent moins de possibilité de progrès, ce qui augmente l'écart de performance entre les élèves réussissant le mieux ou le moins bien.
- Comme les niveaux sont généralement bien espacés (chaque niveau étant d'environ deux ans d'intervalle), on a créé des sous-niveaux afin de pouvoir montrer les progrès accomplis. Toutefois, ceux-ci possèdent peu d'éléments de preuve dans le développement cognitif et servent uniquement à serrer le programme de plus près.

Globalement, les rapports de performance des élèves sous forme de « niveau » de réussite, que celui-ci soit évalué par des tests ou par un jugement des enseignants, sont reconnus pour avoir un impact profond sur la façon dont les enseignants, les parents et les élèves eux-mêmes jugent des progrès accomplis, avec des implications pour la motivation et l'apprentissage des élèves. Ce sont les conséquences involontaires d'un cadre trop normatif pour le programme scolaire et l'évaluation des élèves.

### **Une alternative aux niveaux**

Une alternative aux rapports de niveaux est d'identifier les résultats d'apprentissage à accomplir à la fin des étapes fondamentales de l'enseignement. Les étapes fondamentales pourraient s'étendre sur deux ou trois ans, donc les années de grade 1 à 12 peuvent être divisées en cinq ou six étapes. Les résultats d'apprentissage attendus pourraient bien être analogues aux indications de progression formulées dans le programme scolaire, mais en se concentrant sur quelques compétences clés. Les jugements porteraient sur l'acquisition de ces compétences clés dans le but explicite de garantir la réussite de tous les élèves à la fin de l'étape fondamentale. Les jugements des enseignants pendant l'étape fondamentale permettent d'identifier si les élèves sont en bonne voie pour atteindre les résultats attendus ou s'ils nécessitent une attention particulière. Ainsi, l'évaluation des élèves serait différente de la recherche du « meilleur ajustement » sur l'échelle de niveaux pour le suivi de l'accomplissement des compétences clés.

L'expérience de ces pays ayant des niveaux élevés de réussite et de petits écarts entre les élèves forts et faibles ne suggère pas que la conséquence de cette modalité soit le redoublement des élèves. Ce qui caractérise les systèmes performants semble être leur approche de la progression et de la différenciation des élèves. Au lieu de catégoriser grossièrement les résultats des élèves, ils encouragent tous les élèves à atteindre une compréhension adéquate avant de passer au sujet ou à l'étape suivante. Il n'y a aucune certitude que la capacité limite la réussite, ce qui risque de produire un effet négatif sur les attentes des élèves. Au contraire, on estime que l'engagement et de véritables efforts mènent tout le monde à la compréhension et la réussite.

## Les avantages et les inconvénients des tests et d'une évaluation faite par les enseignants

En résumé et pour conclure, nous rassemblons ici certains des avantages et des inconvénients de l'utilisation de tests ou d'évaluation des élèves par des enseignants à des fins sommatives. Les principaux facteurs concernent le compromis entre la validité et la fiabilité (voir le chapitre 1, page 11). Les tests n'apportent des informations que sur un échantillon d'objectifs mais des dispositions peuvent être prises pour garantir une fiabilité optimale. La lutte pour une grande fiabilité, cependant, conduit à l'inclusion des objectifs les plus faciles à tester de manière fiable, et elle a tendance à exclure des objectifs moins facilement testés comme les compétences et les pratiques de haut niveau. Dans le cas de l'évaluation des élèves par les enseignants, une gamme plus large de réussites et de résultats d'apprentissage peut être inclus, mais la fiabilité peut être faible sauf si des dispositions sont prises pour s'assurer que des normes comparables sont appliquées par tous. Enfin, les procédures de modération soutiennent le développement professionnel des enseignants.

Dans chaque cas particulier où une décision dans un sens ou l'autre doit être prise, quelques facteurs additionnels sont à considérer :

- Les tests apportent aux enseignants des exemples clairs d'objectifs d'apprentissage, mais ils orientent la pédagogie dans des directions spécifiques, identiques pour tous les élèves. L'évaluation des élèves par les enseignants permet une liberté plus grande de ces derniers afin de poursuivre des objectifs d'apprentissage d'une manière convenant à leurs élèves.
- Les tests conçus à l'extérieur de l'école permettent aux enseignants de distinguer leur rôle d'enseignant de celui d'évaluateur. La responsabilité de l'évaluation des élèves, combinant en une même personne le rôle d'enseignant et celui d'évaluateur, peut sembler augmenter la charge de travail des enseignants.
- Dans certains cas, les résultats des tests peuvent être utilisés pour apporter des commentaires qui soutiennent l'apprentissage, mais en général, ces possibilités d'utilisation formative sont limitées. Lorsque les enseignants rassemblent des preuves à partir des travaux en cours des élèves, celles-ci peuvent être utilisées de façon formative pour soutenir l'apprentissage et avoir également des objectifs sommatifs.
- Il est bien connu que les tests provoquent l'anxiété chez de nombreux élèves. Les élèves ne sont pas tous concernés et l'impact est souvent aggravé par la pratique fréquente de tests. L'évaluation des élèves par les enseignants réduit cette source d'erreur et d'inégalité.
- Le temps des enseignants consacré à la préparation des élèves aux tests peut être utilisé plus efficacement pour l'apprentissage lorsque l'évaluation des élèves est constante. Les ressources financières sont également libérées si on achète moins de tests commerciaux.
- Les utilisateurs des données d'évaluation des élèves ont souvent une plus grande confiance dans les tests que dans l'évaluation des élèves par les enseignants, en particulier pour les élèves plus âgés. Toute évolution nécessite donc une plus grande ouverture sur à la fois la précision des tests et les procédures qui peuvent améliorer l'évaluation des élèves par les enseignants.

## Chapitre 7

# Changer les pratiques de l'évaluation des élèves

L'interaction entre l'évaluation des élèves, le programme et la pédagogie, illustrée sur la figure 3 (page 31), montre qu'une évolution de l'évaluation des élèves entraînera des changements dans la pratique de l'enseignement. Il ne s'agira pas seulement de nouvelles procédures appliquées à des pratiques de classes inchangées, tout particulièrement dans le cas où l'évaluation repose davantage sur le jugement des enseignants plutôt que sur les tests et pratiques traditionnelles. Bien qu'aujourd'hui, au delà d'essais pilotes, les expériences portant sur une évaluation différente pour les élèves pratiquant l'ESFI soient encore limitées, il y a beaucoup à apprendre sur les changements dans les pratiques pédagogiques qui accompagnent la mise en place de l'ESFI.

Par conséquent, dans ce chapitre, nous examinons d'abord brièvement les leçons qui peuvent être tirées des tentatives de changements dans le contenu des programmes scolaires et la pédagogie, et de leur impact sur l'évaluation des élèves. Tout en notant que l'expérience ci-dessous se réfère à un contexte de développement professionnel d'enseignants en activité, ses modalités et ses principes sont tout aussi pertinents dans le cas de la formation initiale des enseignants.

### Les modalités du changement de pratiques dans l'éducation

Les différentes modalités du changement de pratiques dans l'éducation peuvent être divisées en deux groupes principaux : la transmission et la transformation.<sup>113</sup> Les diverses formes de *transmission* impliquent la distribution de ressources qui diffusent et publient des idées et des exemples de nouveaux contenus et de nouvelles pratiques. Les ressources produites associent souvent guides écrits et matériaux audio-visuels. Créées pour les nouveaux programmes, ces ressources promettent une simple solution en vue des changements nécessaires. Les enseignants n'ont plus qu'à suivre les guides, car la réflexion nécessaire aura déjà été faite pour eux. Cette modalité « descendante » est toutefois tombée en disgrâce car les messages reçus et réalisés en classe correspondent rarement aux intentions des producteurs.

La nécessité d'une médiation des messages par des séances de perfectionnement professionnel a conduit à une modalité « en cascade ». Elle débute par des séances de formation d'un groupe sur le sujet, les pratiques pédagogiques possibles et parfois une interaction avec la formation d'autres groupes. Ainsi formés, ces premiers stagiaires forment ensuite les autres et la formation passe donc de groupe en groupe, jusqu'à atteindre tous ceux qui ont une classe en charge. Cette modalité présente des inconvénients évidents : les messages présentent un grand risque de distorsion à chaque étape de la cascade. La modalité du « projet pilote et déploiement à grande échelle » est similaire, mais elle débute avec des expériences et l'évolution des idées dans un nombre limité d'écoles pour veiller aux aspects pratiques et assurer la crédibilité auprès des enseignants.

Le succès limité des méthodes de transmission a conduit à des alternatives *transformationnelles*. Les expériences de nombreux changements dans l'éducation, que ce soit dans le domaine des programmes, de la pédagogie, de l'évaluation des élèves ou de l'organisation et la gestion de l'école,

113 Hayward, L. (2010) Moving beyond the classroom, in J. Gardner et al *Developing Teacher Assessment*. Maidenhead, UK: Open University Press.

montrent qu'une participation à l'élaboration de nouvelles procédures ou de documents par ceux-là mêmes qui les mettent en pratique est un moyen plus efficace pour encourager le changement. Lorsque des groupes d'enseignants collaborent avec des chercheurs ou des concepteurs de tests, ils peuvent être créatifs et faire des essais dans un environnement sûr, apprendre les uns des autres, combiner des idées et s'approprier les pratiques émergentes dans une modalité « ascendante ». En amplifiant les occasions de réflexion et de développement de la compréhension des principes sous-tendant le changement, l'expérience peut alors être une forme plus efficace d'apprentissage professionnel. Mais l'exigence en ressources, et particulièrement en temps, est importante. Ce système peut difficilement être étendu à un grand nombre d'enseignants, mais il peut indiquer les types d'expériences susceptibles d'être efficaces dans les cas où le changement exige plus que l'adoption de nouvelles techniques.

Pour les modalités *transformationnelles*, le changement dans la pratique est une question d'apprentissage et un apprentissage efficace par les enseignants possède les mêmes qualités que dans le cas des élèves. Comme les élèves développent la compréhension à travers leur propre activité mentale et physique, les enseignants apprennent mieux lorsqu'ils prennent une part active dans la transformation de leurs pratiques. Hayward<sup>114</sup> cite Bruner<sup>115</sup> pour réclamer la présence de quatre composantes essentielles en vue d'un apprentissage efficace:

- L'action : les individus exercent davantage de contrôle sur leur propre activité d'apprentissage ;
- La réflexion : donner un sens à ce qui est appris, le comprendre et l'intérioriser ;
- La collaboration : en partageant des ressources issues d'une diversité d'êtres humains, impliqués dans l'enseignement et l'apprentissage ;
- La culture : la construction, par les individus et les groupes, de modes de vie spécifiques et de pensée, qu'ils désignent par *la réalité*.

Un autre facteur, révélé à l'évidence à partir de trois projets de développement à grande échelle dans des écoles secondaires danoises, est la mesure dans laquelle une innovation est perçue par les enseignants comme étant conforme à leurs croyances et leurs valeurs (souvent tacites) relatives à l'éducation. Lorsqu'il existe une convergence sur celles-ci, les enseignants adoptent plus facilement les nouvelles idées et effectuent les changements nécessaires dans leurs pratiques. En cas de conflit de valeurs, la participation des enseignants risque d'être faible et le projet aura donc peu d'influence sur leurs pratiques.

Dans les modalités de transformation, plutôt que de considérer toutes les classes comme identiques avec une solution identique pour toutes, l'hypothèse suggère que des environnements d'apprentissage différents nécessitent des solutions différentes. Bien sûr, le problème de couvrir un grand nombre d'enseignants persiste. Les possibilités d'impliquer chaque enseignant dans un développement authentique et d'adapter sur une grande échelle les expériences aux besoins individuels sont clairement limitées. Cependant, il est possible de créer des situations dans lesquelles les enseignants peuvent apprendre en collaboration et travailler sur la façon de mettre en pratique les nouvelles idées et les nouveaux objectifs dans le contexte spécifique de leur propre classe : nous en verrons quelques exemples ci-dessous. Pour les enseignants, l'occasion de visiter d'autres classes ou de montrer des vidéos de leur enseignement s'est avérée efficace dans l'amélioration de l'apprentissage des enseignants. Les projets européens tels que *Fibonacci*<sup>116</sup> et *Pollen*<sup>117</sup> ont montré ce qu'on peut faire lorsque le temps et les ressources sont utilisés pour réunir les enseignants et les formateurs d'enseignants en vue d'apprendre les uns des autres.

114 Hayward, L (2010) op cit p 96

115 Bruner, J. (1996) *The Culture of Education*. Cambridge, MA: Harvard University Press

116 [www.fibonacci-project.eu](http://www.fibonacci-project.eu)

117 [www.pollen-europa.net](http://www.pollen-europa.net)

## Quelques exemples de changement de pratiques dans l'évaluation des élèves

Les pratiques qui consistent à utiliser l'évaluation des élèves pour aider l'apprentissage et de l'évaluation sommative, dans laquelle les enseignants jouent un rôle prépondérant, exigent des stratégies qui diffèrent considérablement des pratiques courantes de nombreux enseignants. Les premiers travaux sur l'évolution des pratiques d'évaluation des élèves étaient axés sur l'évaluation formative des élèves à travers le travail de Black et Wiliam. Ce travail montra qu'une pratique de l'évaluation formative des élèves pouvait augmenter significativement leur niveau de réussite (voir chapitre 3). Pour beaucoup, ce fut une nouvelle façon d'évaluer les élèves que ceux-ci appréciaient. En même temps, il fut reconnu que les modalités existantes d'évaluation sommative des élèves ne fournissaient pas d'informations concordant avec les objectifs d'une éducation moderne, et qu'il fallait une plus grande implication des enseignants dans le processus pour améliorer ce point.

### Le développement des pratiques de l'évaluation formative des élèves par les enseignants : l'expérience en Angleterre

Le projet d'évaluation formative des élèves de Medway Oxford du King's College (KMOFAP) fut le précurseur de plusieurs projets de développement d'utilisation formative de l'évaluation des élèves.<sup>118</sup> Débuté peu après que Black et Wiliam aient terminé leurs recherches en classe, le KMOFAP a reconnu l'importance de trouver des solutions pour permettre aux enseignants d'intégrer l'évaluation formative des élèves dans leur travail. Les chercheurs de l'Université de King's College et des conseillers de deux autorités locales ont collaboré dans la planification du travail avec deux enseignants de mathématiques et deux enseignants de sciences venant de six écoles secondaires. Des financements ont permis aux enseignants d'être disponibles pour des séances de perfectionnement professionnel comprenant sept séances d'une journée entière, réparties sur dix-huit mois. Une fois présentées aux enseignants et aux conseillers les pratiques efficaces pour améliorer l'apprentissage, il est demandé aux enseignants, lors des premières séances, d'élaborer un plan d'action pour la mise en œuvre lors de l'année scolaire suivante. Ceci leur laisse du temps pour « expérimenter certaines stratégies et techniques proposées par les études tels que questionnement abondant, notation sous forme de commentaires, discussion des critères avec les élèves, auto-évaluation et évaluation entre pairs chez ces derniers ».<sup>119</sup>

Au cours du projet, les membres de l'équipe ont visité les classes des enseignants pour observer et associer des situations spécifiques aux plans d'action. Ces visites ont été décrites comme « non-directives, mais plutôt comme une façon d'offrir aux enseignants un miroir pour s'y regarder ».<sup>119</sup> Les enseignants ont été initiés à des stratégies générales et à une certaine idée de ce qu'il faut viser, sans toutefois se voir fournir un modèle indiquant ce qu'il faut faire. La contribution des chercheurs a donc été conçue pour amener les enseignants à identifier les moyens de mettre en pratique les différentes fonctions de l'évaluation formative des élèves. Ce rôle, inhabituel et mettant peut-être mal à l'aise les enseignants, était attribué à des chercheurs et à des conseillers que ces enseignants considéraient comme des experts. Les chercheurs ont noté que :

Dans un premier temps, les enseignants ne semblaient pas y croire. Ils pensaient que les chercheurs opéraient avec un modèle pervers d'un apprentissage par découverte, dans lequel ces chercheurs savaient très bien ce qu'ils voulaient faire faire aux enseignants, tout en ne leur en disant rien car ils voulaient que les enseignants « le découvrent eux-mêmes ». Cependant, après un certain temps, il s'est

118 This account draws on Harlen, W. (2010) On the relationship between assessment for formative and summative purposes, in Gardner, J. et al *Developing Teacher Assessment* Maidenhead, England: Open University Press pages 100 – 129.

119 Wiliam, D. Lee, C., Harrison, C. and Black, P. (2004) Teachers developing assessment for learning: impact on student achievement, *Assessment in Education*, 11 (1) 49-66, p 54.

avéré qu'il n'y avait pas de modèle prescrit d'actions à mettre en œuvre en classe, et que chaque enseignant devait trouver sa propre voie pour mettre en œuvre ces principes généraux dans sa classe.<sup>120</sup>

Au cours d'un essai pratique de mise en œuvre des activités, les enseignants ont pris conscience de la nécessité de comprendre pourquoi ces activités spécifiques étaient importantes et quel était leur mode de fonctionnement. Comme les enseignants utilisaient les activités et voyaient la réaction de leurs élèves, ils voulaient en savoir plus sur la manière dont ces derniers apprennent. Ainsi, environ un an après ce travail sur les actions en classe, l'une des sessions a été conçue pour présenter aux enseignants les théories de l'apprentissage en expliquant l'importance de la participation des élèves dans celui-ci.

Au sujet de la stratégie pour aller des résultats de recherches aux pratiques en classe, les chercheurs ont remarqué que les nouvelles pratiques développées par les enseignants étaient :

beaucoup plus variées et vastes que ce que nous étions en mesure de proposer au début du projet sur la base des travaux de recherches. Elles impliquent toutes des changements dans la façon de travailler avec les élèves et avec le programme.<sup>121</sup>

Au niveau des changements chez les enseignants, on pouvait s'attendre à observer des différences entre les 24 enseignants. Cependant, les chercheurs ont signalé des changements sur plusieurs points, observés chez tous les enseignants impliqués dans le projet, en particulier dans :

- Leur vision des objectifs : aider les élèves à apprendre plutôt que « couvrir le programme » ;
- Leurs attentes concernant les élèves : tout le monde peut apprendre avec le temps et une bonne méthode ;
- Un contrôle plus important laissé aux élèves, en partageant avec eux la responsabilité de l'apprentissage.

Black et ses collègues ont affirmé que les changements dans la pratique ont été lents à apparaître mais ont été durables, et n'auraient probablement pas eu lieu si les chercheurs avaient fourni des recettes de leçons réussies. Selon eux, « le changement des opinions et des valeurs est le résultat d'enseignants qui se considéraient eux-mêmes comme des apprenants en travaillant avec nous pour en savoir plus ». <sup>122</sup> Grâce à des comparaisons diverses des résultats des examens en classe, des examens scolaires internes et externes et des tests du programme national, Wiliam et al<sup>123</sup> ont remarqué un impact positif sur la réussite des élèves après seulement un an d'intervention.

### **Une amplification d'échelle du développement professionnel concernant l'évaluation formative : une expérience aux Etats-Unis**

De 2003 à 2006, Wiliam, de l'ETS aux Etats-Unis, et Leahy, ancien directeur d'école à la retraite en Angleterre, ont expérimenté les nouvelles méthodes pour atteindre des résultats semblables à ceux du projet KMOFAP, ici à une échelle permettant de couvrir un grand nombre de classes. Selon eux,

n'importe quel modèle de développement professionnel des enseignants, évolutif et efficace, a besoin d'un exercice d'équilibre délicat entre deux exigences contradictoires. La première est la nécessité de s'assurer la flexibilité du modèle afin qu'il puisse s'adapter aux circonstances locales de l'intervention, non seulement pour réussir, mais aussi pour capitaliser les potentialités présentes dans le contexte local et qui permettraient d'améliorer l'intervention. La seconde est de s'assurer d'une certaine rigidité du modèle,

120 Wiliam, D. et al (2004) op cit p 51

121 Black, P et al (2003) op cit p 57

122 Ibid, p 98.

123 Wiliam, D. et al (2004) op cit

afin que toute modification reste suffisamment fidèle à la conception originale pour garantir la qualité de l'intervention sans être l'objet d'une « mutation létale »<sup>124</sup>

Les chercheurs ont développé et expérimenté un certain nombre de modèles pour travailler avec des enseignants sur le rythme, la longueur, la durée et le contenu de leurs interventions. Par exemple, ils ont constaté que des intervalles de deux semaines entre les réunions étaient trop courts pour que les enseignants puissent planifier, mettre en œuvre des changements, et avoir des éléments à mettre dans leur rapport pour la prochaine réunion. Des réunions mensuelles étaient considérées comme optimales. Les réunions de deux heures ont été jugées trop longues dans certains cas, mais étant donné que les réunions d'une heure étaient trop courtes, il fallait un compromis entre les deux. Le nombre idéal de participants se situe entre huit et douze personnes dans un groupe composé d'enseignants de différentes matières. L'établissement d'un ordre du jour à chaque réunion a été adopté afin que les enseignants puissent être au courant du programme et de leur rôle dans la procédure.

Les idées issues de ces essais ont été utilisées pour développer des produits à distribuer aux écoles dans l'objectif d'aider les enseignants à développer la pratique d'une évaluation formative des élèves. Leahy et Wiliam ont développé deux ensembles de documents constitués de vidéos, de l'ordre du jour, de notes pour le leader du groupe.<sup>125</sup> Cependant, les chercheurs reconnaissent que l'existence de ces supports au développement professionnel ne garantit pas leur bonne utilisation, qui dépend de la volonté des écoles de mettre en œuvre une évaluation formative des élèves.

### La transformation des pratiques de l'évaluation des élèves en Ecosse

En Ecosse, le projet d'évaluation formative des élèves faisait partie d'un grand programme « Assessment is for Learning » (AifL) qui a examiné l'ensemble du système d'évaluation des élèves : les dossiers des élèves, la planification personnelle, la surveillance du système, l'évaluation systémique de l'école, ainsi que l'évaluation formative et l'évaluation sommative des élèves en classe. Le projet d'évaluation formative des élèves a utilisé une série d'activités de perfectionnement professionnel, notamment des ateliers locaux et des conférences nationales où les exposés ont été donnés par les enseignants déjà impliqués dans la mise en œuvre de l'évaluation formative des élèves en Angleterre et, plus tard, par les enseignants écossais. Il y avait également des discussions moins formelles, mais néanmoins organisées, entre des enseignants locaux et les agents du développement de l'Autorité de l'Éducation. Certaines références faisaient allusion aux publications du Assessment Reform Group.<sup>126</sup> Bien que les chercheurs invités soient inclus dans le programme de développement professionnel, leur participation était clairement décidée, conçue et contrôlée en Ecosse sans être une collaboration formelle avec les concepteurs de l'évaluation formative des élèves en Angleterre ou ailleurs.

Que l'ensemble des activités du projet soit conçu en fonction des besoins variés des différents enseignants ou pas, il s'est avéré positif dans la mise en œuvre d'une évaluation formative des élèves dans leur propre pratique. Voici les trois modalités principales :

- *Essai et ajustement* : les enseignants ont commencé par essayer des stratégies proposées dans les publications ou par des enseignants qui appliquaient déjà l'évaluation formative des élèves. Des stratégies ont été ajustées à la lumière de l'expérience acquise au cours des essais.
- *Examen critique et discussion avant l'essai* : des idées sur la mise en œuvre ont été revues avec des collègues pour augmenter la compréhension avant l'essai.

124 Leahy, S. and Wiliam, D. (2012) op cit, p 54-5

125 Leahy, S. and Wiliam, D. (2009) *Embedding Assessment for Learning – A Professional Development Pack*. London: Specialist Schools and Academies Trust.

Leahy, S. and Wiliam, D. (2010) *Embedding Assessment for Learning – Pack 2*. London: Specialist Schools and Academies Trust.

126 See [www.assessment-reform-group.org/publications](http://www.assessment-reform-group.org/publications)



- *A partir des objectifs et des idées* : les enseignants ont élaboré des solutions pour intégrer les idées de l'évaluation formative des élèves dans la pratique par leur expérience partagée, au lieu de faire comme le premier groupe impliqué dans la KMOFAP.

Même si l'impression générale était positive sur la réussite de la mise en œuvre du projet, les commentaires du rapport<sup>127</sup> mentionnent une réticence apparente de certains enseignants à travailler dans le cadre des théories de l'apprentissage, lesquelles expliquent les raisons pour lesquelles ces stratégies améliorent l'apprentissage. Il a également été souligné que très peu d'enseignants ont développé la participation des élèves aux choix concernant leurs objectifs d'apprentissage. Même si les élèves prenaient plus d'initiative dans la résolution de problèmes et ont semblé être plus réfléchis et plus clairs dans l'apprentissage, les enseignants ont maintenu une emprise sur les objectifs et les cibles des leçons. Ces deux questions étaient peut-être liées : les enseignants permettent aux élèves de prendre plus de responsabilité dans leur apprentissage pendant qu'ils développent en même temps la compréhension des raisons pour lesquelles l'évaluation formative des élèves fonctionne. Comme l'a montré le projet KMOFAP, ceci nécessite du temps.

### **L'importance de l'investigation dans le développement professionnel de l'évaluation formative des élèves**

À travers un certain nombre d'initiatives de recherche et de développement visant à aider les enseignants à élaborer des stratégies de l'évaluation formative des élèves dans divers pays, Pedder et James<sup>128</sup> ont identifié la collaboration entre professeurs, s'appliquant à un apprentissage en classe, comme le facteur le plus fréquemment associé à l'utilisation de certaines stratégies de l'évaluation formative des élèves. Un tel apprentissage collaboratif peut prendre différentes formes comme les discussions informelles entre les enseignants, les visites de classe des enseignants, la discussion de leurs observations, ou la forme plus formelle de Lesson Study développée au Japon. Ils ont également conclu qu'il fallait mettre plus d'accent sur les enseignants qui avaient les possibilités d'utiliser les résultats des recherches pertinentes et de mener des recherches dans leur pratique :

Si les enseignants sont préparés et prêts à s'engager dans un travail plein de risques, s'ils sont capables de problématiser leur propre pratique, de chercher des preuves pour évaluer afin de juger si le changement est nécessaire, puis d'agir sur leurs décisions, ils sont alors engagés dans une évaluation formative des élèves qui est en lien avec leur propre développement professionnel.<sup>129</sup>

### **Le développement de l'évaluation sommative des élèves par des enseignants du secondaire**

Dans l'objectif de trouver une modalité de l'évaluation sommative des élèves qui aurait moins d'effets négatifs sur les enseignants et les élèves que les tests nationaux actuels en Angleterre, des chercheurs et des conseillers de deux autorités locales ont travaillé avec les enseignants pour élaborer l'utilisation des jugements des enseignants dans l'évaluation sommative des élèves.<sup>130</sup> L'objectif visait à trouver des méthodes et des processus pour assurer la comparabilité des jugements entre les enseignants et les écoles. Dans sa phase pilote, le projet a travaillé avec un petit groupe d'enseignants d'anglais et de mathématiques des élèves de la huitième année (âgés de 13 ans). Les enseignants sélectionnés étaient

127 Hayward, L. & Spencer, E. (2010) The Complexities of Change: formative assessment in Scotland, *The Curriculum Journal*, 21 (2) 161-177. Routledge

128 Pedder, D. and James, M. (2012) Professional learning as a condition for assessment for learning. In (ed) J. Gardner *Assessment and Learning*. 2nd edn. London: Sage, pp 33-48.

129 *ibid* p 41/2

130 Black, P., Harrison, C., Hodgen, J., Marshall, B. and Serret, N (2006) Riding the interface: an exploration of the issues that beset teachers as they strive for assessment systems. BERA conference paper see <http://www.kcl.ac.uk/sspp/departments/education/research/crestem/assessment/riding.pdf>

tous versés dans les pratiques d'évaluation formative des élèves. Le projet a porté sur la façon de parvenir à des jugements sommatifs et de transformer des résultats bruts en jugements pertinents. Le développement a été initié par les enseignants qui ont voulu identifier quelles sont « les caractéristiques de réussite à une matière pour des élèves de la huitième année. » Dans la deuxième phase au cours de l'année scolaire 2005-2006, de nouvelles pratiques ont été testées et adaptées. Dans la troisième phase, des enseignants ont diffusé les idées et les pratiques dans leur département.

Les résultats du projet pilote, dérivés des notes de terrain, des observations en classe, des interviews et des comptes rendus de réunions, indiquaient une certaine confusion entre l'évaluation formative et l'évaluation sommative des élèves, et une acceptation passive des tests actuels. Des différences ont été observées dans les réactions des enseignants de mathématiques et d'anglais. Les enseignants d'anglais ont été signalés comme étant à l'aise avec un système de portfolio, en mettant davantage l'accent sur l'oral et l'écoute avec un travail « contrôlé » sur lequel les élèves travaillaient de manière individuelle. Les enseignants de mathématiques ont préféré garder les tests (certains se sont interrogés sur la nécessité du changement) en les améliorant ou en introduisant des « tâches d'évaluation alternatives », à la place d'une modalité plus holistique favorisée par les enseignants d'anglais. Toutefois, la façon de tester en mathématique s'est révélée incapable d'offrir des occasions aux élèves de montrer leur capacité de réussite : ce fut une surprise inattendue pour les enseignants.

### **Le développement de l'évaluation sommative des élèves par les enseignants du primaire**

Depuis 1995, les enseignants en Angleterre espéraient effectuer l'évaluation de leurs élèves ainsi que l'administration de tests nationaux en anglais, mathématiques et sciences en dernière année de l'école primaire (sixième année, 11 ans). En théorie, l'évaluation des élèves par des enseignants a été conçue pour fournir des données sur un plus grand éventail de sujets qui pourraient être inclus dans les tests. En pratique, de nombreux enseignants fondaient l'évaluation des élèves sur les seuls tests nationaux du passé, n'utilisant pas leurs jugements et les données recueillies d'une manière formative et sommative.

En 2006, un projet<sup>131</sup> a été lancé pour tenter de trouver un modèle alternatif pour l'évaluation des élèves et la modération des enseignants statutaires en fin de cycle primaire. Des réunions concernant les jugements de fin d'année ont été organisées avec un petit groupe d'enseignants de la sixième année. Ces jugements devaient être signalés sous forme de « niveau » atteint par chaque élève, les « niveaux » étant identifiés dans le National Curriculum Attainment Targets pour chaque matière. (Voir l'exemple dans l'encadré 19, page 75.) Ces réunions ont permis aux enseignants de réfléchir sur leur expérience de différentes manières de parvenir à des décisions sur les niveaux et de modérer les processus. Au cours des premières réunions, il est devenu évident que les enseignants utilisaient un ensemble restreint de preuves pour porter leurs jugements, probablement en mettant des tests à l'épreuve pour décider ce qu'il faut prendre en compte. Des jugements de niveaux ont été considérés comme problématiques, car les descriptions des niveaux couvraient deux années de travail. Malgré un désir de créer des sous-niveaux, il était difficile d'établir des preuves solides pour les étapes intermédiaires entre les niveaux. En revanche, la collaboration entre les conseillers et les enseignants a conduit à quelques propositions pour élargir la gamme de preuves utilisées et quelques principes pour guider la pratique de jugements sur les niveaux atteints.

Ces documents ont été rendus accessibles à un plus grand groupe d'enseignants de la sixième année venant de vingt-quatre écoles qui ont participé au projet en 2007. Les enseignants ont collecté des preuves ou résultats probants sur une période de deux trimestres et ont apporté des exemples à deux réunions où leurs jugements ont été modérés dans des discussions avec d'autres enseignants. Les réunions leur ont donné non seulement une expérience directe de la modération, mais également la

131 Birmingham City Council Advisory and Support Service (2005) *Effective Assessment*

possibilité de partager leurs expériences en discutant leurs sélections de preuves. Ils ont élargi la base de preuves de leurs jugements et ont commencé à réaliser que les élèves pourraient avoir un rôle dans leur propre évaluation. La réaction des enseignants sur l'expérience de la modération collaborative a été extrêmement positive<sup>132</sup> et ils pensaient que tous les enseignants devraient avoir de telles opportunités. Les enseignants ont attaché de l'importance à cette expérience pour le développement de documents destinés à les guider. Selon eux, pour une diffusion plus large, le processus devrait être « formalisé » afin d'être une modalité plus « descendante » pour les autres.

### **Le changement de la pédagogie et des pratiques de l'évaluation des élèves en France**

Les changements considérables dans l'éducation française entre 2006 et 2009, mentionnés au chapitre 6 (page 78), ont exigé que les écoles mettent en œuvre un socle commun de compétences et de connaissances. La spécification de compétences était une nouveauté, et en sciences elle a demandé à certains enseignants de changer leur méthode d'enseignement et leurs pratiques d'évaluation des élèves. Le projet *La main à la pâte* a soutenu les écoles de manière appropriée à la fois pour les nouveaux changements et les pratiques existantes. Dans un ensemble local [N.d.T. Le Centre pilote *La main à la pâte* de Perpignan en France, recouvrant plusieurs circonscriptions ou districts] les enseignants des écoles primaires ont été aidés à développer des modules d'enseignement sur un thème particulier en faisant une utilisation progressive des compétences d'investigation pour l'école maternelle, les première et deuxième années, et de la troisième à la cinquième année.<sup>133</sup> Chaque module contient une section sur l'évaluation des élèves. Les tests ont également été construits et remis à un échantillon d'élèves à la fin de l'école primaire (cinquième année) afin d'évaluer leurs compétences. La performance n'est pas notée sous forme de note ni de scores, mais avec des mentions comme : « travail fondé sur l'investigation », « progrès en cours vers l'investigation » et « travail pas encore fondé sur l'investigation ». Les rapports non publiés montrent qu'en trois ans, le nombre d'élèves ne travaillant pas par l'investigation est passé de 29% à 8%.

Dans une académie où l'ESFI a déjà été mis en pratique, les écoles ont eu accès à des tâches comme celles décrites dans l'exemple 7<sup>134</sup> (page 79) pour leur permettre d'évaluer les progrès des élèves dans les compétences et les connaissances en investigation. Bien que les tests n'aient pas été conçus comme une évaluation sommative des élèves, les enseignants ont utilisé les résultats ainsi que leurs observations sur les actions des élèves et d'autres preuves afin de compléter le livret de chaque élève de fin d'année. L'effet de l'accès aux tâches et au guide d'utilisation de preuves dans les cahiers, les exposés et les actions des élèves (voir page 78) a été lent à apparaître. Peu de changements ont été observés la première année, mais par la suite, l'utilisation des tâches d'évaluation des élèves et les conseils se sont propagés rapidement dans une large proportion (environ 75%) des écoles du district.

### **Le développement des pratiques de l'évaluation des élèves à travers le programme scolaire en Australie<sup>135</sup>**

En Australie, le projet *Primary Connections* vise à développer les connaissances, les compétences, la compréhension et les capacités des élèves en sciences et en alphabétisation. Il favorise une méthode d'investigation par le biais d'une structure de leçon en cinq phases : s'engager, explorer, expliquer, développer et évaluer. Le développement du projet a été financé par le gouvernement australien de 2004 à 2012 et il est appliqué partiellement par 56% des écoles primaires australiennes. Le programme

132 Harlen, W. (2010) Professional learning to support teacher assessment, in Gardner, J. et al *Developing Teacher Assessment*. Maidenhead, England: Open University Press page 100-129.

133 (see [http://www.crdp-montpellier.fr/cd66/map66/projets\\_federatifs/air/index.php](http://www.crdp-montpellier.fr/cd66/map66/projets_federatifs/air/index.php)).

134 Il s'agit ici de l'Académie de Toulouse (France).

135 Primary Connections <http://science.org.au/primaryconnections/>

comporte deux volets principaux : un programme d'apprentissage professionnel et un ensemble de trente-et-une unités du programme scolaire qui couvrent le programme de sciences de la première à la sixième année en Australie. Les unités du programme marquent les pratiques d'investigation et l'évaluation des élèves de l'apprentissage et pour l'apprentissage. Le programme d'apprentissage professionnel comprend un atelier sur l'évaluation des élèves et plusieurs DVD de formation ont été élaborés pour fournir des exemples pratiques en classes.

Des stratégies fournissant des informations à la fois à l'évaluation formative et à l'évaluation sommative des élèves sont proposées. Les séries de questions des enseignants sont incluses dans les étapes des leçons pour susciter la réflexion des élèves et pour rendre leurs idées accessibles aux enseignants et aux élèves, ceci ayant pour objectif un contrôle de l'apprentissage. Les stratégies pour l'évaluation par les pairs et l'auto-évaluation sont également incluses. L'annexe de chaque unité résume les possibilités d'évaluer les élèves pour chaque phase, et des grilles d'évaluation sont disponibles sur le site internet pour aider les enseignants à surveiller la compréhension conceptuelle et les compétences d'investigation des élèves.

Un rapport de commentaires des enseignants du projet<sup>136</sup> résume les objectifs de la phase d'évaluation :

- donner aux élèves l'opportunité d'examiner et de réfléchir sur leur apprentissage, une nouvelle compréhension et de nouvelles compétences
- fournir des preuves pour des changements dans la compréhension, les opinions et les compétences des élèves.

Au niveau du premier objectif, des preuves et résultats probants solides ont montré que les élèves avaient examiné leur compréhension conceptuelle mais pas leurs compétences, même si d'autres preuves ont indiqué l'utilisation d'une variété de compétences. Il y avait peu de réponses des enseignants par rapport au second objectif qui était probablement négligé par de nombreux enseignants. Les participants ont utilisé un ensemble de stratégies qui ont servi aux deux objectifs de la phase d'évaluation systémique, mais encore une fois seulement au niveau de la compréhension conceptuelle. Les stratégies comprennent des questionnaires, la rédaction d'un article de journal, la création d'un animal imaginaire basée sur des critères, des dessins, des schémas, des boucles de mots, des cartes conceptuelles, des dessins animés conceptuels, des jeux de rôle et des exposés. Les enseignants ont noté qu'un ensemble de tâches qui demandaient énormément de temps de la part des enseignants, incluant le questionnement individuel et des situations nouvelles, était nécessaire pour évaluer la compréhension des élèves.

Dans l'ensemble, bien que l'évaluation des élèves par les enseignants n'ait pas été très développée, les résultats des recherches montrent que certains enseignants s'orientent progressivement vers de meilleures pratiques d'évaluation des élèves en prenant progressivement confiance en eux-mêmes. Les procédures efficaces pour améliorer les pratiques d'évaluation des élèves par les enseignants sont :

- les ressources du programme qui intègrent des exemples d'évaluation des élèves fondées sur la recherche
- l'apprentissage professionnel qui fournit les théories pour expliquer le fonctionnement de la méthode
- les DVD pédagogiques qui montrent à quoi ressemblent les véritables salles de classe
- les expériences pédagogiques positives dans lesquelles le plaisir des élèves et des preuves de l'apprentissage mènent au plaisir et la motivation des enseignants pour s'engager vers un changement de pratique pédagogique en classe.

136 Skamp, K. (2012) Trial-teacher feedback on the implementation of Primary Connections and the 5E model. Australian Academy of Science <http://www.science.org.au/primaryconnections/research-and-evaluation/teaching-ps.html>

## Les implications de l'évaluation des élèves au sein de l'ESFI

Des leçons doivent être tirées de ces tentatives de réformer l'évaluation des élèves qui peut s'appliquer à l'apprentissage et à l'enseignement de toutes les matières. Avant d'examiner les éléments nécessaires dans le cas de l'évaluation des élèves de l'enseignement des sciences fondé sur l'investigation, il est utile de rappeler que l'ensemble de l'enseignement des sciences ne concerne pas l'investigation, il doit donc y avoir une variété de tâches et de procédures d'évaluation des élèves pour répondre à la diversité des objectifs d'apprentissage. Comme mentionné précédemment, il y a une place pour l'enseignement direct dans les conventions, le vocabulaire et les connaissances de base qui peuvent être vérifiés par des tests ou des questionnaires faits par les enseignants. Mais l'évaluation des élèves ne doit pas exiger des réponses qui peuvent être apprises par cœur. Sinon, il n'y aura pas d'informations sur l'étendue des compétences en investigation et la compréhension scientifique des élèves. Ici, notre accent est mis sur les éléments essentiels de l'évaluation des élèves autour des objectifs de l'ESFI.

Malheureusement, l'utilisation de procédures d'évaluation des élèves qui répondent aux objectifs de l'ESFI n'est pas encore répandue, et pratiquement aucune expérience n'existe sur les éléments nécessaires pour changer les pratiques de l'évaluation des élèves afin de mieux s'adapter à ces fins. Pour fournir des données sur les compétences qui font partie des objectifs de l'ESFI, il faut un ensemble de procédures d'évaluation des élèves, certaines impliquent l'observation des actions des élèves, d'autres permettent aux élèves d'exprimer leurs idées oralement et par écrit tout en offrant la possibilité d'utiliser les compétences et le raisonnement d'investigation scientifique. Comme indiqué dans le chapitre 6, des exemples de telles procédures existent mais ne sont pas largement utilisés dans l'enseignement des sciences. Toutefois, par extrapolation à partir des procédures réussies dans d'autres domaines, il est possible de trouver quelques indications sur ce qui aide au mieux les enseignants à changer leur pratique de l'évaluation des élèves en science.

Quatre éléments clés semblent jouer un rôle dans les changements de pratiques d'évaluation des élèves au niveau des classes et des écoles à la fois au niveau de l'évaluation formative et de l'évaluation sommative des élèves :

- la motivation pour un changement
- les objectifs à réaliser
- les occasions pour les enseignants de discuter, comparer et partager des solutions
- les moyens d'évaluer la nouvelle pratique

*La motivation* pour un changement provient du mécontentement issu de la pratique actuelle et de la conviction qu'une amélioration est possible. Dans l'expérience du projet de KMOFAP (page 85), un facteur fondamental pour convaincre les enseignants d'introduire l'évaluation formative des élèves était des preuves de la réussite des élèves. Le facteur de motivation de continuer le projet provenait des preuves de l'impact des changements dans leur enseignement et chez leurs propres élèves. Dans le cas de l'évaluation sommative des élèves, la motivation pour un changement peut venir du fait que des preuves ont montré aux enseignants que l'évaluation sommative conventionnelle par les tests classiques ne donnait pas à leurs élèves l'occasion de montrer ce qu'ils étaient capables de faire et comprendre. Susciter chez les enseignants un intérêt voire un engagement envers des recherches comme celles de Dolin et Krogh, notées dans le chapitre 6 (page 70), pourraient avoir une influence. Les preuves ou résultats probants qui montrent que la procédure d'évaluation des élèves peut faire une différence significative constituent une première étape pour les enseignants qui prennent conscience que, derrière ces méthodes d'évaluation, c'est la vision de l'apprentissage qui fait la différence. Une approche, mettant l'évaluation des élèves plus en cohérence avec le cadre socio-culturel favorable à l'ESFI, permet aux élèves de montrer leurs capacités sur une plus large étendue que les tests ou des tâches individuelles.

*Des objectifs à réaliser* sous-entendent que les enseignants et les écoles ont une idée précise du changement nécessaire ; néanmoins, ils doivent eux-mêmes développer en détail les étapes nécessaires pour réaliser ce changement, en tenant compte de leur contexte spécifique. Dans le cas de l'évaluation formative des élèves, les enseignants peuvent s'informer sur l'ensemble des stratégies efficaces mais peuvent adopter des façons différentes de les mettre en œuvre dans leur propre classe. Au niveau du changement de l'évaluation sommative des élèves, les enseignants peuvent être aidés à affiner leurs idées sur l'identification des réussites et des progrès des élèves et la manière de s'assurer que les élèves peuvent les montrer.

*Les occasions de discuter, comparer et partager des solutions* assurent que même si les pratiques étaient adaptées à des situations particulières, elles maintiennent l'accent sur des objectifs communs. Les discussions avec les autres tentent d'obtenir les mêmes effets dans des contextes différents, elles fournissent aux enseignants des idées qui peuvent ne pas avoir eu lieu avec un travail individuel. L'investigation conduite en collaboration fonctionne pour aider les enseignants à construire et reconstruire des idées comme pour les élèves. Il est important que cette collaboration et cette réflexion partagée soient accessibles à tous. Certains projets qui ont débuté avec le développement collaboratif de nouvelles pratiques d'évaluation des élèves par des enseignants ont conduit à la production de matériaux qui ont été jugés nécessaires pour les autres enseignants. Dans ce cas, ce qui commence comme une transformation pour un petit nombre se termine en transmission pour un plus grand nombre. La solution pour échapper à cela est l'« équilibre délicat », identifié par Leahy et Wiliam, par lequel on peut éviter l'écueil des modèles de transmission du changement : ces modèles qui suggèrent que la pratique peut être changée en suivant une recette fournie par d'autres, fussent-ils enseignants.

*Les moyens d'évaluer les changements* dans les pratiques d'évaluation des élèves permettent aux enseignants de continuer à développer et à ajuster leur pratique aux nouvelles circonstances. Il est important de ne pas changer (au moins sur le long terme) les principes qui guident et apportent la justification des choix. À moins que les raisons du changement soient bien comprises, les techniques doivent être suivies un peu à l'aveuglette ; elles ne peuvent pas être adaptées à toutes les situations et finiront par devenir moins utiles. Comme dans le cas des enseignants du projet KMOFAP (page 85), l'intérêt de connaître *les raisons* pour lesquelles les nouvelles pratiques fonctionnent peut émerger seulement après la prise de conscience du résultat. Toutefois, si le motif sous-jacent est présenté tôt ou tard pendant le changement, les enseignants ont besoin d'avoir à un moment donné une raison pour adopter les nouvelles pratiques et abandonner les anciennes.

Un exemple de principes relatifs à la pratique de l'évaluation des élèves illustré dans l'encadré 21 provient d'un projet qui a étudié les processus de changement dans les pratiques d'évaluation des élèves, notamment en plaçant le rôle des enseignants au centre de l'évaluation.

Le dernier principe implique les normes de qualité qui peuvent être utilisées pour évaluer la pratique. Après l'identification des principes, les acteurs décisifs du projet ont également proposé des normes à respecter pour l'évaluation des élèves en classe, au niveau de l'école, au niveau des collectivités locales et au niveau de la politique éducative nationale. Pour l'élève, l'expérience en classe a l'impact le plus important sur l'apprentissage, donc des normes sont à respecter si l'évaluation des élèves vise à soutenir l'apprentissage. Cependant l'expérience en classe dépend des politiques d'évaluation des élèves de l'école et du discours général sur comment évaluer, signaler ou mesurer les résultats des élèves. À son tour, la politique et la pratique de l'école sont influencées par les conseils des collectivités locales, les politiques et les besoins nationaux. Par conséquent, le projet a identifié des normes à respecter en pratique au sein de chacune de ces communautés qui sont reproduites dans les tableaux 1 - 4. Dans chaque cas, il existe des normes pour l'évaluation des élèves en général, pour l'évaluation formative et l'évaluation sommative des élèves.

*Encadré 21: Les principes de la pratique de l'évaluation des élèves*

- 1 Toute évaluation des élèves doit améliorer l'apprentissage à terme.
- 2 Les méthodes d'évaluation des élèves doivent venir en appui et signaler les progrès vis-à-vis de tous les objectifs d'apprentissage importants.
- 3 Les procédures d'évaluation des élèves doivent inclure des processus explicites pour garantir la validité et la fiabilité des informations par rapport aux objectifs.
- 4 L'évaluation doit favoriser la compréhension par tous d'objectifs d'apprentissage pertinents pour la vie présente et future des élèves.
- 5 L'évaluation des résultats de l'apprentissage doit tenir compte d'erreurs inévitables.
- 6 L'évaluation doit faire partie d'un processus d'enseignement qui permette aux élèves de comprendre les objectifs de leur apprentissage et comment la qualité de leur performance sera jugée.
- 7 Les méthodes d'évaluation doivent promouvoir la participation active des élèves dans leur apprentissage et leur évaluation.
- 8 L'évaluation doit permettre aux élèves de montrer ce qu'ils peuvent faire.
- 9 L'évaluation des élèves doit combiner des informations de différents types, y compris les auto-évaluations des élèves, pour éclairer les décisions à prendre au sujet de l'apprentissage et des résultats des élèves.
- 10 Les méthodes d'évaluation doivent répondre à des normes ou principes qui reflètent un consensus large sur la qualité, depuis la pratique en classe jusqu'à la politique nationale.<sup>137</sup>

---

137 Gardner, J., Harlen, W., Hayward, L. and Stobart, G. with Montgomery, M. (2010). *Developing Teacher Assessment*. Maidenhead: Open University Press pp 48-51.

Tableau 1: Les principes de la pratique de l'évaluation des élèves en classe

Évaluation des élèves en général	Évaluation formative des élèves	Évaluation sommative des élèves
<p>1 L'évaluation des élèves utilise un ensemble de méthodes qui permettent de considérer les différents objectifs d'apprentissage ainsi que la progression dans leur direction</p> <p>2 Les méthodes utilisées pour évaluer concernent les compétences, les connaissances et la compréhension en couvrant l'ensemble du programme dans la matière considérée</p> <p>3 L'enseignement apporte aux élèves l'occasion de montrer ce qu'ils sont capables de faire grâce à des tâches qui reflètent l'ensemble des objectifs de l'apprentissage</p> <p>4 Les enseignants utilisent les preuves ou résultats probants de leur évaluation continue des élèves pour :</p> <ul style="list-style-type: none"> <li>• améliorer l'apprentissage des élèves</li> <li>• être capable faire une synthèse de l'apprentissage par référence aux critères qui leur sont donnés</li> <li>• Réfléchir sur leur enseignement et l'améliorer</li> </ul> <p>5 Les enseignants développent leur pratique de l'évaluation des élèves grâce à leurs activités de développement professionnel, incluant la réflexion et le partage d'expériences avec des collègues</p>	<p>1 Les enseignants recueillent des preuves de l'apprentissage de leurs élèves grâce au questionnement, à l'observation, à la discussion et à l'étude de productions d'élèves reflétant les objectifs de l'apprentissage</p> <p>2 Les enseignants font participer les élèves à des discussions sur les objectifs d'apprentissage et les performances attendues de leur travail</p> <p>3 Les enseignants utilisent l'évaluation des élèves pour les aider à progresser dans leur apprentissage en :</p> <ul style="list-style-type: none"> <li>• adaptant le rythme, les défis et le contenu des activités</li> <li>• donnant des commentaires aux élèves sur la façon de s'améliorer</li> <li>• accordant du temps aux élèves pour réfléchir et évaluer leur propre travail</li> </ul> <p>4 Les élèves utilisent l'évaluation qui est faite de leur travail afin de progresser en :</p> <ul style="list-style-type: none"> <li>• prenant connaissance et utilisant les critères des performances attendues</li> <li>• donnant à et recevant de leurs pairs des commentaires sur la qualité de leur travail et la façon de l'améliorer</li> <li>• réfléchissant sur la façon d'améliorer leur travail en assumant leur propre responsabilité</li> </ul>	<p>1 Les enseignants fondent leur jugement d'après les résultats des élèves, obtenus à partir d'une série d'activités adaptées à l'objet d'étude et à l'âge des élèves ; elles pourraient inclure des tests ou des tâches spécifiques d'évaluation</p> <p>2 L'évaluation des résultats des élèves est basée sur une grande variété de tâches qui permet aux élèves de manifester ce que signifie « être bon élève », s'agissant d'un travail donné</p> <p>3 Les enseignants participent à des discussions entre collègues sur le travail des élèves afin d'harmoniser leurs jugements de niveaux ou de notes au moment voulu</p> <p>4 Les élèves sont conscients des critères sur lesquels leur travail est jugé périodiquement</p> <p>5 Les élèves sont conscients des preuves ou résultats probants utilisés et savent comment les jugements portant sur leurs résultats d'apprentissage sont établis</p> <p>6 Les élèves sont aidés dans l'utilisation des résultats de l'évaluation afin d'améliorer leur apprentissage</p>



Tableau 2: Les principes destinés aux équipes de gestion de l'école

Évaluation des élèves en général	Évaluation formative des élèves	Évaluation sommative des élèves
<ol style="list-style-type: none"> <li>1 Il existe une politique scolaire d'évaluation des élèves qui reflète les principes ci-dessus pour leur mise en pratique dans la classe</li> <li>2 La politique d'évaluation est régulièrement discutée et examinée à la lumière des objectifs convenus</li> <li>3 Les enseignants ont la possibilité d'améliorer leur pratique de l'évaluation des élèves en collaborant et en l'incluant dans leur développement professionnel</li> <li>4 Du temps est mis à la disposition des enseignants pour discuter, réfléchir et, à l'occasion, observer la façon dont chacun pratique l'évaluation des élèves</li> <li>5 La politique d'évaluation et sa mise en pratique sont communiquées aux parents</li> </ol>	<p>Les enseignants collaborent à développer leur pratique en :</p> <ul style="list-style-type: none"> <li>• communiquant aux élèves des objectifs et des critères de qualité</li> <li>• en aidant les élèves à participer à leur auto-évaluation et à l'évaluation mutuelle</li> <li>• en fournissant des commentaires venant à l'appui de l'apprentissage</li> <li>• en rendant les élèves responsables vis à vis de leur travail</li> </ul>	<ol style="list-style-type: none"> <li>1 Les enseignants sont capables d'utiliser une diversité de méthodes d'évaluation des élèves, sans que celle-ci soit toujours sous la pression d'enjeux majeurs</li> <li>2 Les enseignants participent à l'élaboration de procédures d'assurance-qualité afin de donner à leurs jugements la plus grande cohérence possible</li> <li>3 Les résultats des élèves sont discutés en termes de ce qu'ils peuvent faire et non seulement en termes de niveaux ou de notes</li> <li>4 Un système gérable de tenue des dossiers d'élèves est en place, permettant le suivi de l'apprentissage et des appréciations ou rapports</li> <li>5 Les parents reçoivent des rapports écrits et oraux qui leur permettent d'identifier les étapes suivantes attendues de leurs enfants. Ces rapports contiennent des informations sur les processus d'évaluation mis en œuvre afin de leur donner confiance dans l'évaluation faite par les enseignants</li> <li>6 Les jugements sommatifs ne sont nécessaires que lorsqu'il est nécessaire de vérifier les progrès et de faire un rapport sur eux</li> </ol>

Tableau 3: Les principes destinés aux inspections locales et nationales

Évaluation des élèves en général	Évaluation formative des élèves	Évaluation sommative des élèves
<ol style="list-style-type: none"> <li>1 Les politiques et pratiques des écoles en matière d'évaluation des élèves sont examinées en référence aux principes ci-dessus</li> <li>2 Les procédures d'inspection s'assurent que les écoles évaluent leurs pratiques d'évaluation des élèves et élaborent des plans d'action pour les améliorer</li> <li>3 Il existe des possibilités pour les écoles de mettre en commun et de développer des pratiques d'évaluation des élèves</li> <li>4 Le perfectionnement professionnel des enseignants est mis en oeuvre pour élaborer une politique et améliorer les pratiques d'évaluation des élèves</li> <li>5 Des ressources sont disponibles pour permettre aux écoles d'améliorer les pratiques d'évaluation des élèves</li> </ol>	<ol style="list-style-type: none"> <li>1 L'utilisation de l'évaluation des élèves par les écoles, afin de venir en appui à l'apprentissage est considérée comme un facteur clé dans l'évaluation de l'efficacité des écoles</li> <li>2 Les écoles disposent d'aides pour s'assurer que leurs succès bénéficient de l'utilisation de l'évaluation formative des élèves</li> <li>3 Les écoles sont encouragées à utiliser l'évaluation formative des élèves</li> </ol>	<ol style="list-style-type: none"> <li>1 Les écoles disposent d'aide pour élaborer des plans d'action, fondés sur l'auto-évaluation, grâce à une série d'indicateurs allant au-delà des niveaux de performance des élèves</li> <li>2 Les orientations proposées sur les politiques et les pratiques d'évaluation scolaire tiennent compte de ce qui est connu sur la fiabilité et la validité des différentes méthodes d'évaluation des élèves</li> <li>3 Les écoles disposent d'aide pour utiliser au mieux les résultats des évaluations des élèves, en vue de mieux saisir les occasions d'améliorer l'apprentissage</li> </ol>

Tableau 4: Les principes destinés à la formulation des politiques nationales

Évaluation des élèves en général	Évaluation formative des élèves	Évaluation sommative des élèves
<ol style="list-style-type: none"> <li>1 Les politiques exigent des écoles et des conseillers locaux de montrer dans quelle mesure toute évaluation est utilisée pour venir à l'appui de l'apprentissage des élèves</li> <li>2 L'introduction de nouvelles pratiques en matière d'évaluation des élèves est accompagnée de changements dans la formation des enseignants et dans les critères d'évaluation systémique</li> <li>3 Les écoles sont responsables de l'application de l'évaluation formative et de l'évaluation sommative des élèves, afin de maximiser la réalisation des objectifs.</li> <li>4 Les standards nationaux mesurant la performance des élèves sont présentés comme une série de données qualitatives et quantitatives, provenant d'enquêtes faites auprès d'échantillons représentatifs</li> </ol>	<ol style="list-style-type: none"> <li>1 Une évaluation des élèves favorisant l'apprentissage est au cœur des programmes gouvernementaux visant à élever le niveau de réussite</li> <li>2 La formation initiale des enseignants et les actions de développement professionnel s'assurent que les enseignants ont les compétences nécessaires pour que l'évaluation des élèves contribue au soutien de l'apprentissage</li> <li>3 Les systèmes d'inspection des écoles donnent la priorité à l'utilisation de l'évaluation des élèves pour venir en soutien à l'apprentissage</li> <li>4 Les écoles sont encouragées à évaluer et à développer un usage de l'évaluation formative des élèves</li> </ol>	<ol style="list-style-type: none"> <li>1 L'évaluation des élèves, animée par les enseignants, est utilisée pour juger de la performance des élèves tout au long de la scolarité obligatoire</li> <li>2 Une modération collective des jugements des enseignants est nécessaire pour s'assurer d'une interprétation commune des critères, au sein de l'école et entre écoles</li> <li>3 Les règlements s'assurent que les dispositions relatives à l'utilisation de l'évaluation sommative des élèves sont compatibles avec une pratique qui favorise l'apprentissage</li> <li>4 Les objectifs d'amélioration de l'école sont fondés sur une série d'indicateurs. Ils combinent une évaluation systémique externe et une auto-évaluation interne</li> </ol>

## Bibliographie de la version originale en anglais

- Alexander, R. (Ed) (2010) *Children, their World, their Education*. Final report and recommendations of the Cambridge Primary Review. London: Routledge.
- Alexander, R. (2004) *Towards Dialogic Teaching. Rethinking Classroom Talk*. Cambridge: Dialogos.
- Assessment Reform Group (ARG) (2002) *Assessment for Learning: 10 Principles*. [www.assessment-reform-group.org](http://www.assessment-reform-group.org)
- Australian Government Department Education, Employment and Workplace Relations. *Science Education Assessment Resources (SEAR)* <http://cms.curriculum.edu.au/SEAR>
- Barnes, D. (1976) *From Communication to Curriculum*. Harmondsworth: Penguin.
- Birmingham City Council Advisory and Support Service (2005) *Effective Assessment*.
- Black, P. (1998) *Testing: Friend or Foe?* London: Falmer Press.
- Black, P., Harrison, C., Hodgen, J., Marshall, B. and Serret, N (2006) Riding the interface: an exploration of the issues that beset teachers as they strive for assessment systems. BERA conference paper . See <http://www.kcl.ac.uk/sspp/departments/education/research/crestem/assessment/riding.pdf>
- Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2003). *Assessment for Learning: Putting it into Practice*. Maidenhead England: Open University Press.
- Black, P. and Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21 (1). 5-13.
- Black, P. and Wiliam, D. (1998) Assessment and classroom learning, *Assessment in Education*, 5 (1) 7-74.
- Bransford, J.D., Brown, A. and Cocking, R.R. (eds) (2000) *How People Learn, Brain, Mind, Experience and School*. Washington, D.C.: National Academy Press.
- Bruner, J. (1996) *The Culture of Education*. Cambridge, MA: Harvard University Press.
- Budd-Rowe, M. (1974) Relation of wait-time and rewards to the development of language, logic and fate control: Part II, *Journal of Research in Science Teaching*, 11(4) 291-308.
- Butler, R. (1988) Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on interest and performance, *British Journal of Educational Psychology* 58, 1-14.
- Crossouard (2012) Absent presences: the recognition of social class and gender dimensions within peer assessment interactions, *British Educational Research Journal*, 38 (5) 731-748.
- Department for Education (2011). *The Framework for the National Curriculum. A report by the Expert Panel for the National Curriculum review*. London: Department for Education.
- DES, DENI and WO (1985) *APU Science in Schools Age 11 Report no 4*. London: HMSO.
- DES and WO (1988) *National Curriculum Task Group on Assessment and Testing: A Report*. London: HMSO.
- DES, DENI and WO (1981) *Science in Schools Age 11 Report no 1*. London: HMSO.
- Dewey, J. (1933) *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston, MA: D.C. Heath.
- Dolin, J., & Krogh, L. B. (2010): The Relevance and Consequences of Pisa Science in a Danish Context. *International Journal of Science and Mathematics Education*, 8, 565-592.

- Dolin, J., Laursen, E., Raae, P. H., Senger, U. (2005). *Udviklingsprojekter som læringsrum. Potentialer og barrierer for skoleudvikling i det almene gymnasium. Gymnasiepædagogik nr. 54*, Syddansk Universitet. 232 s. (Development projects as learning arenas. Potentials and barriers for school development in the general upper secondary school. University of Southern Denmark). [http://www.sdu.dk/Om\\_SDU/Institutter\\_centre/lkv/Formidling/Tidsskrifter/Gymnasiepædagogik/Udga-ver](http://www.sdu.dk/Om_SDU/Institutter_centre/lkv/Formidling/Tidsskrifter/Gymnasiepædagogik/Udga-ver) Accessed Dec11, 2012.
- Ertl, H. (2006) Educational standards and the changing discourse on education: The reception and consequences of the PISA study in Germany. *Oxford Review of Education*, 32 (5) pp 619–634.
- Gardner, J., Harlen, W., Hayward, L. and Stobart, G. with Montgomery, M. (2010). *Developing Teacher Assessment*. Maidenhead: Open University Press.
- Gipps, C., McCallum, B. and Brown, M. (1996). Models of teacher assessment among primary school teachers in England, *The Curriculum Journal*, 7 (2) 167-183.
- Grigorenko, E. (1998) Mastering tools of the mind in school (trying out Vygotsky's ideas in classrooms), in (eds) R. Sternberg and W. Williams *Intelligence, Instruction and Assessment: Theory and Practice*. Mahwah, NJ: Erlbaum.
- Harlen, W. (2012a) The role of assessment in developing motivation for learning, in (ed) J. Gardner *Assessment and Learning*. London: Sage pp171-184.
- Harlen, W. (2012b) On the relationship between assessment for formative and summative purposes, in (ed) J. Gardner *Assessment and Learning*. London: Sage pp 87-102.
- Harlen, W. (2010) Professional learning to support teacher assessment, in Gardner, J. *et al Developing Teacher Assessment*. Maidenhead England: Open University Press page 100-129.
- Harlen (Ed) (2010) *Principles and Big Ideas of Science Education*. Available from [www.ase.org.uk](http://www.ase.org.uk) in English, from [www.fondation-lamap.org](http://www.fondation-lamap.org) in French, and from [www.innovec.org.mx](http://www.innovec.org.mx) in Spanish.
- Harlen, W. (2007) *Assessment of Learning*. London: Sage.
- Harlen, W. (2006) *Teaching, Learning and Assessing Science 5 – 12*. 4<sup>th</sup> edn. London: Sage.
- Harlen, W. (2004) Trusting teachers' judgements: research evidence of the reliability and validity of teachers' assessment for summative purposes, *Research Papers in Education*, 20(3); 245-270.
- Harlen, W. and Qualter, A. (2009) *The teaching of Science in Primary Schools*. London: Routledge.
- Hattie, J. and Timperley, H. (2007) The power of feedback. *Review of Educational Research*, 77, 81-112.
- Hayward, L. (2010) Moving beyond the classroom, in J. Gardner et al *Developing Teacher Assessment*. Maidenhead, UK: Open University Press.
- Hayward, L. & Spencer, E. (2010) The Complexities of Change: formative assessment in Scotland, *The Curriculum Journal*, 21 (2) 161-177.
- IAP (2012) Taking Inquiry-Based Science Education into Secondary Education. Report of a global conference. <http://www.sazu.si/files/file-147.pdf>
- Jager, J.J., Merki, K.M., Oerke, B. and Holmeier, M. (2012) State-wide low-stakes tests and a teaching to the test effect? An analysis of teacher survey data from two German States, *Assessment in Education*, 19 (4) 451-467.
- James, M. (2012) Assessment in harmony with our understanding of learning: problems and possibilities, in (ed) J. Gardner *Assessment and Learning*, 2<sup>nd</sup> edn. London: Sage 187 – 205.
- Leahy, S. and William D. (2012) From teachers to schools: scaling up professional development for formative assessment, in (ed) J. Gardner *Assessment and Learning*, 2<sup>nd</sup> edn. London: Sage. 49-71.

- Leahy, S. and Wiliam, D. (2009) *Embedding Assessment for Learning – A Professional Development Pack*. London: Specialist Schools and Academies Trust.
- Leahy, S. and Wiliam, D. (2010) *Embedding Assessment for Learning – Pack 2*. London: Specialist Schools and Academies Trust.
- Linn, R. L. (2000) Assessments and accountability, *Educational Researcher*, 29 (2) 4-16.
- Masters, G. and Forster, M. (1996) *Progress Maps*. Camberwell, Victoria, Australia: ACER.
- Maxwell, G. (2004) 'Progressive assessment for learning and certification: some lessons from school-based assessment in Queensland.' Paper presents at the third conference of the Association of Commonwealth Examination and Assessment Boards, March Nidi, Fiji.
- Messick, S.(1989) Validity, in (ed) R. Linn *Educational Measurement* (3<sup>rd</sup> Edn) American Council on Education , Washington: Macmillan, pp 13-103.
- Michaels, S., Shouse, A.W. and Schweingruber, H.A (2008) *Ready, Set, Science! Putting research to work in K-8 Science Classrooms*, Washington: National Academies Press.
- Minner, D.D., Levy, A. J and , Century, J. (2010) Inquiry-Based Science Instruction—What Is It and Does It Matter? Results from a Research Synthesis Years 1984 to 2002, *Journal of Research in Science Teaching*, 47 (4) 474-496.
- National Research Council (2012) *A Framework for K-12 Science Education*. Washington DC: National Academies Press.
- Newton, P. E. (2012) Validity, purpose and the recycling of results from educational assessment, in (Ed) J. Gardner *Assessment and Learning* 2<sup>nd</sup> edition. London: Sage 264-276.
- Noble, T., Suarez, C., Rosebery, A., O'Connor, M.C. ,Warren, B. and Hudicourt-Barnes, J. (2012) "I never thought of it as freezing": How Students Answer Questions on large-scale science tests and what they know about science, *Journal of Research in Science Teaching*, 49 (6) 778–803.
- Nordenbo, S. E., Allerup, P., Andersen, H. L., Dolin, J., Korp, H., Larsen, M. S., et al. (2009). *Pædagogisk brug af test - Et systematisk review*. København: Aarhus Universitetsforlag. (In English: *Pedagogical use of tests – A systematic review*).
- Nuffield Primary Science Teachers' Guide *Materials*. (1995) London: Collins Educational.
- Nusche, D., Laveault, D., MacBeath, J. and Santiago, P. (2012) *OECD Reviews of Evaluation and Assessment in Education: New Zealand 2011*. Paris: OECD.
- OECD (2011) *Towards an OECD Skills Strategy*. Paris: OECD.
- OECD (2006) *PISA released items: Science*. Paris: OECD <http://www.oecd.org/pisa/38709385.pdf>
- OECD 2003, *The PISA 2003 Assessment Framework* Paris: OECD.
- OECD (2000) *Measuring Student Knowledge and Skills: A new Framework for Assessment*. Paris: OECD.
- Osmundson, E., Chung, G., Herl, H., Klein D. (1999) *Knowledge-mapping in the classroom: a tool for examining the development of students' conceptual understandings*. Los Angeles, California: National Centre for Research on Evaluation and Student Testing, University of California. [www.cse.ucla.edu/Reports/TECH507.pdf](http://www.cse.ucla.edu/Reports/TECH507.pdf)
- Osborne, J., Simon, S. and Collins, S.(2003) Attitudes towards science: a review of the literature and its implications, *International Journal of Science Education*, 25, 1049-1079.
- Pedder, D. and James, M. (2012) Professional learning as a condition for assessment for learning. In (ed) J. Gardner *Assessment and Learning*. 2<sup>nd</sup> edn. London: Sage, pp 33-48.
- Pellegrino, J.W., Chudowsky, N. and Glaser, R. (Eds) (2001) *Knowing what Students Know The Science and Design and Educational Assessment*. Washington, DC: National Academy Press.

- Piaget, J (1929) *The Child's Conception of the World*. New York: Harcourt Brace.
- Pine, J., Aschbacher, P., Rother, E., Jones, M., McPhee, C., Martin, C., Phelps, S., Kyle, T. and Foley, B. (2006) Fifth graders' science inquiry abilities: a comparative study of students in hands-on and textbook curricula, *Journal of Research in Science Teaching* 43 (5): 467-484.
- Pollard, A and Triggs, P. (2000) *Policy, Practice and Pupil Experience*. London: Continuum.
- Pollard, A., Triggs, P., Broadfoot, P., Mcness, E. and Osborn, M. (2000) *What pupils say: changing policy and practice in primary education*. London: Continuum.
- Primary Connections <http://science.org.au/primaryconnections/>
- Pryor, J. and Lubisi, C. (2001) Reconceptualising educational assessment in South Africa –testing times for teachers, *International Journal for Educational Development*, 22 (6), 673-686.
- Roderick, M. and Engel, M. (2001) The grasshopper and the ant: motivational responses of low achieving pupils to high stakes testing. *Educational Evaluation and Policy Analysis* 23: 197-228.
- Sadler, D. R. (1989) Formative assessment and the design of instructional systems, *Instructional Science*, 18, 119-44.
- SEED (Scottish Executive Education Department) (2002) *How Good is Our School? Self evaluation using quality indicators*. Edinburgh: HMIE.
- Skamp, K. (2012) *Trial-teacher feedback on the implementation of Primary Connections and the 5E model*. Australian Academy of Science. <http://www.science.org.au/primaryconnections/research-and-evaluation/teaching-ps.html>
- Stobart, G. (2012) Validity in formative assessment, in (ed) J. Gardner *Assessment and Learning*. 2<sup>nd</sup> edn. London: Sage pp 233-242.
- Stobart, G. (2008) *Testing Times. The uses and abuses of assessment*. London: Routledge.
- Streeter, L., Bernstein, J., Foltz, P. and DeLand, D. (2011) *Pearson's Automated Scoring of Writing, Speaking, and Mathematics*. Pearson  
<http://kt.pearsonassessments.com/download/PearsonAutomatedScoring-WritingSpeakingMath-051911.pdf>
- Tymms, P. (2004) Are standards rising in English primary schools? *British Educational Research Journal*, 30 (4) 477-94.
- Vygotsky, L.S. (1978) *Mind in Society: The Development of Higher Psychological Process*. Cambridge, MA: Harvard University Press.
- Watkins, C. (2003) *Learning: A Sense-Maker's Guide*. London: Association of Teachers and Lecturers.
- Welford, G., Harlen, W. and Schofield, B. (1985) *Practical Testing at ages 11, 13, and 15*. London: DES, WO and DENI.
- White, R. T. (1988) *Learning Science*. Oxford: Blackwell.
- William, D. (2001) Reliability, validity and all that jazz, *Education 3-13*, 29 (3): 17-21.
- William, D. (2009) An integrative summary of the research literature and implications for a new theory of formative assessment, in (eds) H. L. Andrade and G. J. Cizek, *Handbook of Formative Assessment*, New York: Taylor and Francis.
- William, D. Lee, C., Harrison, C. and Black, P. (2004) Teachers developing assessment for learning: impact on student achievement, *Assessment in Education*, 11 (1) 49-66.

## Bibliographie ajoutée pour l'édition française

- Bell D. Deves R. Dyasi H. Fernandez de la Garza G. Harlen W. Léna P. Millar R. Reiss M. Rowell P. Wei Yu, 10 notions-clés pour enseigner les sciences de la maternelle à la 3ème. Le Pommier, traduit de l'anglais Harlen W (Ed) (2010) *Principles and Big Ideas of Science Education* (2009).
- Charpak G dir. (1996) La main à la pâte. Les sciences à l'école primaire, O. Jacob
- Charpak G. Léna P. Quéré Y. (2005) L'enfant et la science, l'aventure de la main à la pâte O. Jacob
- Coppens N. Munier V. Evaluation d'un outil méthodologique, le double « QCM », pour le recueil de conceptions et l'analyse de raisonnements en physique Didaskalia n°27, 2005
- Ruffenbach M. La démarche d'investigation au collège ; mission impossible... BUP n° 886(1) 2006
- Venturini P. les démarches d'investigation : enjeux pour l'enseignement et objets de recherche pour la didactique L'harmattan pp 9-13. 2012
- Vygotski L Pensée et langage (1933). La dispute, Paris 1997
- MEN - DEPP. 2009. "L'expérimentation D'un «Enseignement Intégré de Sciences et de Technologie En Classe de Sixième et Cinquième En 2008 et 2009." *Note D'information*.  
[http://media.education.gouv.fr/file/2011/70/3/DEPP-NI-2011-19-experimentation-enseignement-integre-science-technologie-college\\_201703.pdf](http://media.education.gouv.fr/file/2011/70/3/DEPP-NI-2011-19-experimentation-enseignement-integre-science-technologie-college_201703.pdf).
- MEN - DEPP. 2010. "L'évolution Des Compétences Générales Des Élèves En Fin de Collège de 2003 à 2009." *Note D'information* 10-22 (Décembre).  
[http://media.education.gouv.fr/file/2010/99/4/NIMENJVA1022\\_161994.pdf](http://media.education.gouv.fr/file/2010/99/4/NIMENJVA1022_161994.pdf).
- MEN - DEPP. 2011. "Les Compétences Des Élèves En Sciences Expérimentales En Fin D'école Primaire." *Note d'Information* 11-05 (Janvier). <http://www.education.gouv.fr/cid54836/les-competences-des-eleves-en-sciences-experimentales-en-fin-d-ecole-primaire.html>.
- MEN - DEPP. 2011. "Les Compétences Des Élèves En Sciences Expérimentales En Fin de Collège." *Note d'Information* 11-06. <http://www.education.gouv.fr/cid54837/les-competences-des-eleves-en-sciences-experimentales-en-fin-de-college.html>.
- Site de La main à la pâte. [www.fondation-lamap.org](http://www.fondation-lamap.org). Recherche avec "évaluation". Par exemple :  
<http://www.fondation-lamap.org/fr/page/14205/evaluation-des-l-ves>  
<http://www.fondation-lamap.org/fr/page/14203/evaluation-de-dispositifs>