



MODULE 2

Évaluer autrement pour s'attaquer aux effets délétères des stéréotypes

I. CE QUE NOUS APPREND LA RECHERCHE	2
A. QUELS SONT LES DETERMINANTS DE LA REUSSITE SCOLAIRE ET LEURS CAUSES ?.....	2
1. <i>Les principaux déterminants sociaux de la réussite scolaire</i>	2
a. Le sexe	2
b. L'origine sociale.....	3
2. <i>Deux sources d'explication : des causes externes, difficiles à changer et des causes sur lesquels les enseignants peuvent agir</i>	4
a. Les causes externes	4
b. Les causes sur lesquelles on peut agir	5
B. QUELS SONT LES EFFETS DES STEREOTYPES ET COMMENT L'ECOLE PEUT-ELLE LES CONTREFAIRE ?	6
1. <i>Le rôle des stéréotypes dans la réussite scolaire</i>	6
a. Nous sommes tous stéréotypés	6
i. Les parents	6
ii. Les élèves.....	6
iii. Les enseignants.....	7
b. Les effets auto-réalisateurs de ces stéréotypes	9
i. La menace du stéréotype perturbe la performance lors des évaluations et crée une désidentification à long terme	9
ii. L'effet Pygmalion rend réelles les attentes biaisées des professeurs.....	12
2. <i>L'évaluation formative pour favoriser la réussite de tous les élèves</i>	13
a. L'évaluation sommative peut, sous certaines formes, accroître les inégalités à l'école	13
b. Privilégier l'évaluation formative pour faire progresser tous les élèves	15
c. La théorie à la base de l'évaluation formative.....	19
II. PRATIQUES QUOTIDIENNES ET ACTIVITES POUR LA CLASSE	22
A. FAVORISER L'EVALUATION FORMATIVE EN CLASSE	22
1. <i>Fixer des objectifs d'apprentissage très clairs</i>	22
2. <i>Chercher des preuves de compréhension et de connaissances des élèves</i>	24
3. <i>Offrir des retours précis et détaillés concernant le travail de l'élève</i>	25
4. <i>Adapter l'enseignement et les stratégies d'apprentissage</i>	26
5. <i>Impliquer des élèves dans le processus d'évaluation et d'apprentissage</i>	27
B. LUTTER CONTRE LA MENACE DU STEREOTYPE DE MANIERE DIRECTE	28
1. <i>L'affirmation de valeurs</i>	28
2. <i>La présentation de rôles modèles</i>	28
3. <i>Intervention pour déconstruire les stéréotypes</i>	29
III. ACTIVITES POUR LA FORMATION.....	30
A. NOUS POSSEDONS TOUS DES STEREOTYPES	30
B. LA MENACE DES STEREOTYPES	32
C. QUELQUES RISQUES DES NOTES	33
LISTE DES ENCADRES.....	34
LISTE DES FIGURES	34
LISTE DES TABLEAUX.....	34
BIBLIOGRAPHIE.....	35



I. Ce que nous apprend la recherche

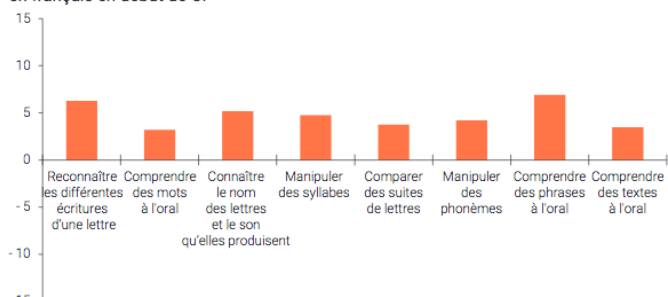
A. Quels sont les déterminants de la réussite scolaire et leurs causes ?

1. Les principaux déterminants sociaux de la réussite scolaire

a. Le sexe

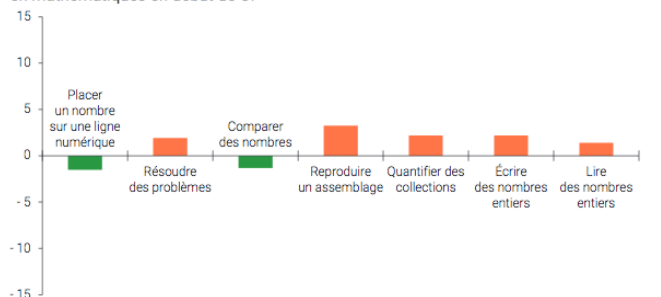
Selon la DEPP (2020), les filles présentent de meilleures performances en français que les garçons dans tous les domaines évalués en début de CP en français. En mathématiques, les filles ont une meilleure maîtrise que les garçons dans 5 domaines sur 7.

Écart entre la part des filles et la part des garçons maîtrisant les domaines évalués en français en début de CP



Lecture : à la rentrée scolaire 2019, les filles sont proportionnellement plus nombreuses que les garçons à comprendre des mots à l'oral. La différence est de 3,2 points de pourcentage.

Écart entre la part des filles et la part des garçons maîtrisant les domaines évalués en mathématiques en début de CP



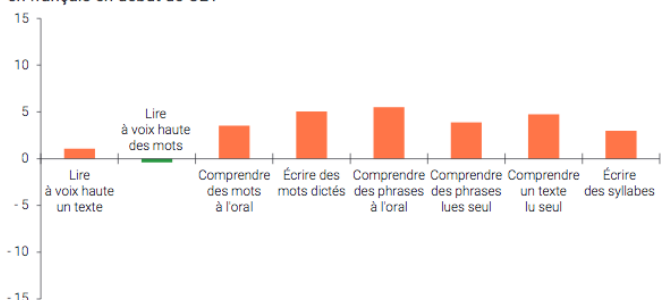
Lecture : à la rentrée scolaire 2019, les garçons sont proportionnellement plus nombreux que les filles à savoir comparer des nombres. La différence est de 1,3 point de pourcentage.

Champ : France métropolitaine + DOM - Polynésie française et Saint-Pierre-et-Miquelon, public + privé sous contrat. Source : MEN-J-MESRI-DEPP - Repères CP, septembre 2019.

Figure 1 - Ecart de compétences en mathématiques et en français entre les filles et les garçons en début de CP

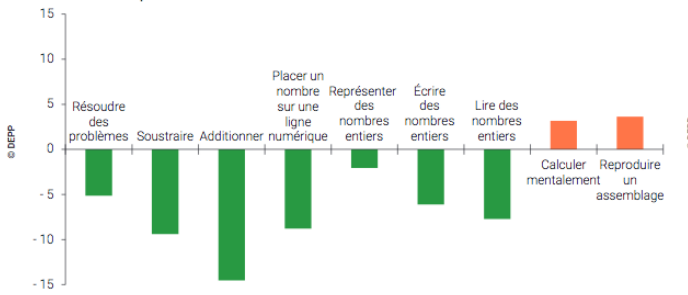
Pourtant, en début de CE1, bien que les filles continuent à présenter de meilleures performances en français que les garçons dans les différents domaines évalués, leur performance en mathématiques est dépassée par celle des garçons dans 7 domaines sur 9 avec parfois un écart considérable.

Écart entre la part des filles et la part des garçons maîtrisant les domaines évalués en français en début de CE1



Lecture : à la rentrée scolaire 2019, les filles sont proportionnellement plus nombreuses que les garçons à comprendre des phrases à l'oral. La différence est de 5,5 points de pourcentage.

Écart entre la part des filles et la part des garçons maîtrisant les domaines évalués en mathématiques en début de CE1



Lecture : à la rentrée scolaire 2019, les garçons sont proportionnellement plus nombreux que les filles à savoir additionner. La différence est de 14,5 points de pourcentage.

Champ : France métropolitaine + DOM - Polynésie française et Saint-Pierre-et-Miquelon, public + privé sous contrat. Source : MEN-J-MESRI-DEPP - Repères CE1, septembre 2019.

Figure 2 - Ecart de compétences en mathématiques et en français entre les filles et les garçons en début de CE1

Que se passe-t-il entre les années CP et CE1 en mathématiques pour les filles et les garçons ? Quels sont les facteurs qui conduisent à cette évolution dans leurs performances ?



Encadré 1. Les apports des comparaisons internationales sur les différences filles/garçons à l'école

L'enquête TIMSS est une enquête pour comparer les compétences en mathématiques des CM1 de différents pays. Les résultats de l'édition 2019 devraient arriver fin 2020. Les résultats de l'enquête 2015 constatent que les filles et garçons obtiennent le même score en sciences, mais que les garçons obtiennent un score plus élevé en mathématiques de manière significative mais peu marquée. (DEPP, 2019)

L'enquête PIRLS est une enquête pour comparer les compétences en compréhension de l'écrit des CM1 de différents pays. L'écart entre filles et garçons pour la compréhension de l'écrit est un des moins marqués : 8 points en France, contre 13 en moyenne dans l'Union européenne et 19 au niveau international. (DEPP, 2017)

L'enquête PISA est une enquête menée par l'OCDE chaque trois ans depuis 2000. Elle a comme objectif de comparer la performance des élèves âgés de 15 ans en mathématiques, lecture et sciences dans des pays à travers le monde. Les résultats de PISA 2018 montre que :

- Les scores des garçons et des filles sont comparables en culture scientifique. En revanche, en culture mathématique les performances des garçons sont supérieures à celles des filles. L'écart de score associé à la variation d'une unité de l'indice SESC (socio-économique) est de 47 points, contre 36 points pour les pays de l'OCDE.
- Les filles obtiennent un meilleur score que les garçons en compréhension de l'écrit. Leur score est en moyenne de 25 points supérieur, contre 30 points en moyenne dans les pays de l'OCDE. Depuis 2009, en France la tendance est à la diminution des écarts.¹
(OECD, 2019)

b. L'origine sociale

Le genre n'est pas le seul déterminant de la réussite. Selon la DEPP (2020), on a une répartition inégale de la proportion d'élèves en retard à l'entrée en 6^e selon la CSP des parents.

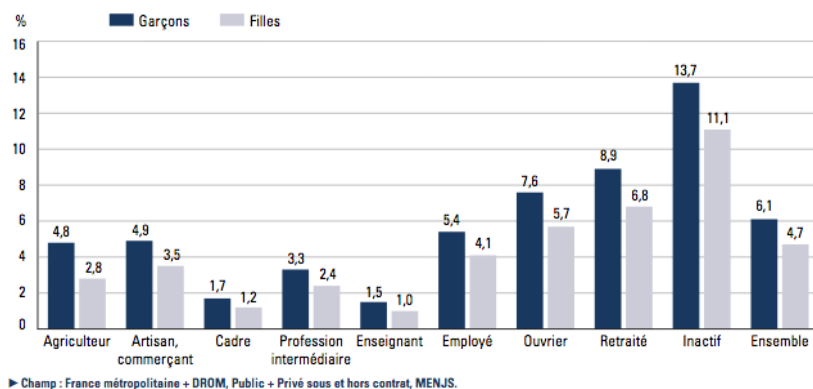


Figure 3 - Proportion d'élèves en retard à l'entrée en sixième à la rentrée 2019 selon l'origine sociale de l'élève (%)

A l'âge de 15 ans, les élèves français ont un score moyen en culture mathématique tout comme en culture scientifique qui restent parmi les plus fortement corrélés, parmi les pays de l'OCDE, avec le niveau socio-économique et culturel des familles, même si cette corrélation est en baisse (OECD, 2019). En compréhension de l'écrit, les résultats de la France sont parmi les plus fortement corrélés des pays de l'OCDE avec le niveau socio-économique et culturel des familles (107 points d'écart de score en compréhension de l'écrit en France contre 89 pour la moyenne de l'OCDE).

¹ Ces résultats montrent que l'écart se creuse entre le 3^e et le début de 6^e quand la DEPP évalue que les filles et les garçons ont le même niveau en mathématiques et les filles un meilleur niveau en lecture.



Ci-dessous, le graphique du taux de réussite au baccalauréat en 2020 selon l'origine sociale des parents et le baccalauréat, construit à partir de données de la DEPP (2020). On voit que, malgré des différences entre baccalauréats, une tendance générale se dessine, avec un plus fort taux de réussite pour les professions intellectuelles supérieures, et un plus faible taux de réussite pour les enfants d'ouvriers.

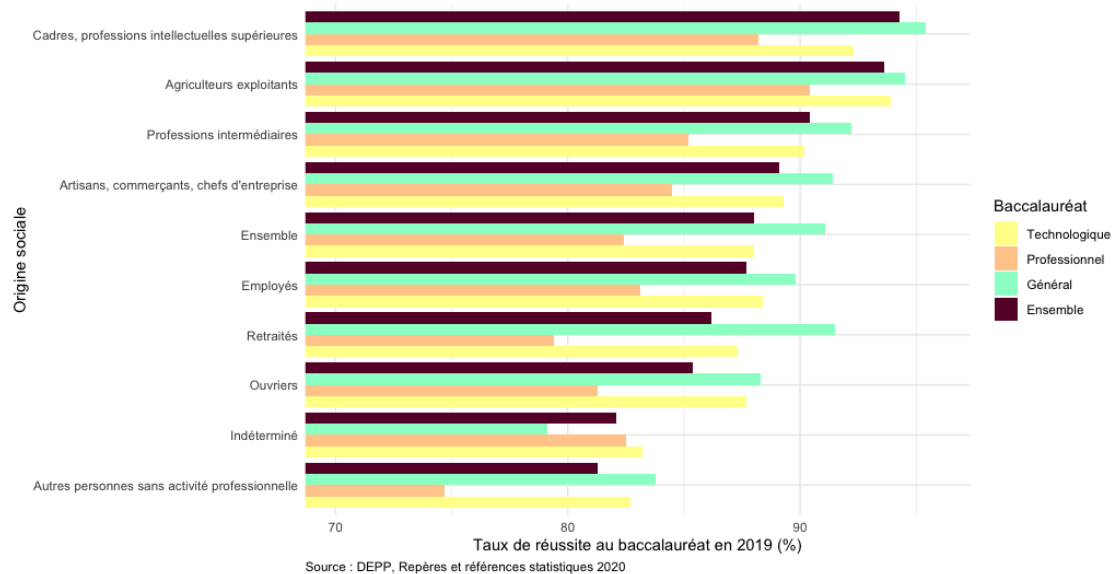


Figure 4 - Taux de réussite au baccalauréat en fonction de la CSP des parents

Hoff (2013) passe en revue la littérature sur les différences d'acquisition du langage en fonction de la classe sociale. La taille du vocabulaire, le développement grammatical mais aussi la capacité à élaborer un récit sont, dès l'âge de 3 ans, différenciés en fonction de l'origine sociale des parents, et en particulier le niveau d'études de la mère. Pour un exemple, Hart et Risley (1995) montrent qu'à 3 ans, des enfants d'origine sociale favorisée ont déjà produit 1000 mots, contre 500 mots pour les enfants d'origine sociale défavorisée. Et c'est un effet robuste que l'on trouve dans la littérature. Cela a des conséquences à long terme puisque la taille du vocabulaire à 3 ans est un bon estimateur des compétences en lecture et en orthographe, de la maternelle au CE2. **Il y a donc des déterminants mais tous ne semblent pas prédéterminés dès l'entrée à l'école puisque cette différence de performance ne semble pas caractériser les élèves en début de CP pour le genre notamment.**

2. Deux sources d'explication : des causes externes, difficiles à changer et des causes sur lesquels les enseignants peuvent agir

a. Les causes externes

Pour identifier les causes sur lesquels les enseignants et autres peuvent agir, il faut d'abord identifier celles qui sont innées ou non modifiables. Plomin et Von Stumm (2018) s'intéressent, dans une revue de littérature, à l'estimation de l'impact de l'héritage génétique sur la réussite scolaire. Ils estiment que l'on pourrait expliquer 11 à 13% des différences de réussite scolaire à partir des gènes. Cependant, il faut bien noter que les scientifiques sont divisés sur le sujet². Ceux qui défendent une héritabilité génétique insistent sur deux faits :

1. L'environnement joue un rôle très important. En effet, des facteurs génétiques peuvent plus ou moins s'exprimer selon l'environnement dans lequel se développe l'enfant.
2. Cette part génétique explique très peu de variance comparativement à l'influence de l'environnement.

De la même manière, les capacités cognitives sont déterminées par les **premières années de la vie** et de la stimulation reçue durant ces années. Par exemple, les capacités acquises précocement en langage oral ont une

² Il est notamment très difficile de distinguer la variabilité génétique de facteurs familiaux, même environnementaux. En effet, l'héritage génétique des enfants est hérité par ses gènes mais aussi par son environnement familial, lui-même déterminé par l'éducation des parents et leurs ressources, déterminées par leurs capacités cognitives, elles aussi déterminées par les gènes



influence sur l'apprentissage et la maîtrise de la lecture. L'étude de Hart et Risley (1995) aux Etats-Unis a souligné un écart considérable de développement du langage selon l'origine sociale des enfants : les enfants des familles vulnérables ont entendu 30 millions de mots de moins que les enfants issus de milieux socio-culturels favorisés et maîtrisent trois fois moins de mots. Ils montrent que c'est souvent lié à un niveau d'interactions différentes : les mères de milieu social défavorisé s'adressent souvent moins à leur enfant directement, et notamment pour seulement discuter. Des programmes, comme Parler Bambin, tentent de compenser ce déséquilibre dès la crèche.

Le contexte familial explique aussi les différences d'orientation. Guyon et Huillery (2016) expliquent notamment qu'un des facteurs de la différenciation d'orientation est le manque d'informations sur l'orientation. Les ressources (capital culturel et capital économique de Bourdieu) auxquelles les familles ont accès renforcent aussi ces différences (cours particuliers, culture familial).

b. Les causes sur lesquelles on peut agir

Guyon et Huillery (2016) se sont demandées pourquoi les enfants d'origine défavorisée font de moins grandes études. Elles s'intéressent aux aspirations différentielles des élèves de collège. Outre le manque d'information, elles mettent en avant un biais d'autoévaluation et autocensure en ce qui concerne :

- Les différentes stratégies académiques utilisées et leur efficacité³ ;
- Les biais dans la perception du retour sur investissement dans l'effort ;
- L'influence de la famille.

Elles montrent que les élèves surestiment le lien entre milieu social et réussite académique ultérieure. Parmi les élèves défavorisés, les immigrés sont plus ambitieux que les non-immigrés. Les aspirations à la fin de la 3^e sont corrélées aux résultats scolaires de l'année et au brevet. Cependant, les élèves ne sont pas toujours conscients du lien entre choix d'orientation et perspectives d'emploi.

D'autres explications peuvent être mises en avant. Quinn (2018), dans une méta-analyse, met en évidence le fait que les garçons sont plus souvent diagnostiqués avec des troubles de la lecture, peut-être à cause de troubles comportementaux plus fréquents (biais de constatation).

Ainsi, contrairement à l'étude de Benbow et Stanley (1980), il semble que des sciences cognitives à l'économie, en passant par la sociologie, la recherche tende à montrer des déterminants sociaux qui créent un biais **d'autocensure et d'autoévaluation**, perpétuant des désavantages sociaux. Il y a une intériorisation du stéréotype :

- Chronique : par une perception différenciée de ses aptitudes scolaires :
 - Sentiment d'efficacité personnelle (Gecas, 1989)
 - Estime de soi⁴
 - Peur d'échouer
- Fonctionnelle : Certains contextes conduisent les individus à répondre de manière conforme à ce que les stéréotypes prescrivent

Les stéréotypes, qui sont seulement des croyances sur la réussite, semblent donc avoir une part importante dans la détermination de la réussite concrète des élèves

³ Callan et al. (2017) montrent que le CSP de la famille et de l'école sont positivement corrélées à l'usage de stratégies académiques efficaces, qui sont reliées à la réussite en lecture, en sciences et en maths.

⁴ Von Soest et al. (2018) montrent à nouveau un fort lien entre SES et l'estime de soi avec un panel d'adultes norvégiens, un effet validé par la littérature (notamment par une méta-analyse de Twenge et Campbell (2001)). Ils trouvent aussi une estime de soi plus importante chez les garçons que chez les filles. Des expériences de vie, comme le handicap ou le chômage, ont également un impact important.



B. Quels sont les effets des stéréotypes et comment l'école peut-elle les contrefaire ?

1. Le rôle des stéréotypes dans la réussite scolaire

a. Nous sommes tous stéréotypés

i. Les parents

De nombreuses études (Lummis & Stevenson, 1990 ; Eccles & Jacobs, 1986 ; Andre et al., 1999) évoquent **une influence des parents sur la construction des attentes et stéréotypes de leurs enfants**. Pour Eccles et Jacob, cela viendrait des médias qui affectent les croyances des mères dans les capacités de leurs filles. Andre et al. (1999) notent que même en maternelle, les parents perçoivent les garçons meilleurs en math que les filles. La science est souvent présentée comme plus importante pour les garçons. Notamment, Yee et Eccles (1988) font passer des questionnaires à des parents de 48 élèves âgés de 12 à 14 ans (7th grade). Ils montrent que les parents attribuent plus les bons résultats en mathématiques au travail fourni par leur fille qu'au talent, mais les attribuent plus au talent inné pour leurs fils. Lorsque leurs enfants ont de moins bons résultats, les parents attribuent cela à un manque de travail, que ce soient leurs fils ou leurs filles.

ii. Les élèves

Maass et Cadinu (2003) montrent que les filles pourraient souffrir de la menace du stéréotype dès 5-6 ans vis-à-vis de leur performance en mathématiques, dont l'apogée serait atteinte durant le collège. Cela passerait par une anxiété, des pensées intrusives, une éviction du risque et, sur le long terme, un désengagement et une baisse de l'effort. **La date d'apparition des premiers effets de la menace du stéréotype fait débat mais de nombreuses études semblent trouver une première empreinte des stéréotypes, même implicite, dès l'âge primaire.**

De nombreuses études ont étudié l'émergence des stéréotypes chez des élèves en école primaire (de la Grande section à la 6^e). Au **niveau implicite**, toutes les études trouvent un niveau d'association des filles à la lecture et des garçons aux mathématiques, dès le CP. Au **niveau explicite**, Del Rio & Strasser (2013) trouvent des traces d'attentes stéréotypées à propos de la réussite académique dès l'âge de 5 ans, mais Martinot et al. (2012) n'en trouvent pas chez des enfants de CM1. Une perception différenciée semble émerger au cours de l'école primaire. Andre et al. (1999) notent une perception différenciée genrée des compétences à partir du CM1, ainsi qu'un goût différent aussi en CM1. Cvencek et al. (2011) ont des résultats qui laissent à penser que l'identité de genre et la conscience des stéréotypes arrivent dès le CP-CE1 mais que la conception de soi liée aux mathématiques émergent plus tard. On peut noter que l'étude de Martinot et al., qui ne trouvent pas de trace de stéréotype explicite, est la seule étude réalisée en France. Cependant, Lummis & Stevenson (1990) se sont intéressés aux différences interculturelles entre pays au niveau de la performance en maths et lecture de la maternelle au CM2. Ils montrent que les différences observées entre garçons et filles sont relativement stables d'un pays à l'autre, et que les croyances sur les capacités ont une influence plus forte que le vrai impact de ces capacités.

On peut noter **un phénomène de stratification du stéréotype** (Régner et al., 2014 ; Martinot et al., 2012) pour lequel les enfants dissocient les stéréotypes sur les adultes (hommes et mathématiques, femmes et lettres) des stéréotypes sur les enfants. En effets, ils se considèrent comme également doués, voire ont une tendance à valoriser leur propre groupe, vers le CM1, dans une sorte de chauvinisme de genre.

Plus tard dans leur scolarité, les élèves connaissent explicitement ces stéréotypes. Ainsi, Muntoni et al. (2020) s'intéressent aux stéréotypes liés à la lecture. Ils montrent que les filles ont des stéréotypes forts concernant la lecture (sur des CM2-6^e en Allemagne). Guyon et Huillery (2016), par des questionnaires auprès d'élèves de collège, notent que **tous les élèves surestiment le lien entre succès académique futur et milieu social**.



iii. Les enseignants

De nombreuses études ont montré des **traitements différenciés par les professeurs, inconsciemment**. Un rapport du National Women Service du Chili (SERNAM) de 2008 a montré que les enseignants de grande section interagissent plus avec les garçons que les filles, et ce, non pas seulement à cause des réponses spontanées plus fréquentes des garçons [Citation]. De plus, les enseignants auraient une attitude plus paternaliste envers les filles, leur procurerait plus d'aide et donneraient des exemples stéréotypés. Sadker et Sadker (1986) notent aussi que les enseignants accordent plus d'attention aux garçons, qu'aux filles, auxquelles on demande plus de passivité et de conformité. Cela recoupe l'observation de Grésy et Georges (2013) sur la crèche qui étudient les crèches françaises. Selon leurs observations, le personnel de la crèche demande plus souvent son avis aux garçons qu'aux filles, coupe plus souvent la parole aux filles. Il demande plus souvent aux filles de faire attention à leur apparence et tolère moins que les filles se mettent en colère. Le personnel incite plus les garçons à participer à des jeux collectifs. Le personnel valorise la modestie et la grâce chez les filles et l'esprit de compétition chez les garçons. Duru-Bellat (1990), par un travail sur l'école primaire, montre qu'en moyenne, **un enfant est classé selon son sexe vingt fois par jour. Il s'agit d'un rappel à l'ordre symbolique : il y a des comportements attendus pour les garçons et pour les filles**. Les filles sont davantage perçues comme un groupe indifférencié (« les filles, arrêtez de parler »), alors que les garçons sont plus souvent individualisés (appelés par leur prénom). En moyenne, les enseignants consacrent 44% de leur temps individualisé aux filles et 56% aux garçons. En mathématiques, les deux tiers des interactions verbales des professeurs sont avec les garçons.

Des préférences genrées pour certaines matières scolaires sont influencées par les stéréotypes. Gentrup et Rjosk (2018) notent que certains élèves de CP sont plus ou moins vulnérables aux estimations des profs en maths (c'est-à-dire les filles), mais pas en lecture. Lavy et Sand (2015) suivent une cohorte d'élèves israéliens de l'entrée en sixième à la fin du lycée. En plus de la notation dans la classe, les élèves passent des tests corrigés anonymement. Au collège, en mathématiques, les filles obtiennent en moyenne de meilleures notes que les garçons quand elles sont corrigées anonymement et de moins bonnes notes quand elles sont corrigées par leurs professeurs. Paradoxalement, il n'y a pas de biais de notation dans les matières littéraires.

Cela passe aussi par le biais des appréciations selon le travail de Marie Duru-Bellat (1990) dans l'école primaire. Par une observation de la salle de classe, elle met en avant que quand les filles sont mauvaises dans les matières scientifiques, les enseignants considèrent plus souvent qu'elles ne sont pas douées pour les sciences : cette absence de résultats serait liée à leur nature et non à la socialisation. Quand un garçon a de mauvais résultats dans les matières scientifiques, l'enseignant considère plus souvent qu'il n'a pas suffisamment travaillé.

Des études à travers le monde montrent également que **le traitement des élèves dépend aussi de ses origines ethnique et sociale**. Alesina et al. (2018) travaillent sur des élèves et enseignants de 4^e en Italie. A performance constante (calculée par des tests standardisés), les enfants d'origine immigrée reçoivent des notes plus basses quand ils sont notés de manière non anonyme. Cet effet se retrouve principalement sur les professeurs de mathématiques qui ont un score IAT élevé. Quel que soit le score, elle ne trouve pas d'effet pour les professeurs de littérature. De la même manière, Botelho et al. (2015), au Brésil, montrent un traitement différencié en 4^e qui résultent sur une sous-notation des Noirs. **Ils notent que les professeurs utiliseraient un modèle de discrimination statistique avec un « type d'élève » qui se construirait le long de la scolarité**, ce qui devient d'autant plus important que les parents se reposent sur les professeurs pour l'expertise scolaire.



Encadré 2. Une étude qui montre l'impact des stéréotypes d'origine sociale des enseignants sur la notation

Une étude de Autin et al. (2019), s'est intéressée à l'effet de la classe sociale sur la notation dans le cadre d'une évaluation formative ou une évaluation normative). Ainsi, ils donnent, dans deux expériences, des dictées d'élèves en demandant d'abord de souligner les fautes puis, soit de donner une note dans le cadre d'une évaluation normative soit de commenter les fautes dans le cadre d'une évaluation formative. Ils manipulent la classe sociale, c'est-à-dire qu'ils modifient le profil de l'élève pour le faire appartenir à une classe sociale favorisée ou défavorisée. Ils montrent que **les erreurs mènent à une évaluation plus négative quand l'élève est de classe sociale défavorisée**.

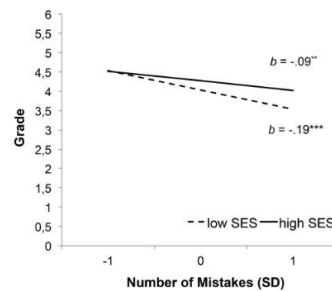


Figure 5 - Expérience 1 : lien entre nombre d'erreurs et la note comme fonction de l'origine sociale de l'élève

Dans la seconde expérience, ils ne font pas annoter les copies mais arrêtent l'expérience après que les fautes aient été soulignées. Ils trouvent un effet significatif de l'objectif de l'évaluation. Quand les participants sont dans le cadre d'une évaluation normative, **ils soulignent plus de fautes pour les enfants de classe sociale défavorisée que favorisée**. Cet effet n'est pas significatif pour l'évaluation formative.

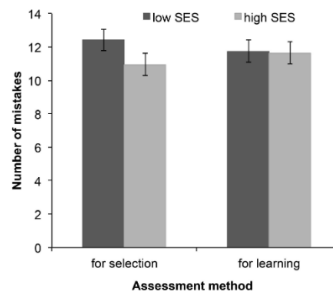


Figure 6 - Expérience 2 : Nombre de fautes trouvées dans la dictée comme une fonction de la classe sociale de l'élève et du type d'évaluation

Enfin, dans une troisième expérience, ils cherchent à savoir pourquoi cet effet existe. Ils rajoutent une variable : **l'évaluation sert soit à classer les élèves, soit à leur apprentissage**. Ils montrent que cet effet remplace l'effet de l'évaluation formative vs normative.

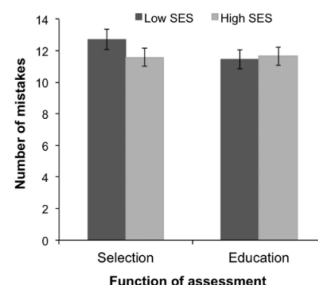


Figure 7 - Expérience 3 : Nombre d'erreurs trouvées dans la dictée en fonction de la classe sociale de l'élève et de la fonction de l'évaluation

Dans une méta-analyse sur leurs expériences, ils trouvent **un effet petit mais significatif du type d'évaluation sur l'écart de notation entre classes sociales**.



Encadré 3. Une étude qui montre l'impact des stéréotypes d'origine sociale sur l'orientation scolaire

Les mêmes auteurs (Batruch et al., 2019) réalisent aussi une étude similaire vis-à-vis des conseils en matière d'orientation scolaire. Ainsi, ils mènent deux études (l'une sur des étudiants, l'une sur des enseignants) en Suisse. A partir d'un même bulletin, ils font varier le genre et la classe sociale (par des indices sur l'élève) et demandent aux participants de dire, pour différents cursus scolaires, à quel point l'élève convient à ce parcours. Ils observent que les enfants de classe défavorisée sont notés comme significativement plus adaptés à des carrières éducatives plus petites que les enfants de classe favorisée ; et que les enfants de classe favorisée le sont significativement plus pour des plus grandes carrières éducatives plus grandes (résultats répliqués avec les étudiants et les enseignants).

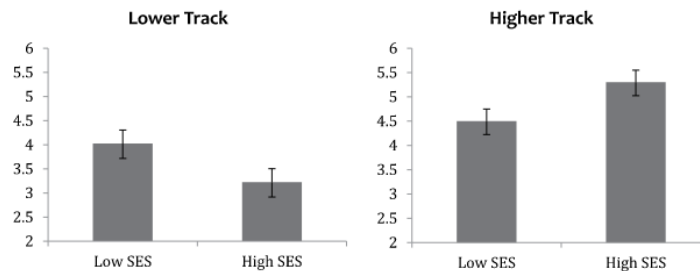


Figure 8 - Adéquation moyenne d'après les participants pour des carrières éducatives petites ou grandes en fonction de la classe sociale de l'élève cible

Dans la même lignée, ils s'intéressent à la fonction de l'école et de l'évaluation comme une explication d'une partie de cette différence. Ainsi, ils rajoutent un facteur de présentation de l'école qui doit, soit classer les élèves (Sélection), soit les aider à progresser (Education). Ils montrent que seulement changer la présentation de l'évaluation et de l'école réduit cet écart, que ce soit pour les grandes ou petites carrières éducatives.

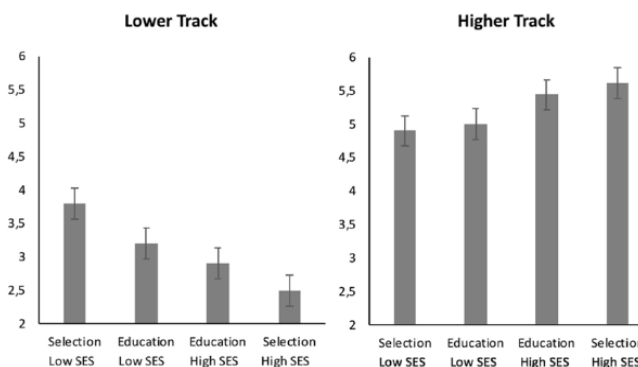


Figure 9 - Adéquation moyenne d'après les participants pour des carrières éducatives petites ou grandes en fonction de la classe sociale de l'élève cible et la fonction de l'école

b. Les effets auto-réalisateurs de ces stéréotypes

i. La menace du stéréotype perturbe la performance lors des évaluations et crée une désidentification à long terme

La menace du stéréotype constitue une littérature abondante depuis les années 1990. Le terme est issu de l'article de Steele et Aronson (1995) qui s'intéressaient initialement à l'explication de la différence de réussite entre étudiants Noirs et Blancs lors de tests standardisés. Dans leur étude, ils font passer un test à un ensemble d'étudiants en présentant le test, soit comme diagnostique de l'intelligence, soit en ne disant rien. Ils notent que les étudiants Afro-américains voient leur performance réduite lorsque le test est présenté comme une mesure diagnostique. La menace du stéréotype passe pour eux par : (1) une stimulation à **court terme** qui réduit la gamme d'indices que les participants peuvent utiliser, une dispersion de l'attention du fait des craintes provoquées par la peur de valider un stéréotype négatif, et une interférence de la conscience de soi lors du test ; et (2) une désidentification à long terme vis-à-vis du sujet à propos duquel l'individu subi le stéréotype.



Encadré 4. La menace du stéréotype ne se limite pas aux membres des groupes historiquement discriminés

De nombreux groupes sociaux ont été étudiés vis-à-vis de cette menace, en essayant de répondre à la question : *est-ce nécessaire de faire partie d'un groupe historiquement discriminé (les Afro-américains notamment) ou est-ce que cet effet peut toucher toute personne affectée par un stéréotype ?* Deux études semblent apporter une réponse intéressante à cette question.

Les hommes peuvent être subi à une menace du stéréotype non cognitif

Leyens et al. (2000) étudient l'effet du stéréotype non cognitif sur la capacité des étudiants américains hommes à gérer l'information émotionnelle. Lorsque les hommes sont mis face à une information stéréotypique, leur capacité se réduit. Cela passe par une stratégie de « zéro coup raté » c'est-à-dire que les hommes font plus de fautes car ils surinterprètent les informations émotionnelles.

Le sexe et l'origine ethnique peuvent générer des résultats mixtes quant à la menace du stéréotype

Ambady et al. (2001) se penchent sur les stéréotypes de genre en maths à l'école primaire et au collège en combinaison avec des stéréotypes ethniques. Ils trouvent des signes de menace de stéréotype dès la maternelle, à 5 ans. Ils exposent les enfants soit à un renforcement de leur identité asiatique (stéréotype positif vis-à-vis des maths), soit à un renforcement de leur genre (stéréotype positif ou négatif selon garçon/fille), soit à aucun renforcement. Ils montrent que quand l'identité asiatique est activée, les filles réussissent mieux que quand aucune identité est activée, qui est une meilleure performance que lors de l'activation du genre. Dans le cadre des garçons, ils réussissent mieux lorsque leur identité de genre ou ethnique est activée que lorsqu'aucune identité n'est activée. Ils ne trouvent pas de conscience explicite des stéréotypes mais des associations implicites entre origine asiatique et maths et garçons et mathématiques, chez les garçons comme chez les filles.

Il est important de se pencher sur cette menace puisqu'elle touche les enfants dès l'âge primaire, et par différents stéréotypes. Ci-dessous quelques exemples qui peuvent s'appliquer à la classe primaire.

Les filles et les mathématiques

Les méta-analyses (Stoet & Geary, 2012; Flore & Wicherts, 2015) s'accordent sur un effet de menace de stéréotype sur les performances des filles en maths.⁵ Sur l'école primaire, on peut noter l'étude de Muzzati et Agnoli (2007) en Italie. Ils notent que les différences de performance entre garçons et filles apparaissent au collège mais les différences de confiance en soi apparaissent avant. Ainsi, Ils montrent que la conscience en soi des filles et garçons et la difficulté qu'ils perçoivent par rapport aux mathématiques est identique en CP mais supérieur pour les garçons à partir du CE2. De la même manière, ils semblent que les stéréotypes de genre liés aux mathématiques apparaissent dès le CM1. A partir du CM2, les filles disent que les maths sont moins utiles, s'appliquent moins et aiment moins les maths. Les auteurs ne trouvent cependant pas de trace de menace du stéréotype sur les performances. **Ainsi, même si on n'observe pas encore de réussite différentielle, c'est dès l'âge primaire que se forment les bases des stéréotypes.**

Les garçons et la lecture

C'est en quelque sorte le pendant masculin du stéréotype précédemment mentionné. Selon l'enquête PISA 2018, les filles réussissent mieux en compréhension de l'écrit que les garçons (+25 points). Pansu et al. (2016), avec des élèves de CE2 français, montrent que la menace du stéréotype affecte aussi la performance en lecture des garçons, qui réussissent moins bien quand le test est présenté comme mesure d'habilité).⁶ Il convient tout de même de faire attention à cette expérience menée sur seulement 4 classes.

⁵ Cependant, ils mettent en garde à propos d'un effet potentiellement surestimé par un biais de publication, et à propos du fait qu'ils ne trouvent pas un effet significatif de l'âge.

⁶ On peut noter que cette expérience a été menée sur seulement 4 classes, et il y a globalement un problème d'ordre de grandeur des études citées qui concernent souvent un petit nombre de classes ou d'élèves.



L'origine ethnique

Des études d'Alter et al. (2010), d'Aronson et al. (2002), Cohen et al. (2006) ou encore Sherman et al. (2013) rapportent une menace de stéréotype quant à l'origine ethnique (Afro-américains pour les trois premiers, latino-américains pour le dernier). Cet effet est étudié à l'université (Aronson et al., 2002) mais aussi au collège (les trois autres études étant menées du CM2 à la 3^e). En dehors de ces études américaines, Chateignier et al. (2009), se sont intéressés à **la menace du stéréotype sur les étudiants franco-arabes en France**. Ils montrent qu'ils souffrent de cette menace puisque, quand le test est présenté comme diagnostique de l'intelligence, ils réussissent moins bien que les français non issus de l'immigration, mais aussi moins bien que dans la condition non diagnostique, où leur performance égale celle des français non issus de l'immigration.

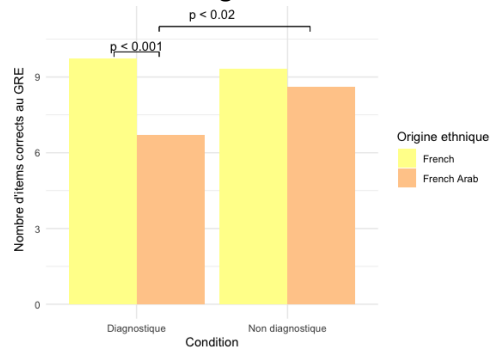


Figure 10 - Nombre d'items corrects en fonction du groupe expérimentale et l'origine ethnique (Chateignier et al., 2009)

Les enfants de milieux défavorisés

En France, Croizet et Dutrévis (2004) se sont penchés sur l'**effet Volvo** (le lien entre classe socio-économique et résultats aux tests standardisés). Parmi d'autres explications (moins d'accès à l'éducation préscolaires, aux ressources scolaires, moins d'intérêts pour le milieu académique, intelligence génétique), ils montrent qu'il existe un effet de menace du stéréotype. Ainsi, quand le test est décrit comme une mesure diagnostique de l'intelligence les étudiants de milieux défavorisés, qui font l'objet d'une réputation de plus faibles capacités, réussissent moins bien alors qu'ils réussissent tout aussi bien que les autres en l'absence de cette description. Ils mettent en avant le **rôle discriminatoire de la situation de test**. De la même manière, dans une étude antérieure, Croizet et Claire (1998) avaient montré que des étudiants d'université de milieux défavorisés réussissaient moins bien un test standardisé que ceux d'un milieu favorisé lorsque le test était présenté comme une mesure de capacités intellectuelles, alors qu'ils obtenaient les mêmes résultats dans la condition contrôle. Hoff et Pandey (2004) mettent en place une étude en Inde sur des élèves en 6^e et 5^e. La différence de performance observée entre les castes serait pour eux due à une menace du stéréotype.

Cette menace a donc un effet sur les performances des élèves, reproduisant les inégalités supposément existantes. Walton & Spencer (2009) dans une méta-analyse, parle de « **capacité latente** » pour les groupes **stéréotypés**. Ils trouvent une sous-performance significativement plus faible pour les groupes stéréotypés (ethniques, ou genre, dans 39 études différentes de 5 pays différents), mais aussi une performance égale lorsque le stéréotype est levé.

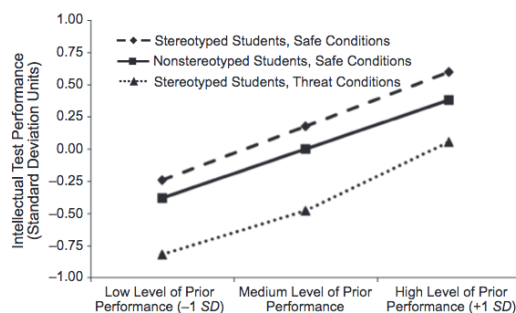


Figure 11 - Performance à un test d'intelligence d'élèves stéréotypés et non stéréotypés en fonction de la condition de menace de stéréotype et de la performance passée



Ainsi, comme présenté par Cohen et al. (2006), un effet négatif des stéréotypes se traduit en termes de notes, donc en de moins bons résultats qui eux-mêmes agissent en validant à leur tour le stéréotype négatif : c'est un cercle vicieux.

Les effets de pairs semblent renforcer cet effet. Salikutluk & Heyne (2017) montrent sur des élèves de 3^e en Allemagne que **plus les élèves (garçons ou filles) adhèrent aux normes de la masculinité, moins ils ont de bonnes notes**. Ils supposent que c'est en raison de la menace du stéréotype chez les filles, et de l'effet « mauvais élève » (transmis par ce même stéréotype) chez les garçons. Et, de plus, plus ils adhèrent à ces normes, et plus l'écart entre garçons et filles s'agrandit. Il y a donc une influence de l'adhésion des pairs aux normes de la masculinité. **Il existerait aussi une relation significative et négative entre les résultats en lecture des garçons et les stéréotypes de leurs camarades** (Muntoni et al., 2020).

ii. L'effet Pygmalion rend réelles les attentes biaisées des professeurs

Les professeurs **par la présentation de modèles stéréotypés** (SERNAM, 2008 ; Sadker & Sadker, 1986) par exemple, ou **par une prophétie auto-réalisatrice** contribuent aussi à ce phénomène (effet Pygmalion). C'est à propos de ce point que nous cherchons à agir. Les professeurs agissent de deux manières : par le renforcement du stéréotype conscient (modèles stéréotypés notamment), ou par une évaluation ou soutien faussé lié à ses stéréotypes qui agit sur la perception de ses aptitudes scolaires.

Sans leur transmettre directement un stéréotype, c'est **aussi par des pratiques différenciées, en classe ou dans la notation**, qu'ils peuvent contribuer à la reproduction des stéréotypes grâce à une prophétie auto-réalisatrice (Merton, 1948). Duru-Bellat (1990) met en avant qu'il y a des comportements attendus en fonction du sexe, et que les parcours se construisent selon les stéréotypes de sexes. Les garçons sont censés être plus doués et préférer les matières scientifiques. Les filles doivent être moins curieuses, plus sensibles et préférer les matières littéraires. Ces attentes ont des effets de réel : les enseignants considèrent comme moins grave que les filles ne suivent pas en sciences.

L'**effet Pygmalion** est débattu depuis sa première occurrence dans l'article de Rosenthal et Jacobson (1968). Si l'on peut s'interroger sur la magnitude de cet effet, Jussim et Harber (2005), dans une revue de la littérature, concluent que les prophéties auto-réalisatrices existent mais ont de faibles effets, ne s'accumulent pas et ont tendance à se dissiper. Cependant, certaines peuvent être plus puissantes pour des groupes sociaux stigmatisés. Ainsi, Jussim et al., 2009 montrent qu'il est nécessaire de cumuler trois conditions pour parler d'effet Pygmalion ou de prophétie auto-réalisatrice dans la salle de classe :

- Le professeur doit avoir de fausses attentes par rapport aux élèves ;
- Le professeur change la manière dont il traite les élèves selon ses attentes ;
- Les élèves réagissent en confirmant ses attentes.

Encadré 5. Un article fondateur : *Pygmalion dans la classe* de Rosenthal et Jacobson

Dans cet article, les auteurs mènent une étude dans des écoles primaires aux États-Unis. En fin d'année, ils font passer aux élèves un test présenté aux enseignants comme destiné à mesurer l'intelligence des élèves. Au début de l'année suivante, l'enseignant reçoit les résultats de ces tests, mais les élèves désignés comme supérieurement intelligents sont choisis de manière aléatoire. Le but est de créer deux groupes dans la classe : un groupe que l'enseignant pense supérieurement intelligent et de l'autre un groupe témoin.



Or, en fin d'année scolaire, ceux qui ont été désignés par l'expérience ont plus progressé que la moyenne. Ils expliquent cela par le fait que les enseignants ont un certain nombre d'attentes sur le travail des élèves qui déforment la perception et influence la notation. En retour, un élève qui se sent valorisé tente de répondre aux attentes de l'enseignant tandis qu'un élève qui se sent stigmatisé (comme mauvais élève) se sent dévalorisé et peut chercher à endosser ce rôle de mauvais élèves. Les enseignants se font par avance une idée des capacités des élèves. Ces attentes peuvent varier selon un certain nombre de paramètres : comportement, 1e copie... Le risque est que les attentes peuvent être liées à des caractéristiques sociales : il y a un impact des stéréotypes sur les attentes des professeurs (sexe, origines ethniques et sociales...). L'effet Pygmalion peut être source de discrimination parce que ce sont ceux qui ont de plus grandes chances d'obtenir de bons résultats scolaires qui suscitent les plus grandes attentes. L'effet Pygmalion est un effet d'autoréalisation des attentes.

Deux études trouvent aussi un impact de ces prophéties auto-réalisatrices sur les garçons (Muntoni & Retelsdorf, 2018 ; Retelsdorf et al., 2015). **L'image de soi par rapport à la lecture est plus dégradée pour les garçons de CM2 dont les enseignants ont des stéréotypes plus forts.** De la même manière, les attentes des profs de CM2 sont corrélées aux résultats en lecture de leurs élèves, ce qui peut impliquer soit une précision de la prévision, soit un effet auto-réalisateur. **On peut noter que les enseignants ont des attentes plus élevées pour les filles que pour les garçons**, même si ce lien varie selon le niveau de stéréotypes détenus par l'enseignant (fort ou non).

Cet effet Pygmalion a des résultats sur l'apprentissage et la poursuite d'études. Lavy et Sand (2015) montrent que la notation différenciée entre filles et garçons donnée par les professeurs de 6^e ont des conséquences sur le lycée : **toutes choses égales par ailleurs, plus une fille a eu des professeurs biaisés au collège, moins elle a de chances de s'orienter vers les filières scientifiques au lycée. L'effet est inverse pour les garçons.** À la fin du lycée, les garçons sont effectivement meilleurs que les filles en mathématiques. Il est intéressant de voir que l'effet Pygmalion passe par un processus inconscient, latent. Les professeurs sous-notent et discriminent sans le savoir, même s'ils s'opposent à la discrimination dans leurs discours. De la même manière, Duru-Bellat (1990) montre que les différences d'appréciation scolaire créent des stéréotypes de sexe qui construisent des réussites scolaires différentes pour les filles et les garçons. À note équivalente, on considère comme moins grave qu'une fille s'oriente vers la filière littéraire plutôt qu'un garçon.

Cette prophétie auto-réalisatrice a des effets de long terme, impactant la valeur assignée aux matières (Anderman et al., 2004) mais aussi la réussite à long terme. Papageorge et al. (2019) montrent que les attentes des profs au lycée ont un impact sur la réussite à l'université, limitant le risque d'abandonner l'université, augmentant les notes, le temps passé à travailler ainsi que les attentes des élèves. Lavy et Sand (2015) montrent aussi que l'évaluation positive (ou négative) par le prof a un effet positif (ou négatif) sur la réussite future des garçons (filles) en maths.

Il y a donc une nécessité d'agir sur ces stéréotypes.

2. L'évaluation formative pour favoriser la réussite de tous les élèves

a. L'évaluation sommative peut, sous certaines formes, accroître les inégalités à l'école

L'évaluation sommative utilise une métrique commune afin d'indiquer le niveau de performances des élèves. La métrique la plus classique est la note chiffrée, mais elle peut aussi reposer sur des lettres, des couleurs, ou toute autre échelle de valeurs comme des smileys. **L'évaluation sommative a pour fonction l'attestation des apprentissages.**

Selon la littérature sur la menace des stéréotypes, telle que détaillée dans la première partie de cette revue, c'est la présentation de l'évaluation comme une mesure diagnostique de l'intelligence ou des capacités intellectuelles de l'élève qui conduit à des résultats différenciés selon le genre, l'origine socio-économique ou



l'origine ethnique (voir notamment la revue de littérature de Walton & Spencer). Un environnement qui favorise la comparaison des élèves – entre eux ou avec une mesure externe – risque de produire des effets négatifs sur leur motivation et *in fine* leurs apprentissages.

Dans une revue de la littérature, Harlen et Deakin (2002)⁷ trouvent que les élèves qui obtiennent de mauvais résultats (quelle qu'en soit la raison) sont attrapés dans un cercle vicieux et sont doublement désavantagés par l'évaluation sommative. En effet, être qualifié d'échec n'a pas seulement un impact sur les sentiments actuels d'un élève concernant sa capacité à apprendre, mais **diminue encore son estime de soi**, déjà faible, et réduit son effort futur et sa probabilité de réussir. Après l'introduction des National Curriculum Tests en Angleterre pour les élèves dans la deuxième année (équivalent au cours préparatoire en France), les moins performants ont vécu une baisse significative de leur estime de soi par rapport aux élèves les plus performants, alors qu'auparavant il n'y avait pas de corrélation entre l'estime de soi et les résultats (Davies & Brember, 1998).

La nature de la motivation des élèves semble aussi jouer un rôle déterminant. Pulfrey et al. (2013) démontrent que pour 89 collégiens, la présence de note réduit l'intérêt pour le contenu en faveur d'un focus sur la note, ce qui a comme conséquence la **réduction de la motivation intrinsèque** de l'élève. Selon Butler (1988), lors d'un exercice noté, l'ego devient la source principale de motivation. Les bons élèves gardent un haut niveau d'engagement tant qu'ils anticipent de futures notes car ils sont en mesure de réaffirmer leurs capacités, à l'inverse l'engagement des moins bons élèves s'effondre à l'introduction d'une note chiffrée. Pour les plus et moins performants, on observe un **effet négatif des notes sur l'engagement à des activités créatives**.

L'**auto-handicap** est aussi un mécanisme possible. Jones et Berglas (1978) le définissent comme une stratégie de protection de l'estime de soi lorsque la valeur personnelle ou les compétences d'un individu doivent être démontrées. Pour éviter un possible échec, l'individu s'inflige des obstacles et des difficultés externalisant ainsi la raison de l'échec, le rendant encore plus probable. Ainsi, Zuckerman et al. (1998) montrent que les étudiants recourant le plus aux stratégies d'auto-handicap obtenaient de moins bons résultats à résultats académiques antérieurs égaux. Les constats de Harlen et Deakin (2002) semblent suggérer que l'auto-handicap s'installe au fur et à mesure car les plus jeunes élèves (ayant moins de 11 ans) évitent une baisse d'estime de soi résultants de mauvaises notes en les attribuant à des forces externes ou à de la pratique.

De nombreuses études témoignent à une forte corrélation négative entre l'anxiété et la performance, tant dans les tests réalisés par l'enseignant que dans les tests standardisés (Hill and Wigfield, 1984 ; Cassady & Johnson, 2002). En montrant que les conditions dans lesquelles les élèves passer des tests peuvent modifier la performance des élèves anxieux, Hill et Eaton (1977) suggèrent que c'est l'anxiété qui cause de mauvais résultats sur des examens. Effectivement, Cruzevialle et Butera (2013) prouvent que chez des étudiants à l'université, la présence de notes est vécue comme une menace. L'anxiété et le désir de surpasser les autres induisent des pensées parasites qui accaparent les capacités cognitives verbales des étudiants entraînant des performances plus faibles sur des tâches complexes. Dans le domaine de mathématiques, Ma et Xu (2004) fait preuve de l'existence d'une relation réciproque entre l'anxiété et la performance chez les collégiens et lycéens. Ce modèle est également soutenu par les résultats d'une étude auprès des étudiants post-secondaires dans le domaine de chimie (Gibbons et al., 2018). Ma et Xu (2004) montrent également **que la présence et l'impact de l'anxiété diffèrent selon le genre de l'élève**. Bien que la présence de l'anxiété chez les filles reste assez stable, son impact négatif sur la performance augmente lors des périodes de transition, c'est-à-dire entre l'école élémentaire et le collège, ou le collège et le lycée. Chez les garçons, l'impact de l'anxiété reste stable, tandis que sa présence s'accumule au fil du temps. Plusieurs revues de littérature (Crooks, 1988 ; McDonald, 2001 ; Harlen & Deakon, 2002) convergent sur la conclusion que **l'anxiété liée aux tests a un effet débilant sur les résultats** en raison des comparaisons entre les élèves et de l'utilisation de

⁷ Cette revue de la littérature a identifié 183 études pertinentes à leurs questions de recherche, dont 19 études empiriques correspondant à l'ensemble de leurs critères d'inclusion et une quarantaine d'autres études pouvant être utile pour comprendre et contextualiser les résultats.



notes. McDonald (2001) va plus loin et conclut que l'effet est suffisamment important pour faire la différence entre la réussite et l'échec à un test pour au moins un cinquième de tous les élèves.

Lorsque les élèves savent qu'ils sont en compétition pour des bonnes notes, **le travail coopératif entre élèves, dans le vrai sens du terme, en prend un coup**. C'est-à-dire que les élèves ont très peu d'intérêt pour partager des informations et pour se soutenir mutuellement pour réaliser les objectifs d'apprentissage. Ceci est vrai autant pour des élèves en primaire (Hayek et al., 2017) que des étudiants en université (Hayek et al., 2015). Effectivement, selon la littérature sur le partage d'informations en groupes, l'utilisation de notes dans des situations d'interdépendance amènent aux membres des groupes à faire face à des incitations concurrentielles, c'est-à-dire celles de la réussite individuelle versus la réussite collective (Davis et al., 1976 ; Wittenbaum et al., 2004 ; Hayek et al., 2015 ; Hayek et al., 2017). Et malheureusement, c'est la réussite individuelle qui finit par gagner.

En plus de ses effets négatifs sur des facteurs de réussite indirects (anxiété, motivation, estime de soi, coopération), l'évaluation sommative a également un **impact négatif direct sur les apprentissages**. Lorsque leur performance sur le test est favorisée, les élèves ont tendance à favoriser l'apprentissage des faits et des stratégies qui leur font réussir sur le test. La littérature montre que la plupart des examens réalisés par les enseignants dépendent sur les connaissances de base, qui ne nécessitent que de la mémorisation de faits et de processus (Butler, 1987; Crooks, 1988 ; Butler & Neuman, 1995). De plus, Gordon et Reese (1997) ont montré que les enseignants peuvent former les élèves à passer n'importe quel type de test, même ceux qui ont l'intention d'évaluer des capacités de réflexion supérieures, ce qui frustrer ceux qui considèrent que l'enseignement à des tests bien conçus peut influencer l'enseignement dans une manière positive (par exemple, Yeh, 2001). Kellaghan et al. (1996) ont remis en cause l'affirmation selon laquelle de meilleurs tests conduisent à un meilleur enseignement et à un meilleur apprentissage. Ce qui va de pair avec la conclusion de Elton et Laurillard (1979) disant que la manière la plus rapide de transformer les apprentissages des élèves est de transformer le système d'évaluation, ce qui ressemble presque une loi de comportement d'apprentissage.

b. Privilégier l'évaluation formative pour faire progresser tous les élèves

Si les notes chiffrées ont fortement battu en brèche au cycle 2 (CP-CE1-CE2) au profit des points de couleur (vert-jaune-rouge), des lettres (A-B-C-D), et des appréciations (TB-B-AB-à revoir), il ne s'agit pas d'un changement réel d'approche, car il s'agit toujours d'une évaluation du niveau de performance des élèves (Algan et al., 2018). En réalité ce n'est pas l'absence ou la présence de notes qui distinguent les deux formes d'évaluation, mais l'objectif recherché, et par voie de conséquence la façon dont l'enseignant dialogue avec l'élève au cours du processus d'évaluation, l'encouragement, et élabore avec lui des stratégies pour s'améliorer.

Effectivement, **l'évaluation formative a pour fonction de favoriser la progression des apprentissages**. Elle est réalisée en cours d'activité ou entre activités et unités et vise à faire état des progrès des élèves ainsi que des difficultés rencontrées. L'évaluation formative repose sur deux piliers : premièrement le diagnostic, qui identifie les progrès effectués et les points à améliorer, et deuxièmement la remédiation, qui porte sur les moyens de dépasser les difficultés. Emprunter la définition de l'évaluation formative telle qu'élaborée par Black et William (2009), nous considérons que l'évaluation est formative si les informations qu'elle produit sont utilisées par les enseignants et les élèves pour modifier leur activité.

Fuchs et Fuchs (1986) présentent une méta-analyse de 21 études différentes, portant sur des élèves de l'école maternelle jusqu'au lycée. En tout, ces études ont abouti à 96 tailles d'effet différentes. L'accent a été mis sur le travail des enfants souffrant de handicaps légers et sur l'utilisation du retour d'information par et pour les enseignants. Les études ont été rigoureusement sélectionnées : toutes comportaient une comparaison entre un groupe expérimental et un groupe témoin et des activités d'évaluation se déroulant entre 2 et 5 fois par semaine. La taille d'effet moyenne obtenue était de 0,70. Certaines des études ont également inclus des enfants non handicapés : elles ont donné une taille d'effet moyenne de 0,63 sur 22 séries de résultats (sans différence



significative par rapport à la moyenne de 0,73 pour les groupes handicapés). Les auteurs ont noté que dans environ la moitié des études, les enseignants ont été obligés de mettre en place des règles prédéterminées concernant l'examen des données et les actions à suivre, tandis que dans l'autre moitié, les actions étaient laissées au jugement des enseignants. Les premières ont produit une taille d'effet moyenne de 0,92 contre 0,42 pour les dernières. De même, les études dans lesquelles les enseignants ont entrepris de produire des graphiques des progrès de chaque enfant pour orienter et stimuler l'action ont fait état de gains moyens plus importants que celles dans lesquelles cela n'a pas été fait (taille moyenne de l'effet de 0,70 contre 0,26). (Voir l'Encadré 5 pour davantage d'informations sur une taille d'effet.

Encadré 6. Comment interpréter une taille d'effet

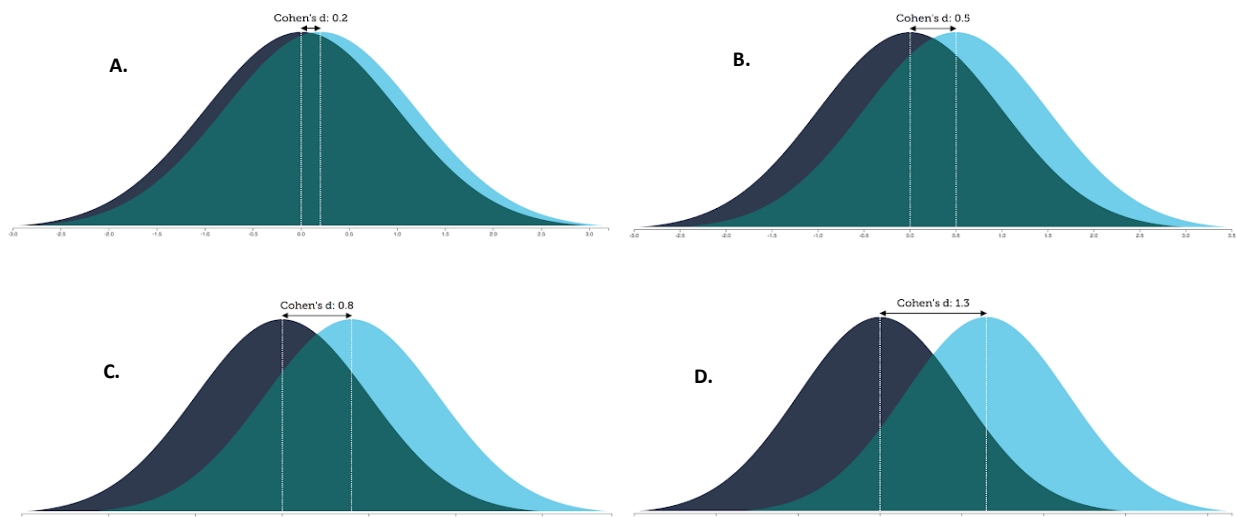
En statistique, la taille d'effet mesure la force de la relation entre deux variables sur une échelle numérique. **On l'appelle aussi « un pourcentage d'écart-type »**. Plus la taille d'effet est grande, plus la différence de performance entre les deux groupes est importante. La taille d'effet statistique nous aide à déterminer si la différence est réelle ou si elle est due à des facteurs tiers. Dans les tests d'hypothèses, la taille de l'effet, la puissance, la taille de l'échantillon et le niveau de signification sont liés les uns aux autres.

Le d de Cohen, une taille d'effet

La prise en compte de la variabilité de l'amélioration mesurée peut aider à interpréter l'ampleur et amener à une valeur qui exprime, dans la situation considérée, un rapport et non plus une différence de deux grandeurs. **Les mesures de tailles d'effet sont particulièrement utiles pour conduire des méta-analyses, car elles permettent de la comparaison des résultats issus de différentes études scientifiques pour en faire la synthèse.** Une approche de la taille d'effet commune en éducation est le d de Cohen, qui permet de caractériser la magnitude d'un effet associé dans une population donnée par rapport à une hypothèse nulle. **Le d de Cohen se calcule comme la différence entre deux moyennes, divisée par l'écart-type combiné.**

$$d = \frac{\mu_2 - \mu_1}{\sigma}$$

La question qui suit est de savoir comment interpréter le résultat. Cohen propose une grille d'interprétation assez arbitraire, telle que développée dans les graphiques ci-dessous.



Jacob Cohen avançait qu'une taille d'effet moyen de 0,5 est visible à l'œil nu (B). Une taille d'effet faible de 0,2 est sensiblement plus petite que moyenne, mais pas assez petite pour être triviale et sans importance (A). Un effet fort de 0,8 (C) est la même distance au-dessus du moyen que le petit est en dessous. Arriver à une taille d'effet de 1,3 (D) est très rare dans la littérature.

La taille d'effet en termes réels

Supposons que nous souhaitons mesurer l'impact d'une intervention pour laquelle les élèves du groupe « test » ne font plus de devoirs à la maison et ceux du groupe « témoin » continuent comme à l'habitude. Au but de



l'année scolaire, nous comparons les scores des deux groupes sur des tests standardisés en mathématiques et en lecture. Il s'avère que ne plus faire de travail à la maison conduit à une taille d'effet de 0,07 (7 % d'un écart-type) sur les scores en mathématiques et 0,11 (11 % d'un écart-type) en lecture. Mais qu'est-ce que ça veut dire en termes réels ? Pour contextualiser les tailles d'effet de manière lisible, l'*Education Endowment Fund* au Royaume-Unis traduit les tailles d'effet en « mois de progrès » supplémentaires.

Taille d'effet	Mois de progrès	Taille d'effet	Mois de progrès	Taille d'effet	Mois de progrès
0,02 – 0,09	1	0,36 – 0,44	5	0,70 – 0,78	9
0,10 – 0,18	2	0,45 – 0,52	6	0,79 – 0,87	10
0,19 – 0,26	3	0,53 – 0,61	7	0,88 – 0,95	11
0,27 – 0,35	4	0,62 – 0,69	8	0,96 – 1,00	12

Dans notre exemple, ne plus faire de travail à la maison correspond à environ 1 mois de progrès supplémentaire en maths et à environ 2 mois supplémentaires en lecture pour le groupe ayant subi l'intervention.

Limites à l'interprétation

Ces désignations grandes, moyennes et petites ne tiennent pas compte d'autres variables comme l'exactitude de la méthode d'évaluation ni de la diversité de la population étudiée. Elles fournissent un guide général qui doit être éclairé par le contexte. Les résultats des études doivent être interprétés en fonction de leur signification pratique et clinique. **Des facteurs tels que la qualité de l'étude, l'incertitude de l'estimation, l'âge des élèves, la durée de l'expérimentation et les résultats de travaux antérieurs doivent être considérés.** Ensuite, on peut interpréter ou rejeter l'importance d'une taille d'effet.

De la même manière, l'interprétation en termes de mois de progrès des élèves est une interprétation générale, qui ne prend pas en compte les spécificités propres à chaque niveau scolaire. Cette interprétation fait l'hypothèse qu'un mois d'enseignement en CP est équivalent à un mois d'enseignement en 4^e. Ces interprétations sont donc **utiles pour visualiser** mais il convient donc d'être prudent pour ne pas les prendre au pied de la lettre.

Dans une revue de la littérature de 14 domaines de recherche, Crooks (1988) met en lumière la relation entre les pratiques d'évaluation en classe et leurs conséquences sur les élèves. Il trouve notamment que l'évaluation dans la classe est très importante dans le jugement implicite qu'il porte sur ce qui est important à apprendre et dans la motivation dans la durée pour apprendre et se développer. Pour éviter des effets négatifs, il montre que l'évaluation doit, dès le début de la scolarité, faire l'objet d'une planification bien réfléchie, favoriser l'approche « apprendre à apprendre », éviter faire la comparaison entre élèves, et être effectué de manière régulière afin d'alimenter le processus d'enseignement et d'apprentissage.

Black et Wiliam (1998) ont examiné la recherche produit entre 1988 et 1998 sur l'impact de l'évaluation sur les apprentissages des élèves de tous niveaux scolaires. A la suite de leur analyse, ils considèrent que la recherche montre de manière concluante que l'évaluation formative a un impact positif et significatif sur l'apprentissage. Les gains en termes de résultats semblent être considérables et parmi les plus importants existant dans le cadre d'interventions éducatives.

Loganathan *et al.* (2020) ont choisi d'analyser les résultats de la recherche quant à l'impact de l'évaluation formative sur les élèves ayant des bons et des mauvais résultats scolaires. Ils se sont concentré sur les 14 études les plus pertinentes réalisées dans la période de 2013 à 2018. Certaines des résultats de la recherche confirment l'idée que les élèves peu performants en tirent plus de bénéfices que les élèves performants de l'évaluation formative. Mais il semble que l'évaluation formative ait un impact positif sur les deux groupes, notamment en ce qui concerne les domaines cognitif et affectif (motivation). En ce qui concerne les avantages affectifs, il semble qu'ils se manifestent par dans une implication plus importante de la part des élèves dans leur propre apprentissage. En outre, les gains cognitifs sont proportionnels aux gains affectifs – Ainsi engagés, les apprenants finissent par obtenir de meilleurs résultats, ce qui aboutit dans des gains cognitifs proportionnels aux gains affectifs.



Klute *et al.* (2017) ont fait une méta-analyse de la recherche effectuée entre 1988 et 2014 portant sur l'évaluation formative et son impact sur la performance scolaire des élèves. En tout, 23 études répondaient aux exigences du *What Works Clearinghouse* pour être retenu dans leur analyse. Globalement, ils ont identifié une taille d'effet moyenne de 0,26, variant entre 0,46 et 1,22. En comparant les effets dans différentes disciplines académiques, ils ont trouvé que l'évaluation formative utilisée lors des séances de mathématiques a été la plus efficace en comparaison à celle utilisée dans les séances de lecture ou d'écriture.

Encadré 7. Des exemples d'évaluation formative ayant des résultats positifs sur les élèves

Retours précis et détaillés (Butler & Nissan, 1986). Cette étude montre l'effet positif de l'évaluation formative au cours d'une expérience randomisée sur 261 élèves âgés de 11 à 13 ans (6th grade) en utilisant des tests de raisonnement divergent et convergent. Les élèves recevant des retours précis et détaillés affichaient des résultats plus élevés à des tests quantitatifs et sur des épreuves faisant appel à la créativité que ceux recevant des notes et ceux ne recevant aucun retour. L'évaluation formative avait aussi réduit l'anxiété, et augmenté l'intérêt des élèves pour l'exercice.

Stratégie en écriture, définition des objectifs et suivi des progrès avec l'enseignant (Schunk & Swartz, 1993). Cette expérience teste l'effet de la définition des objectifs et des retours sur les progrès réalisés sur la performance en écriture, l'auto-efficacité et le transfert des compétences acquises de 40 élèves âgés de 9 à 11 ans (4th grade). Tous les élèves ont été initiés à une stratégie d'écriture identique mais contextualisée de trois manières différentes : une avec un objectif « de processus » (apprendre la stratégie), une avec un objectif « de production » (réaliser le produit final) et une avec un objectif général (faire de bon travail, le groupe de contrôle). La moitié des élèves travaillant sous l'objectif de processus ont également reçu des retours sur leurs progrès dans la maîtrise de la stratégie d'écriture. Les résultats de cette expérience montrent que l'évaluation formative, caractérisée par la définition des objectifs orientée vers le processus et des retours sur les progrès réalisés, favorisent la performance, le transfert de stratégie et l'auto-efficacité des élèves.

Évaluation en mathématiques par carte-éclair et suivi des progrès avec les pairs (Menesses & Gresham, 2009). Cette étude a utilisé une évaluation formative dirigée par l'élève pour les élèves âgés de 7 à 11 ans (2nd à 4th grades) à risque d'échec scolaire. L'évaluation s'est concentrée sur le renforcement de la fluidité et de la précision des élèves quant à des faits de mathématique de base. Des paires d'élèves ont mis en œuvre les sessions en deux parties. La première partie de chaque session consistait en trois minutes de pratique ; la deuxième partie de chaque session était une évaluation formative. Pour faire l'évaluation formative, l'un des élèves a présenté à l'autre des cartes-éclair sur les faits qu'il venait de pratiquer. L'élève « évaluateur » a compté le nombre de faits corrects et l'a enregistré sur un tableau. Lorsqu'un élève a fourni 10 faits corrects durant la partie « évaluation » en deux sessions consécutives, cet élève a été initié à une nouvelle série de faits sur les cartes-éclair lors de la session suivante. Les sessions ont eu lieu trois fois par semaine, pour un total de 15 sessions. Les résultats de cette étude montrent que le tutorat en pairs conduit à une amélioration significativement plus importante pour le groupe « test » que pour le groupe « témoin ».

Parmi la diversité d'interventions éducatives existant dans le monde, **l'évaluation formative et ses composants reçoivent des éloges quasi universels pour leur efficacité et leur efficience**. La méta-analyse des méta-analyses réalisée par l'*Education Endowment Foundation* évalue un impact du *feedback* (retours d'informations sur le travail de l'élève et le composant le plus important de l'évaluation formative) de 0,64, c'est-à-dire un impact fort pour un coût très faible (Higgins *et al.*, 2018). Cet impact correspond à **l'équivalent de 8 mois en classe**.

Encadré 8. Remettre en perspective l'effet avec la méta-analyse de l'*Education Endowment Foundation*

L'*Education Endowment Foundation* est une association caritative anglaise liée à la recherche, et qui a pour but de s'attaquer au lien entre ressources familiales et réussite scolaire. Pour cela, ils financent régulièrement des enquêtes d'impact en éducation et publient une fois tous les deux ans une variété de méta-analyses qui visent à permettre la comparaison de différents dispositifs en éducation.

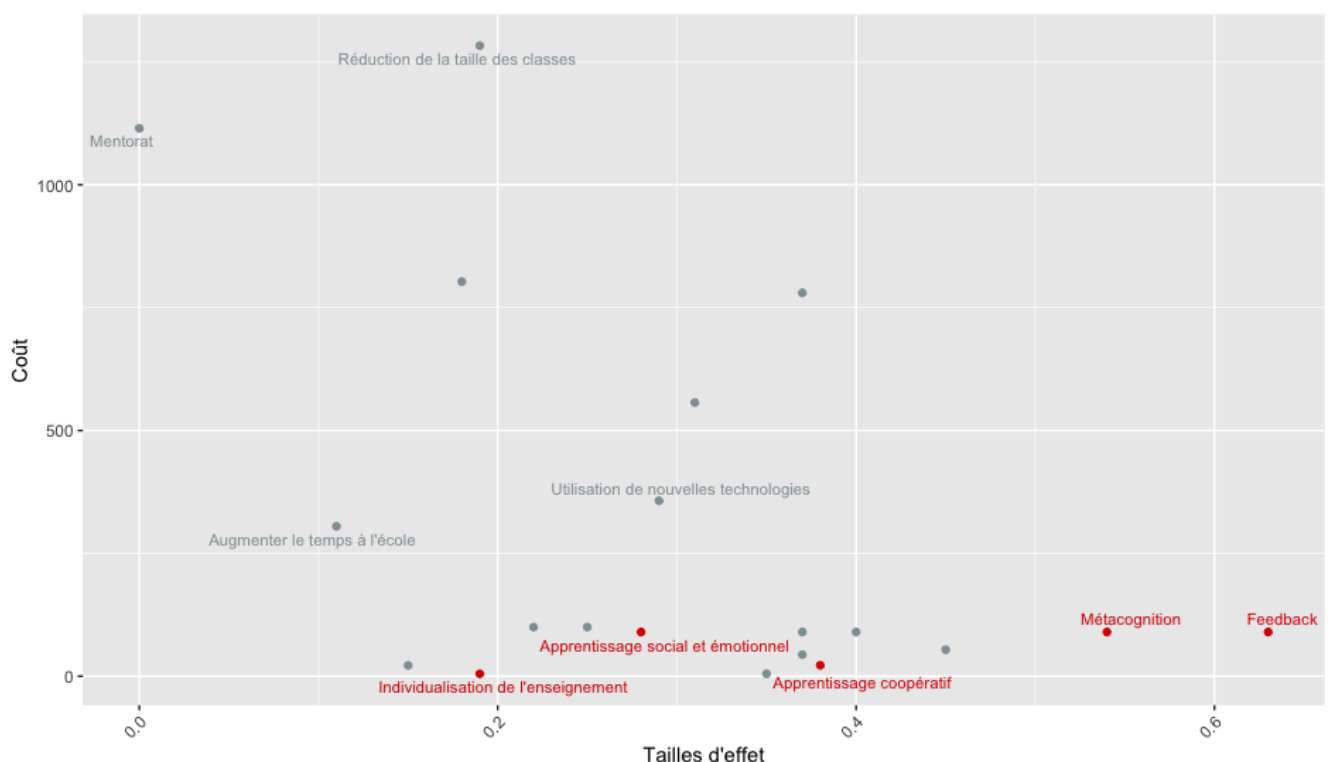


Pour cela, ils présentent trois indicateurs :

- Le coût, évalué en livre sterling mais présenté dans le graphique ci-dessous en euros.
- La taille de l'effet, selon la méthodologie présentée ci-dessus
- La robustesse du résultat. Nous avons fait le choix de présenter uniquement les programmes dont la méthodologie était supérieure ou égale à la moyenne.

On peut voir sur la Figure 1 que le **feedback (retours d'informations sur le travail de l'élève et aspect central de l'évaluation formative)** représente un coût très faible pour une efficacité très bonne ($d=0,64$). Nous avons fait le choix de présenter en addition l'effet de l'individualisation de l'enseignement ($d=0,19$) qui peut être un effet qui complète l'évaluation formative notamment, et qui a un effet plus faible mais aussi à un coût mineur.

Pour une comparaison, on voit que la réduction de la taille des classes a un effet équivalent à celui de l'individualisation ($d=0,19$) pour un coût multiplié par 250. De la même manière, l'utilisation de nouvelles technologies a un effet moyen ($d=0,29$) mais nettement inférieur au feedback, pour un coût multiplié par 15.



Note : effets calculés par l'Education Endowment Foundation (Higgins et al., 2016).
On a enlevé deux points : le redoublement d'une année ($d=-0,32$, 6690€) et interventions pré-scolaires ($d=0,38$, 6470€)

Figure 12 - Graphique des tailles d'effet des différentes interventions revues par l'EEF en fonction de leur coût (€)

c. La théorie à la base de l'évaluation formative

C'était Bloom (1968) qui a initialement fait la distinction entre l'évaluation formative et l'évaluation sommative dans le cadre de l'apprentissage des élèves en classe. Il a également conclu son papier en éclaircissant la complémentarité entre ces deux formes d'évaluation dans le développement de l'auto-efficacité chez les élèves. En examinant la contribution de Bloom, Cardinet (1977) a proposé que la fonction de l'évaluation formative soit la régulation de l'apprentissage des élèves par de boucles d'adaptation intégrées dans l'enseignement à chaque étape des activités pédagogiques en classe.⁸

⁸ Au fur et à mesure, Cardinet (1986) est arrivé à l'idée que la démarche évaluative nécessite des interactions entre l'enseignant et les élèves, si elle va contribuer à la progression des apprentissages.



Dans le prolongement tu teste de Cardinet, Allal (1979) a présenté une comparaison systématique entre une conception de l'évaluation formative basée sur le modèle néo-comportementaliste de l'apprentissage par la maîtrise de Bloom⁹ et une approche de l'évaluation formative basée sur une conception cognitive et constructiviste de l'apprentissage, comme le suggèrent la théorie de l'apprentissage de Bruner (1966)¹⁰ et le concept d'équilibration de Piaget (1975)¹¹. Au lieu de se concentrer uniquement sur la remédiation, elle propose une approche cognitive et constructiviste impliquant d'autres façons de mener l'évaluation formative. En effet, elle a mis l'accent sur l'observation et l'interaction avec les élèves lorsqu'ils effectuent des tâches afin de comprendre leurs représentations cognitives et leurs processus de pensée. Par ailleurs, elle a proposé que la régulation de l'apprentissage en classe vise à créer un « décalage optimal » entre les exigences de la tâche d'apprentissage et les connaissances et compétences que l'apprenant possède déjà (Allal, 1979). Si l'écart est trop important, l'apprenant sera incapable de participer à la tâche. Si, en revanche, l'écart est trop faible, comme dans les séquences d'apprentissage comportemental où l'apprenant réussit immédiatement l'étape suivante, alors le processus d'équilibre et de restructuration cognitive –nécessaire pour un apprentissage profond – ne se produira probablement pas. La notion de décalage optimal implique une tension synergique entre une tâche qui provoque la réflexion de l'apprenant et un apprenant qui manipule activement la tâche et ses paramètres.

Allal (1988) a décrit trois modalités de régulation associées à l'évaluation formative qui diffèrent en fonction de la temporalité de leurs contingences. Elles sont :

- La **régulation rétroactive** (c'est-à-dire la « remédiation » de Bloom) se dessine autour les résultats d'une évaluation formative qui identifient les objectifs d'apprentissage que l'élève n'a pas maîtrisés lors d'une phase antérieure de l'enseignement, conduisant à des « correctifs » qui permettent à l'élève de revisiter ses objectifs pour surmonter ses difficultés ;
- La **régulation proactive** se produit lorsque les informations issues de l'évaluation formative sont utilisées pour planifier et préparer de nouvelles activités pédagogiques, y compris l'enrichissement et la consolidation de l'apprentissage, en fonction des besoins et des intérêts de tous les élèves ;
- La **régulation interactive** se base sur les interactions de l'élève avec l'enseignant, ses pairs, et le matériel pédagogique et les outils d'évaluation, *pendant* les activités pédagogiques en cours.

Bien que ce soit l'enseignant qui met en place les structures permettant à une approche systématique pour réaliser ces différentes modalités de régulation dans le cadre d'une évaluation formative, l'apprenant a également un rôle à jouer. Avec des informations sur les buts à atteindre et des retours rétroactifs sur ses avancées auparavant, l'apprenant peut également jouer un rôle actif et important dans son apprentissage. Ce processus dynamique par lequel l'apprenant planifie, surveille et évalue ses propres apprentissages s'appelle **l'autorégulation** et apparaît être un déterminant important de l'amélioration des résultats scolaires et de la motivation (Cosnefroy, 2011 ; Efklikes, 2006 ; Laveault, 2007).

Carver et Sheier (1999) propose un modèle d'autorégulation focalisant sur le rôle des rétroactions sur l'éventuelle efficacité de l'autorégulation. Dans ce modèle, l'autorégulation se concerne avec l'écart entre le but à atteindre et le résultat des dernières actions mises en place pour l'atteindre. Les rétroactions sont effectivement des retours de l'environnement (de l'enseignant, des pairs, d'un outil numérique, d'un rubrique d'auto-évaluation) qui permet à l'apprenant de se repérer où il est et où il doit aller pour arriver à son objectif.

⁹ Bloom a proposé que l'enseignant peut utiliser de « correctifs » pour éliminer les erreurs et conduire à la maîtrise des objectifs d'apprentissage.

¹⁰ Selon la théorie de Bruner, les individus forment des hypothèses, mettent en œuvre des stratégies pour les tester, révisant suite à la fois les hypothèses et les stratégies pour attendre les concepts.

¹¹ Selon Piaget, ce concept représente un processus qui conduit de certains états d'équilibre approché à d'autres, qualitativement différents, en passant par de multiples déséquilibres et rééquilibrations.



	<i>Où les élèves vont-ils ?</i>	<i>Où en sont-ils dans leurs apprentissages ?</i>	<i>Comment, et par quelles étapes, peuvent-ils atteindre les buts visés ?</i>
Enseignant	Clarifier, partager, et faire comprendre les intentions d'apprentissage et les critères de réussite	Organiser de véritables discussion, activités et tâches qui produisent des preuves sur les apprentissages	Donner un feedback qui fait progresser les élèves
Pairs		Inciter les élèves à être des ressources pour leurs pairs	
Apprenants		Inciter les élèves à être responsables de leurs apprentissages	

Tableau 1 - Aspects de l'évaluation formative

Source : Lepareur, Grandit et Grangeat (2017)

Dans une étude de cas portant sur l'autorégulation et les rétroactions, Lepareur, Grandit et Grangeat (2017) présente un modèle qui conceptualise l'évaluation formative en différentes stratégies génériques qui résultent des trois processus centraux liés à des rétroactions et aux rôles de l'enseignant, des apprenants et des pairs (Tableau 1).



II. Pratiques quotidiennes et activités pour la classe

A. Favoriser l'évaluation formative en classe

Selon Cauley et McMillan (2010), les enseignants efficaces utilisent l'évaluation formative pendant l'enseignement pour identifier les malentendus spécifiques des élèves, fournir un retour d'information aux élèves sur leur travail pour les aider à corriger leurs erreurs, et identifier et mettre en œuvre des correctifs pédagogiques. L'évaluation formative continue se fait principalement par le biais d'observations informelles et de questions orales posées aux élèves pendant l'enseignement ou la révision du contenu. Si les informations tirées des observations et des questions posées aux élèves sont exactes, l'enseignant peut identifier les ajustements pédagogiques qui sont susceptibles d'aider à améliorer les apprentissages. De cette façon, l'évaluation formative est intégrée à l'enseignement et, idéalement, elle fournit un processus continu d'évaluation suivi d'un enseignement, puis d'une évaluation et d'un enseignement supplémentaire, etc. En établissant ce cercle vertueux d'évaluation et d'enseignement, l'enseignant a la possibilité d'augmenter la motivation, l'engagement, l'auto-efficacité et, *in fine*, les résultats scolaires de ses élèves (voir Figure 13).

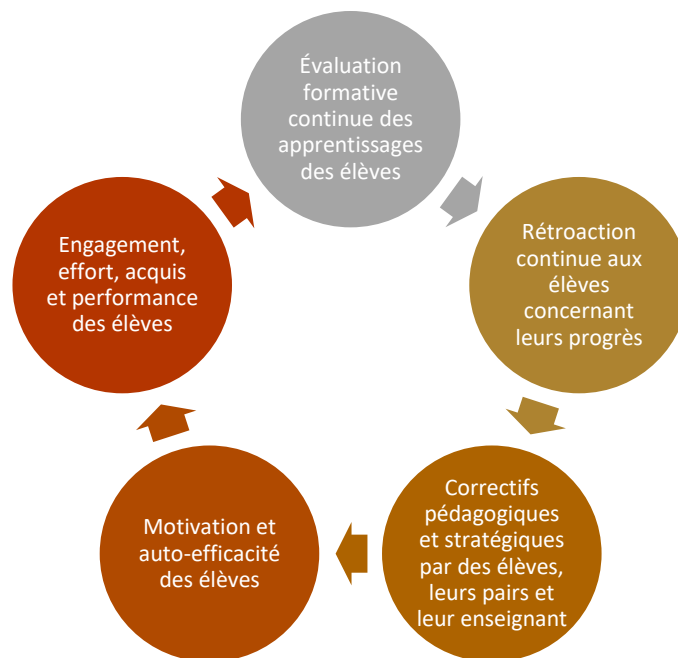


Figure 13 - Cycle d'évaluation formative

Pour favoriser ce cycle d'évaluation, Cauley et McMillan (2010) proposent aux enseignants de fournir des objectifs d'apprentissages très clairs, chercher des preuves de compréhension et de connaissance, adapter la manière d'enseigner et les stratégies d'apprentissage proposées aux élèves, offrir des retours précis concernant la réalisation des objectifs d'apprentissage, et impliquer les élèves dans le processus d'évaluation et d'apprentissage.

1. Fixer des objectifs d'apprentissage très clairs

Pour l'enseignant un objectif d'apprentissage représente ce qu'il souhaite que ses élèves apprennent pour un sujet donné. Pour l'élève, il représente ce qu'il devra être capable de faire. Il n'est que possible d'identifier le chemin à prendre, si l'objectif, « la destination » est connu.

Il existe de nombreuses études montrant que fixer des objectifs concrets axés sur l'apprentissage et non la performance favorise la motivation et l'apprentissage (Bandura, 1986 ; Locke & Latham, 1990 ; Schunk, 1996 ; Schunk, 2003). Les effets des objectifs d'apprentissage ne sont cependant pas automatiques, mais dépendent plutôt des propriétés des objectifs : leurs spécificité, proximité et difficulté. Les objectifs qui dénotent des normes de



performance spécifiques sont plus susceptibles d'améliorer la performance que les objectifs généraux (par exemple, « Faites de votre mieux »). Les objectifs proches dans le temps augmentent les performances mieux que les objectifs éloignés. Les objectifs ambitieux mais réalisables augmentent la motivation et l'apprentissage mieux que les objectifs perçus comme très faciles ou trop difficiles (Schunk, 1991). Les objectifs de processus impliquent des techniques et des stratégies que les étudiants utilisent pour apprendre (Weinstein & Mayer, 1986).

Les effets des objectifs sur la performance dépendent également du sentiment d'auto-efficacité de l'apprenant, c'est-à-dire ses croyances concernant sa capacité à effectuer des actions à des niveaux désignés (Bandura, 1986). L'auto-efficacité peut affecter le choix des activités, l'effort et la persévérance. Lorsque les élèves adoptent un objectif, ils peuvent éprouver un sentiment d'auto-efficacité pour l'atteindre. Leur sentiment initial d'efficacité est confirmé lorsqu'ils travaillent sur la tâche et observent la progression de l'objectif, car les progrès perçus indiquent qu'ils deviennent habiles. À leur tour, les évaluations des capacités renforcées soutiennent la motivation et amènent les apprenants à se fixer de nouveaux objectifs lorsqu'ils atteignent leurs objectifs actuels (Schunk, 1991).

En France, le programme scolaire propose déjà des objectifs d'apprentissages très clairs. C'est donc le rôle de l'enseignant de communiquer ces objectifs à ses élèves de manière adaptée à leur niveau de compréhension actuelle.

Encadré 9. Fixer des objectifs d'apprentissage

Avant de construire les objectifs d'apprentissage pour un sujet donné, l'enseignant doit bien comprendre son public (la composition de sa classe), le point de départ de ses élèves pour un sujet de donné (les acquis préalables), les choses qu'il souhaite que ses élèves soient capables à faire, et les indicateurs que cet objectif a été atteint.

Les objectifs spécifiques

Pour commencer, il peut être utile de construire une phrase du type « A la fin de l'activité, l'élève sera capable de... ». Pour compléter cette phrase, l'enseignant peut garder en tête les règles générales suivantes :

- Un objectif reste un énoncé court, exprimé avec un verbe d'action.
- Une description d'un contenu (une liste de concepts, par exemple) ne constitue pas un objectif.
- Un objectif visant l'utilisation d'un concept est, en revanche, acceptable.

Les verbes à utiliser

Les verbes employés pour rédiger les objectifs spécifiques sont plus précis que ceux utilisés pour les objectifs généraux. Le verbe qui décrit l'objectif spécifique doit décrire de façon précise l'effet attendu, ils traduisent des gestes, des actions, des performances observables. On conseille d'éviter des verbes vagues pour lesquels il sera difficile d'imaginer comment vérifier si l'objectif est atteint : connaître, comprendre, apprendre, savoir, percevoir... sont trop vagues, la difficulté se reportera sur l'évaluation. On peut utiliser les verbes de la taxonomie de Bloom qui associe à six niveaux de compétences des verbes d'actions. (Voir le tableau ci-dessous)

	<i>Niveau</i>	<i>Verbes indicatifs des capacités</i>
1	Connaissance	définir, identifier, nommer, énumérer, dire avec ses propres mots ...
2	Compréhension	décrire, résumer, expliquer, interpréter ...
3	Application	utiliser, résoudre, construire, démontrer, calculer, dériver...
4	Analyse	analyser, distinguer, comparer, faire le choix ...
5	Synthèse	concevoir, rédiger, planifier, réaliser, faire un exposé, produire, mettre au point, ...
6	Évaluation	justifier, défendre, juger, argumenter, critiquer, évaluer ...

Source : INSA



2. Chercher des preuves de compréhension et de connaissances des élèves

« Si vous pouvez à la fois écouter les enfants et accepter leurs réponses non comme des choses à juger comme vraies ou fausses mais comme des éléments d'information qui peuvent être révélateur des pensées de l'enfant, vous aurez fait un pas de géant pour devenir un maître enseignant plutôt qu'un simple diffuseur d'informations. » (Easley & Swoyer, 1975, p. 25)

Les enseignants utilisent un mélange d'approches pour évaluer l'apprentissage et la compréhension des élèves. Selon Klute et al. (2017), les enseignants ont parfois recours à l'évaluation diagnostique pour déterminer le niveau d'un élève lorsqu'il entre dans une nouvelle école ou à des moments précis de l'année scolaire afin de les aider à élaborer leurs stratégies d'enseignement et à adapter les programmes. Le plus souvent, cependant, les enseignants utilisent une évaluation formative à cycle court au cours des leçons et entre celles-ci pour ajuster l'enseignement en temps réel. C'est ce qu'Allal (1979) appelle la "régulation interactive" de l'apprentissage.

Dans une étude menée auprès de sept enseignants du premier degré expérimentés, Reynolds et al. (1995) mettent en évidence les indicateurs implicites de la compréhension chez les jeunes élèves qui ont émergé de leurs discussions avec ces enseignants. Il est important de noter que les enseignants considèrent que ces indicateurs représentent des éventuels indices leur permettant de diagnostiquer le niveau de compréhension de leurs élèves quant à un sujet donné. Les indicateurs qu'ils ont tous identifiés sont :

- 1) **changements de comportement** : les élèves qui avaient compris avait des yeux « brillants » tandis que ceux qui n'avaient pas compris avait des yeux « glauques » ;
- 2) **extension d'un concept** : les élèves qui ont compris quelque chose ont tendance, de leur propre initiative, d'aller plus loin avec l'idée ;
- 3) **modification d'un schéma** : les élèves qui comprennent commencent spontanément à apporter leurs propres modifications au concept, tandis que ceux qui ne comprennent pas ont tendance à imiter ou suivre soigneusement les règles ;
- 4) **utilisation des processus dans un contexte différent** : les élèves qui ont compris une idée particulière commencent souvent à voir les mêmes schémas ailleurs ;
- 5) **utilisation des raccourcis** : seuls les étudiants qui sont sûrs d'avoir une vue d'ensemble peuvent raccourcir une procédure de sorte que le fait de penser ou d'utiliser un raccourci est considéré comme une preuve de compréhension ;
- 6) **capacité d'expliquer** : les élèves qui ont compris quelque chose sont généralement capables de l'expliquer ;
- 7) **capacité de concentrer son attention** : la persistance dans une tâche est considérée comme un signe de compréhension.

Selon Black et al (2004), les enseignants ont du mal à mener un dialogue en groupe classe qui sert à aider les élèves à apprendre. Par exemple, des recherches ont montré qu'après avoir posé une question, très peu d'enseignants attend plus d'une seconde pour la réponse de leurs élèves. Si aucune réponse est donnée, ils ont tendance à procéder à une autre question ou à répondre eux-mêmes à la question posée. Une conséquence d'un « temps d'attente » aussi court est que les seules questions qui « marchent » sont celles auxquelles on peut répondre rapidement, sans réflexion - c'est-à-dire les questions concernant les faits mémorisés. Par conséquent, le dialogue se situe à un niveau superficiel.

Un questionnement efficace fait également partie des interventions impromptues menées par les enseignants une fois que ses élèves sont engagés à une activité. Poser des questions simples, telle que « Qu'en penses-tu? » ou « Peux-tu expliquer ton raisonnement ? », peuvent être une partie de la dynamique interactive en classe et peuvent fournir une occasion inestimable d'appuyer les réflexions des élèves aux travers des retours précis et détaillés concernant leur travail.



Encadré 10. Quelques techniques pour vérifier la compréhension des élèves

La technique de « feux de circulation »

Lorsque les enseignants veulent vérifier la compréhension des élèves concernant un sujet donné, ils leur demandent de tenir un panneau vert, jaune ou rouge pour indiquer s'ils comprennent, s'ils pensent comprendre mais ne sont pas tout à fait sûrs ou s'ils ne comprennent pas. Ils passent plus de temps avec les élèves qui montrent le jaune et le rouge. Selon le concept à apprendre, ils peuvent travailler directement avec les élèves ayant une compréhension en « rouge » et organiser des pairs d'élèves « vert/jaune » pour travailler indépendamment en parallèle. En tout cas, ces panneaux permettent à l'enseignant d'adapter immédiatement son enseignement à sa classe. (OECD, 2005)

Un temps de réflexion au lieu de lever les mains

Il est possible d'appliquer une politique de « non intervention » qui néanmoins favorise de manière importante la progression des apprentissages. Dans cette situation, l'enseignant pose une question, fait une pause de trois secondes à plusieurs minutes, puis appelle un élève. Il est compris que l'enseignant peut appeler n'importe quel élève. Ainsi tous les élèves doivent être prêts à répondre à la question posée. Il s'avère que la qualité des réponses s'améliore beaucoup lorsque les élèves ont le temps de réfléchir. De plus, les élèves les plus timides ou ceux qui se considèrent eux-mêmes comme les plus faibles sont amenés à participer à la discussion en classe. Ce n'est plus le cas que les élèves les plus performants sont les seuls à s'engager aux différentes discussions. (Black et al., 2004)

Suivi des élèves par l'enseignant

L'un des aspects clés d'une évaluation formative continue c'est le suivi des élèves tout au long de l'année scolaire, ce qui représente un vrai défi pour l'enseignant qui a une vingtaine ou trentaine d'élèves dans sa classe. Pour relever ce défi, l'enseignant peut créer un carnet de suivi comportant d'une section pour chacun de ses élèves où il ou elle peut prendre des notes ou cocher une liste de contrôle, par exemple. Il est également possible de faire appel à l'auto-évaluation de l'élève pour compléter ces informations. Par exemple, l'enseignant peut demander à l'élève s'il estime qu'il a atteint la maîtrise d'une compétence donnée pour compléter sa propre évaluation formative de l'élève. (IES/REL, 2020)

Pour changer cette situation et favoriser la vérification de la compréhension, Black et al (2004) proposent de :

- **Formuler des questions qui valent la peine d'être posées**, c'est-à-dire des questions qui explorent des sujets essentiels au développement de la compréhension des élèves.
- **Prolonger le temps d'attente** afin de donner aux élèves le temps de réfléchir, et s'attendre à ce que chacun ait une réponse et contribue à la discussion. Ensuite, toutes les réponses, bonnes ou mauvaises, peuvent être utilisées pour développer la compréhension. L'objectif est d'améliorer la réflexion plutôt que d'obtenir une réponse correcte du premier coup.
- **Planifier des activités de suivi** qui favorisent l'approfondissement de l'apprentissage et de la compréhension des élèves.

De manière générale, les indicateurs de Reynolds et al. (2017) et les changements proposés par Black et al. (2004) permettent aux enseignants de diagnostiquer le niveau de compréhension de leurs élèves, d'identifier des éventuels trous ou malentendus, et d'adapter leur enseignement aux besoins de leurs élèves.

3. Offrir des retours précis et détaillés concernant le travail de l'élève

Dans une étude sur l'écart persistant entre les élèves issus de milieux sociaux différents, Autin et al. (2019) montrent qu'une évaluation sommative ayant comme objectif de classer les élèves peut générer la création de l'écart entre les élèves de différents milieux sociaux en raison du biais inconscient de la personne évaluant leurs résultats. Le fait d'utiliser l'évaluation pour la sélection cause la personne faisant l'évaluation du travail de trouver plus d'erreurs lorsque le test est attribué à un étudiant de milieux défavorisé plutôt qu'à un étudiant de milieu favorisé. **Cette étude montre que ce n'est pas nécessairement l'outil qui est mauvais, mais plutôt son usage.**



Encadré 11. Le rôle des tests réalisés par l'enseignant dans l'évaluation formative

La recherche montre que les examens peuvent être bénéfiques quant à l'acquis des nouvelles connaissances ou compétences et à leur rétention (Karpicke & Roediger, 2008 ; Roediger & Karpicke, 2006). En effet, tester régulièrement les connaissances pourrait être plus efficace que de les ré-enseigner, car cela implique un processus de récupération active en mémoire (retrieval) avec des effets positifs sur la mémorisation. En matière de « pratiques pédagogiques », le test est très facile à mettre en place, pas très coûteux en terme de temps et de matériel et avec une efficacité avérée.

Tests au service de l'évaluation formative, l'apprentissage par la maîtrise

L'enseignant donne des leçons sur les compétences de décodage à l'ensemble des élèves de sa classe. Dans chaque leçon, il procède à des évaluations formatives sur les objectifs de la leçon. Les élèves qui répondent correctement à 80 % ou plus des éléments de l'évaluation participent à des activités d'enrichissement, tandis que les autres élèves participent à des activités de ré-enseignement. Une deuxième évaluation formative parallèle est ensuite administrée pour recadrer les prochaines étapes. Cette façon d'utiliser l'évaluation sommative au service de la progression a conduit à des progressions plus importantes dans le groupe test versus le groupe témoin (Null, 1990).

Les retours d'information évaluatif sous forme de notes, de lettres, de coches et de commentaires non spécifiques (par exemple, « très bien » ou « bon travail ») n'aident pas les élèves à identifier les points forts et les points faibles sur lesquels ils peuvent capitaliser ou s'améliorer et risquent même de miner la confiance, la motivation et la performance des élèves lors des possibilités d'apprentissages ultérieures (Schunk & Swartz, 1993 ; Black & Wiliam, 1998). Pour faire progresser les élèves dans leurs apprentissages, les retours de l'enseignant doit se porter sur les tâches d'apprentissage. Le retour d'information descriptif dans le cadre de l'évaluation est largement reconnu comme un mécanisme de soutien essentiel à l'apprentissage des élèves (Hattie & Timperley, 2007). Il permet aux apprenants de comprendre comment ils se débrouillent ou se sont débrouillés, ainsi que ce qu'ils pourraient faire à l'avenir pour améliorer leurs connaissances et leurs performances (Callingham, 2008).

En effet, Hattie (2008) a réalisé une grande méta-analyse de la littérature scientifique dédiée aux facteurs de la réussite scolaire. Il en déduit que la présence de retours pertinents de l'enseignant sur les travaux de l'élève constitue le facteur le plus puissant de réussite scolaire, avec une taille d'effet moyenne de 0,73. En comparaison, des interventions plus répandues, mais aussi plus coûteuses, telles que la réduction de la taille des classes ou les devoirs, mènent à des tailles d'effet de 0,21 et 0,29 respectivement.

4. Adapter l'enseignement et les stratégies d'apprentissage

Les recherches menées dans différents domaines montrent que les étudiants auxquels on enseigne des stratégies d'apprentissage ont tendance à améliorer leurs compétences (Borkowski, 1985 ; Pressley et al., 1990), et que l'utilisation de stratégies efficaces est positivement associée à l'auto-efficacité (Pintrich & De Groot, 1990 ; Zimmerman & Martinez-Pons, 1990). Selon Crooks (1988), de nombreuses stratégies peuvent être utilisées en fonction des besoins individuels des élèves. Différentes attentes ou tâches peuvent être fixées pour chaque élève, une grande flexibilité dans les parcours d'apprentissage peut être prévue (par exemple avec des approches d'apprentissage par la maîtrise), ou l'apprentissage coopératif peut être utilisé pour réduire la pression sur les individus et compenser les forces et faiblesses individuelles. Les élèves les plus faibles peuvent bénéficier de l'identification d'objectifs intermédiaires plus faciles à atteindre, ce qui permet d'obtenir un modèle de succès répétés qui conduit à une meilleure auto-efficacité.

Encadré 12. Des méthodes pour adapter l'enseignement à l'apprenant

Diversifier les méthodes d'enseignement

Fournir une mallette à stratégies pouvant arriver au même but mais de différentes manières

Voir aussi Chan et al 2014_ The Critical Role of Feedback in Formative Assessment for some techniques



5. Impliquer des élèves dans le processus d'évaluation et d'apprentissage

Les enfants ont tendance à juger leur propre travail en fonction de sa netteté, de son exactitude et de son exhaustivité, selon les critères qu'ils perçoivent comme correspondant à ceux de leurs enseignants. Perry (1998) montre que ces critères peuvent être modifiés par une action délibérée de l'enseignant. Benmansour (1999) souligne l'importance de cultiver la motivation intrinsèque et l'auto-efficacité des élèves. Ce sont les facteurs qui conduiront à l'utilisation de meilleures stratégies d'apprentissage et à l'amélioration des performances.

On suppose que l'auto-efficacité influence le choix des tâches, l'effort, la persistance et la réussite (Bandura, 1986, 1997 ; Schunk, 1995). Par rapport aux élèves qui doutent de leurs capacités d'apprentissage, ceux qui se sentent efficaces pour apprendre ou exécuter une tâche participent plus facilement, travaillent plus dur, persistent plus longtemps lorsqu'ils rencontrent des difficultés et obtiennent de meilleurs résultats.

Fixer des objectifs d'apprentissage personnalisés (c'est-à-dire pour et par chaque élève) fait partie intégrante de la motivation et de l'apprentissage. Au début d'une activité, les élèves ont des objectifs tels que l'acquisition de compétences et de connaissances, l'achèvement d'un travail et l'obtention de bonnes notes. Au cours de l'activité, les élèves observent, jugent et réagissent à leur perception des progrès réalisés par rapport à ces objectifs (Bandura, 1988 ; Locke & Latham, 1990 ; Schunk, 1990). Lorsqu'ils s'engagent à tenter d'atteindre un objectif, ils sont susceptibles de comparer leurs performances avec les objectifs au fur et à mesure qu'ils travaillent sur la tâche. Les auto-évaluations des progrès augmentent l'auto-efficacité et maintiennent la motivation. Un écart perçu entre les performances et l'objectif peut créer une insatisfaction temporaire qui conduit à une augmentation d'effort.

Selon Schunk (2003), les élèves ne peuvent atteindre un objectif d'apprentissage que s'ils comprennent cet objectif pour eux-mêmes et peuvent évaluer ce qu'ils doivent faire pour l'atteindre. Ainsi, l'auto-évaluation est essentielle à l'apprentissage. De nombreux enseignants qui ont essayé de développer les compétences de leurs élèves en matière d'auto-évaluation ont constaté que la première et la plus difficile des tâches est d'amener les élèves à penser à leur travail *en fonction* d'un ensemble d'objectifs. Une fois que cela a été compris par les élèves, il devient possible pour les élèves de développer une vue d'ensemble de leur travail, leur permettant de le contrôler par eux-mêmes. Autrement dit, les élèves développent la capacité de travailler à un niveau métacognitif. (Voir Tableau 2.)

Auto-efficacité liée aux objectifs d'apprentissage	→	Engagement dans la tâche Auto-évaluation	→	Auto-efficacité Motivation Performance
--	---	---	---	--

Tableau 2 - L'auto-efficacité et l'apprentissage

Source : Schunk, 2003

Schunk (1996) a réalisé deux études pour examiner la façon dont les objectifs et l'auto-évaluation affectent la motivation et les résultats scolaires. Dans les deux études, des élèves âgés de 9 à 11 ans (4th grade) ont reçu un enseignement et un entraînement sur les fractions au cours de plusieurs séances de mathématiques. Les élèves ont travaillé dans des conditions impliquant soit un objectif d'apprentissage de la résolution de problèmes (objectif d'apprentissage), soit un objectif de simple résolution (objectif de performance). Dans la première étude, la moitié des élèves de chaque condition d'objectif ont évalué leurs capacités à résoudre des problèmes. L'objectif d'apprentissage *avec ou sans* auto-évaluation et l'objectif de performance *avec* auto-évaluation ont conduit à une meilleure auto-efficacité, compétence, motivation et orientation vers la tâche que l'objectif de performance sans auto-évaluation. Dans la deuxième étude, tous les élèves de chaque condition d'objectif ont évalué leurs progrès dans l'acquisition de compétences. L'objectif d'apprentissage a conduit à une plus grande motivation et à des résultats plus élevés que l'objectif de performance. Ces constats témoignent à la fois l'importance des objectifs d'apprentissage, tels que proposés par l'enseignant et adaptés par les élèves, ainsi que la valeur de l'auto-évaluation dans le processus d'évaluation formative.



Encadré 13. Les techniques d'auto-évaluation

La technique de « feux de circulation » revisitée

La technique de « feux de circulation » proposée ci-avant représente également une façon de mettre en place l'auto-évaluation par les élèves. En effet, en demandant aux élèves de se situer leur compréhension d'un sujet donné sur le feu de circulation (de rouge à vert), ils font une auto-évaluation de ce qu'ils ont appris ou non. Pour aller plus loin avec cette technique, on pourrait même demander aux élèves de justifier leur place sur le feu de circulation à l'un de leur camarade de classe, ce qui conduit au développement des compétences d'auto-évaluation et d'un certain détachement qui lui permettent de mettre l'évaluation au service de son apprentissage. (Black et al., 2004)

Portfolios, journaux de bord et grilles d'évaluation

Les portfolios et les carnets de bord utilisés dans les écoles faisant l'objet d'études de cas permettent des dialogues écrits entre l'enseignant et l'élève, et donnent également aux élèves l'occasion de réfléchir à leur processus d'apprentissage. Les grilles d'évaluation sont des lignes directrices spécifiques avec des critères permettant d'évaluer la qualité du travail des étudiants, généralement sur une échelle de points. Les élèves peuvent utiliser des grilles d'évaluation pour évaluer leur propre travail, le corriger et l'améliorer. (OECD, 2005)

Voir aussi Schunk 1996 (Goal and Self-Evaluative Influences During Children's Cognitive Skill Learning)

B. Lutter contre la menace du stéréotype de manière directe

1. L'affirmation de valeurs

Utilisée par Sherman et al sur des 6^e-4^e (Sherman et al., 2013) ou Cohen et al sur des 5^e (Cohen et al., 2006). Affirmer ses valeurs permet de protéger leur motivation. Le test n'est pas moins stressant que pour les autres mais ils ne se sentent pas liés à la menace, qui n'a pas d'impact sur leur sentiment d'auto-efficacité. Cette technique est utilisée dans ces deux études pour minimiser l'écart dû à l'origine ethnique (Afro-américains et latino-américains. Cohen et al. voient une augmentation des notes pour 70% des Afro-américains du groupe traitement, ainsi qu'un taux de mauvaises performances diminuer (20% dans le groupe contrôle, 9% dans le groupe traitement).

- ⇒ *Pratiques possibles* : cours d'expressions de ses sentiments à la manière de la Belgique par exemple. Ou exercice d'affirmation des valeurs (choisir une valeur qui est importante pour si et écrire en quelques lignes pourquoi elle l'est).

2. La présentation de rôles modèles

La création d'environnement peut protéger l'identité. Marx et al. ont montré avec des étudiants que lorsque qu'un stéréotype est activé, l'identité de groupe est mise en avant (Marx et al., 2005). Mais quand des modèles positifs sont présentés au sein du groupe, les membres réussissent mieux. Plusieurs études françaises se sont penchées sur la possibilité de présenter des rôles modèles à l'opposé des stéréotypes. Cela a été reproduit selon différents contextes et méthodologies. McIntyre et al. font lire à certains étudiants des textes à propos de femmes qui ont réussi (McIntyre et al., 2003). Cela augmente significativement les résultats des étudiants et n'impacte pas les étudiants. Bagès et Martinot présentent à des CM2 des modèles se présentant comme ayant réussi grâce à leurs efforts ou grâce à leurs dons de manière innée garçons ou filles issus de la classe de 6^e (Lockwood et Kunda montrent que les élèves sont plus motivés par des élèves plus vieux que de la même année (Lockwood & Kunda, 1997)). Elles montrent qu'avec un modèle travailleur, les garçons et les filles réussissent autant (Bagès & Martinot, 2011). Quand il n'y a pas d'explication ou une justification par le talent, les garçons et les filles réussissent mieux quand on leur présente un modèle féminin. Enfin, Galdi et al. mettent en place une intervention auprès des CP : ils montrent que même sans connaissance explicite des stéréotypes, il y a une sous-performance des filles en



maths qui pourrait être due à des associations automatiques (Galdi et al., 2014). Leur intervention consiste en des coloriages de filles qui font des maths. Cela diminue l'effet des associations automatiques et les filles ont dans cette condition les meilleurs résultats. Ils ne trouvent pas d'effet sur les garçons, suggérant une absence d'associations automatiques. Stephens et al. organisent une intervention pour les étudiants rentrant à l'université en leur présentant des personnes de milieux et passés différents pour présenter les difficultés que l'on rencontre, et que l'on surmonte (Stephens et al., 2014). En utilisant les histoires personnelles des élèves d'université, ils trouvent une diminution du fossé de performances entre étudiants de classe sociale favorisée (dont au moins un parent a déjà suivi un cursus à l'université) et les étudiants de classe sociale défavorisée (dont aucun parent n'est allé à l'université) de 63% et une amélioration de plusieurs variables psychologiques (engagement social et académique notamment).

- ⇒ *Pratiques possibles* : attention à utiliser des exemples variés, vérifier le matériel de classe pour s'assurer qu'il y ait des filles dans les exercices de maths, des garçons pour la lecture par exemple ou dans les coloriages. Ou, faire des activités sur les femmes qui ont réussi en maths, des hommes de lettres, des personnes de toutes les origines...

3. Intervention pour déconstruire les stéréotypes

Walton et al. convainquent des ingénieurs que leurs collègues femmes ont des capacités latentes (Walton et al., 2015). Dans la classe, Vallée et al. initient un débat avec des CE2-CM1 sur les stéréotypes (Vallée et al., 2020). Ils montrent que dans les groupes ayant reçu les débats, les filles réussissent mieux, et ont une plus grande confiance dans leurs capacités. Les garçons ne sont pas impactés négativement par le débat et, au contraire, cela crée une identité de groupe dans la classe.

- ⇒ *Pratiques possibles* : activités de déconstruction des stéréotypes avec les élèves.

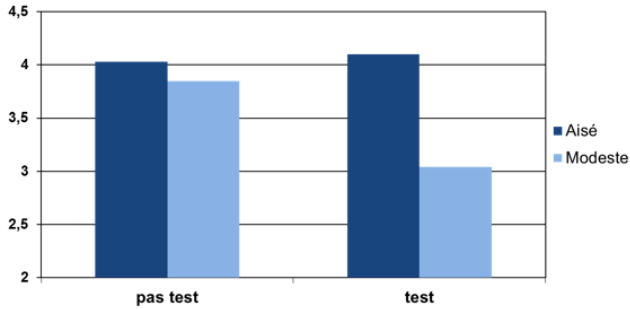


III. Activités pour la formation

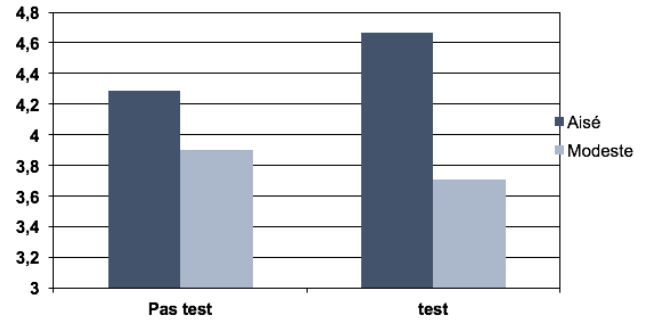
A. Nous possédons tous des stéréotypes

Darley et Gross montrent un film qui présente Hannah, une petite fille qui joue soit dans un environnement de milieu aisé ou de milieu modeste (Darley & Gross, 1983). Puis, certains regardent une situation où Hannah est en situation de test (ambigu, on a du mal à savoir sa performance), ou certains ne regardent pas.

Niveau de mathématique estimé



Niveau de français estimé



« Non, moi je ne juge pas sur ce type d'information... »

Un élève qui échoue à un exercice est nécessairement un élève qui n'a pas acquis la compétence requise pour réaliser cet exercice.

VRAI ou FAUX ?

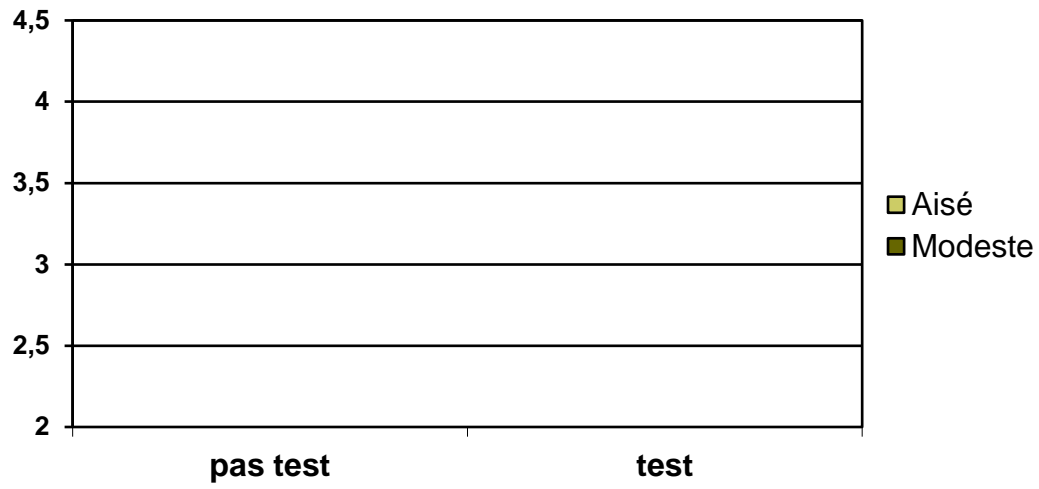
Un élève qui réussit un exercice est nécessairement un élève qui a acquis la compétence requise pour réaliser cet exercice.

VRAI ou FAUX ?

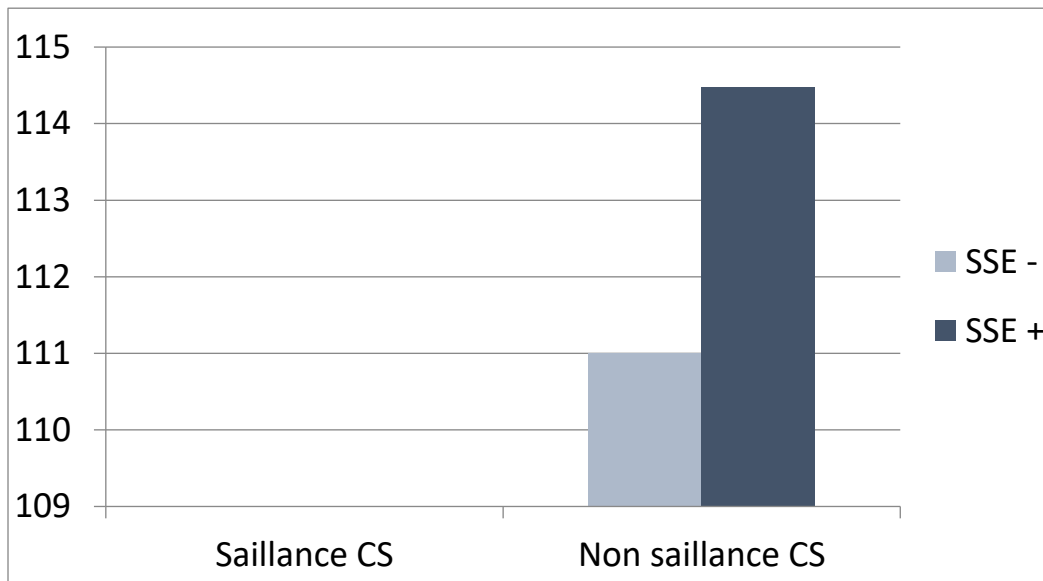
L'égalité des chances existe-t-elle ?



Expérience de Darley & Gross (1983) : Niveau de mathématiques estimé



Expérience de Kudrna et al (2010) : Intelligence estimée

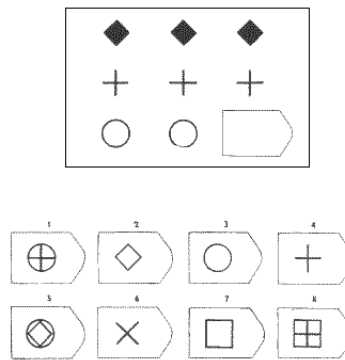




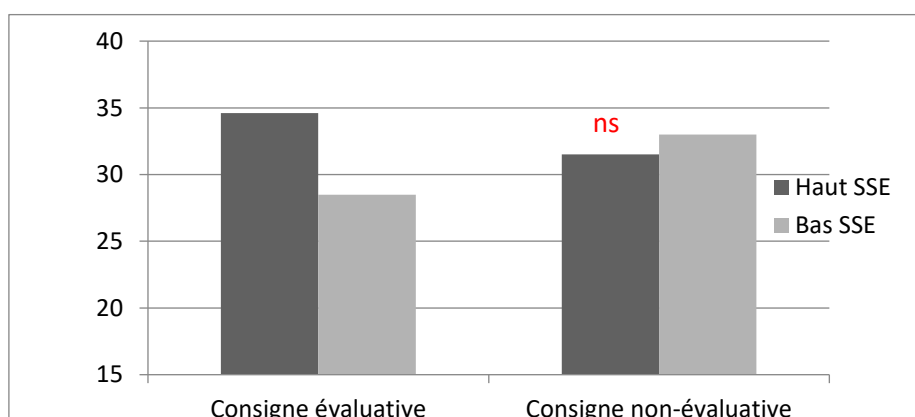
B. La menace des stéréotypes

Désert, M., Préaux, M., & Jund, R. (2009). So young and already victims of stereotype threat: Socio-economic status and performance of 6 to 9 years old children on Raven's progressive matrices. *European Journal of Psychology of Education, 24*, 207-218.

Les sujets de cette expérience étaient des enfants de CP et de CE2 (6 à 9 ans environ). Ils devaient réaliser la tâche des matrices de Raven, une tâche souvent utilisée par les psychologues pour mesurer les aptitudes intellectuelles des enfants, en particulier les capacités de raisonnement. Cette tâche consiste à présenter aux enfants une séquence logique. L'enfant doit trouver l'élément manquant. Un exemple d'items utilisés dans ce test est présenté ci-dessous. Dans ce test, les items sont présentés par ordre de difficulté croissante.



Dans cette étude, le test était présenté soit avec la consigne évaluative habituellement utilisée par les psychologues lorsqu'ils font passer ce test (« Nous allons faire ensemble une série d'exercices de manière à évaluer quels sont les domaines dans lesquels tu réussis et ceux dans lesquels tu échoues, afin de connaître tes forces et tes faiblesses ») soit avec une consigne non-évaluative (« Nous allons faire ensemble une série de jeux parce que nous avons créé ces jeux et que nous voudrions savoir s'ils sont adaptés aux enfants de ton âge »). La performance au test des matrices de Raven était ensuite mesurée. Les résultats sont représentés dans la figure ci-dessous. Ils montrent que sous consigne évaluative, les élèves issus de milieux favorisés ont mieux réussi le test que les élèves issus de milieux défavorisés. Cela pourrait être compris comme une indication de l'exactitude du stéréotype. Toutefois, les résultats obtenus sous consigne non évaluative indiquent clairement qu'il n'en est rien. En effet, lorsque la pression évaluative est retirée, tous les élèves réussissent aussi bien, quel que soit leur milieu d'origine.





C. Quelques risques des notes

Koestner, R., Zuckerman, M., & Koestner, J. (1987). Praise, involvement, and intrinsic motivation. *Journal of Personality and Social Psychology*, 53(2), 383-390.

Pour les tenants de la théorie de l'autodétermination, les enfants naissent avec une forte envie de découvrir le monde qui les entoure. Ainsi, pour eux, la motivation intrinsèque correspondrait-elle en quelque sorte à « l'état naturel » des individus. Toutefois, l'environnement joue un rôle fondamental dans le maintien ou la dégradation de cette motivation. En particulier, beaucoup de facteurs environnementaux ont pour effet de réduire la motivation intrinsèque en « désapprenant » aux enfants à réaliser des activités pour leur valeur intrinsèque. Pour ces auteurs, l'évaluation est l'un de ces facteurs.

Une expérience de Koestner, Zuckerman et Koestner (1987) permet d'illustrer ce propos. Dans cette recherche, des étudiants étaient amenés à réaliser une tâche (un puzzle). A la moitié des participants, le puzzle était présenté comme un jeu. A l'autre moitié, le puzzle était présenté comme étant diagnostique des capacités et comme mesurant en partie l'intelligence. Bien sûr, le puzzle était exactement le même dans les deux cas. Pour mesurer la motivation intrinsèque, les expérimentateurs faisaient en sorte de laisser aux participants une période de « temps libre » de 6 minutes au cours de laquelle ils pouvaient, s'ils le souhaitaient, poursuivre l'activité, mais également faire d'autres activités, ou ne rien faire. Les participants étaient observés à leur insu. Le temps passé à poursuivre les puzzles était utilisé comme mesure de motivation intrinsèque. Les résultats montrent que les personnes qui pensent que le puzzle est un test passent significativement moins de temps à poursuivre l'activité que celles qui pensent qu'elle est un jeu.

Si cette induction expérimentale de quelques secondes a suffi à créer des différences de motivation intrinsèque, alors on peut aisément imaginer à quel point, sur l'ensemble d'un cursus scolaire, la pression évaluative peut contribuer à altérer la motivation intrinsèque des élèves.



Liste des encadrés

Encadré 1. Les apports des comparaisons internationales sur les différences filles/garçons à l'école	3
Encadré 2. Une étude qui montre l'impact des stéréotypes d'origine sociale des enseignants sur la notation	8
Encadré 3. Une étude qui montre l'impact des stéréotypes d'origine sociale sur l'orientation scolaire	9
Encadré 4. La menace du stéréotype ne se limite pas aux membres des groupes historiquement discriminés.....	10
Encadré 5. Un article fondateur : <i>Pygmalion dans la classe</i> de Rosenthal et Jacobson.....	12
Encadré 6. Comment interpréter une taille d'effet.....	16
Encadré 7. Des exemples d'évaluation formative ayant des résultats positifs sur les élèves.....	18
Encadré 8. Remettre en perspective l'effet avec la méta-analyse de l' <i>Education Endowment Foundation</i>	18
Encadré 9. Fixer des objectifs d'apprentissage	23
Encadré 10. Quelques techniques pour vérifier la compréhension des élèves	25
Encadré 11. Le rôle des tests réalisés par l'enseignant dans l'évaluation formative.....	26
Encadré 12. Des méthodes pour adapter l'enseignement à l'apprenant	26
Encadré 13. Les techniques d'auto-évaluation	28

Liste des figures

Figure 1 - Ecart de compétences en mathématiques et en français entre les filles et les garçons en début de CP	2
Figure 2 - Ecart de compétences en mathématiques et en français entre les filles et les garçons en début de CE1 ..	2
Figure 3 - Proportion d'élèves en retard à l'entrée en sixième à la rentrée 2019 selon l'origine sociale de l'élève (%)	3
Figure 4 - Taux de réussite au baccalauréat en fonction de la CSP des parents	4
Figure 5 - Expérience 1 : lien entre nombre d'erreurs et la note comme fonction de l'origine sociale de l'élève	8
Figure 6 - Expérience 2 : Nombre de fautes trouvées dans la dictée comme une fonction de la classe sociale de l'élève et du type d'évaluation	8
Figure 7 - Expérience 3 : Nombre d'erreurs trouvés dans la dictée en fonction de la classe sociale de l'élève et de la fonction de l'évaluation.....	8
Figure 8 - Adéquation moyenne d'après les participants pour des carrières éducatives petites ou grandes en fonction de la classe sociale de l'élève cible	9
Figure 9 - Adéquation moyenne d'après les participants pour des carrières éducatives petites ou grandes en fonction de la classe sociale de l'élève cible et la fonction de l'école	9
Figure 10 - Nombre d'items corrects en fonction du groupe expérimentale et l'origine ethnique (Chateignier et al., 2009).....	11
Figure 12 - Performance à un test d'intelligence d'élèves stéréotypés et non stéréotypés en fonction de la condition de menace de stéréotype et de la performance passée	11
Figure 12 - Graphique des tailles d'effet des différentes interventions revues par l'EEF en fonction de leur coût (€)	19
Figure 13 - Cycle d'évaluation formative	22

Liste des tableaux

Tableau 1 - Aspects de l'évaluation formative	21
Tableau 2 - L'auto-efficacité et l'apprentissage	27



Bibliographie

- Alesina, A., Carlana, M., Ferrara, E. L., & Pinotti, P. (2018). *Revealing Stereotypes: Evidence from Immigrants in Schools* (Working Paper No. 25333; Working Paper Series). National Bureau of Economic Research. <https://doi.org/10.3386/w25333>
- Algan, Y., Constantin, J., Delpeuch, S., Huillery, É., & Prost, C. (2018). Impact de l'évaluation par compétence. *Focus Du CAE*, 027–2018.
- Allal, L. K. (1979). *Stratégies d'évaluation formative: Conceptions psychopédagogiques et modalités d'application*. Institut Romand de recherches et de documentation pédagogiques.
- Allal, L. K. (1988). Quantitative and qualitative components of teachers' evaluation strategies. *Teaching and Teacher Education*, 4(1), 41–51.
- Alter, A. L., Aronson, J., Darley, J. M., Rodriguez, C., & Ruble, D. N. (2010). Rising to the threat: Reducing stereotype threat by reframing the threat as a challenge. *Journal of Experimental Social Psychology*, 46(1), 166–171. <https://doi.org/10.1016/j.jesp.2009.09.014>
- Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype Susceptibility in Children: Effects of Identity Activation on Quantitative Performance. *Psychological Science*, 12(5), 385–390. <https://doi.org/10.1111/1467-9280.00371>
- Andre, T., Whigham, M., Hendrickson, A., & Chambers, S. (1999). Competency beliefs, positive affect, and gender stereotypes of elementary students and their parents about science versus other school subjects. *Journal of Research in Science Teaching*, 36(6), 719–747. [https://doi.org/10.1002/\(SICI\)1098-2736\(199908\)36:6<719::AID-TEA8>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1098-2736(199908)36:6<719::AID-TEA8>3.0.CO;2-R)
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence. *Journal of Experimental Social Psychology*, 38(2), 113–125. <https://doi.org/10.1006/jesp.2001.1491>
- Autin, F., Batruch, A., & Butera, F. (2019). The function of selection of assessment leads evaluators to artificially create the social class achievement gap. *Journal of Educational Psychology*, 111(4), 717–735. <https://doi.org/10.1037/edu0000307>
- Bagès, C., & Martinot, D. (2011). What is the best model for girls and boys faced with a standardized mathematics evaluation situation: A hardworking role model or a gifted role model? *British Journal of Social Psychology*, 50(3), 536–543.
- Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology*, 4(3), 359–373.
- Batruch, A., Autin, F., Bataillard, F., & Butera, F. (2019). School Selection and the Social Class Divide: How Tracking Contributes to the Reproduction of Inequalities. *Personality and Social Psychology Bulletin*, 45(3), 477–490. <https://doi.org/10.1177/0146167218791804>
- Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science*, 210(4475), 1262–1264. <https://doi.org/10.1126/science.7434028>
- Benmansour, N. (1999). *Motivational orientations, self-efficacy, anxiety and strategy use in learning high school mathematics in Morocco*.



- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8–21.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (Formerly: Journal of Personnel Evaluation in Education)*, 21(1), 5.
- Bloom, B. S. (1968). Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation Comment*, 1(2), n2.
- Borkowski, J. G. (1985). Signs of intelligence: Strategy generalization and metacognition. *The Growth of Reflection in Children*, 7, 105–144.
- Botelho, F., Madeira, R. A., & Rangel, M. A. (2015). Racial Discrimination in Grading: Evidence from Brazil. *American Economic Journal: Applied Economics*, 7(4), 37–52.
- Bruner, J. S., & others. (1966). *Toward a theory of instruction* (Vol. 59). Harvard University Press.
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, 79(4), 474.
- 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), 1–14.
- Butler, R., & Neuman, O. (1995). Effects of task and ego achievement goals on help-seeking behaviors and attitudes. *Journal of Educational Psychology*, 87(2), 261.
- Butler, R., & Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *Journal of Educational Psychology*, 78(3), 210.
- Callan, G. L., Marchant, G. J., Finch, W. H., & Flegge, L. (2017). Student and school SES, gender, strategy use, and achievement. *Psychology in the Schools*, 54(9), 1106–1122. <https://doi.org/10.1002/pits.22049>
- Callingham, R., & others. (2008). Dialogue and feedback: Assessment in the primary mathematics classroom. *Australian Primary Mathematics Classroom*, 13(3), 18.
- Cardinet, J. (1977). *Objectifs éducatifs et évaluation individualisée*. Institut Romand de recherches et de documentation pédagogiques.
- Cardinet, J. (1986). *L'évaluation en classe: Mesure ou dialogue?* IRDP.
- Carver, C. S., & Scheier, M. F. (1999). Themes and issues in the self-regulation of behavior. *Advances in Social Cognition*, 12(1), 1.
- Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, 27(2), 270–295.
- Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 83(1), 1–6.
- Chateignier, C., Dutrévis, M., Nugier, A., & Chekroun, P. (2009). French-Arab students and verbal intellectual performance: Do they really suffer from a negative intellectual stereotype? *European Journal of Psychology of Education*, 24(2), 219. <https://doi.org/10.1007/BF03173013>
- Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the Racial Achievement Gap: A Social-Psychological Intervention. *Science*, 313(5791), 1307–1310. <https://doi.org/10.1126/science.1128317>



- Cosnefroy, L., Perret, C., Huard, V., Demougeot-Lebel, J., & Conjard, P. (2011). *L'apprentissage autorégulé*. Editions L'Harmattan.
- Croizet, J.-C., & Claire, T. (1998). Extending the Concept of Stereotype Threat to Social Class: The Intellectual Underperformance of Students from Low Socioeconomic Backgrounds. *Personality and Social Psychology Bulletin*, 24(6), 588–594. <https://doi.org/10.1177/0146167298246003>
- Croizet, J.-C., & Dutrévis, M. (2004). Socioeconomic status and intelligence: Why test scores do not equal merit. *Journal of Poverty*, 8(3), 91–107.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438–481.
- Crouzevialle, M., & Butera, F. (2013). Performance-approach goals deplete working memory and impair cognitive performance. *Journal of Experimental Psychology: General*, 142(3), 666.
- Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–Gender Stereotypes in Elementary School Children. *Child Development*, 82(3), 766–779. <https://doi.org/10.1111/j.1467-8624.2010.01529.x>
- Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44(1), 20–33. <https://doi.org/10.1037/0022-3514.44.1.20>
- Davies, J., & Brember, I. (1998). National Curriculum Testing and Self-esteem in Year 2—The first 5 years: A cross-sectional study. *Educational Psychology*, 18(4), 365–375.
- Del Río, M. F., & Strasser, K. (2013). Preschool children's beliefs about gender differences in academic skills. *Sex Roles*, 68(3–4), 231–238.
- Direction de l'évaluation, de la prospective et de la performance. (2016). *TIMSS 2015 mathématiques et sciences—Évaluation internationale des élèves de CM1*. Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche.
- Direction de l'évaluation, de la prospective et de la performance. (2017). *PIRLS 2016—Évaluation internationale des élèves de CM1 en compréhension de l'écrit—Évolution des performances sur quinze ans*. Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche.
- Direction de l'évaluation, de la prospective et de la performance. (2020). *Repères et références statistiques sur les enseignements, la formation et la recherche 2019*. Ministère de l'Éducation Nationale de la Jeunesse et des Sports. <https://www.education.gouv.fr/reperes-et-references-statistiques-sur-les-enseignements-la-formation-et-la-recherche-2019-3806>
- Duru-Bellat, M. (1990). *L'école des filles: Quelle formation pour quels rôles sociaux?* l'Harmattan Paris.
- Eccles, J. S., & Jacobs, J. E. (1986). Social Forces Shape Math Attitudes and Performance. *Signs*, 11(2), 367–380.
- Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, 1(1), 3–14.
- Elton, L., & Laurillard, D. (1979). Trends in research on student learning. *Studies in Higher Education*, 4(1), 87–102.
- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, 53(1), 25–44. <https://doi.org/10.1016/j.jsp.2014.10.002>
- Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53(3), 199–208.



- Galdi, S., Cadinu, M., & Tomasetto, C. (2014). The roots of stereotype threat: When automatic associations disrupt girls' math performance. *Child Development, 85*(1), 250–263.
- Gecas, V. (1989). The Social Psychology of Self-Efficacy. *Annual Review of Sociology, 15*(1), 291–316. <https://doi.org/10.1146/annurev.so.15.080189.001451>
- Gentrup, S., & Rjosk, C. (2018). Pygmalion and the gender gap: Do teacher expectations contribute to differences in achievement between boys and girls at the beginning of schooling? *Educational Research and Evaluation, 24*(3–5), 295–323. <https://doi.org/10.1080/13803611.2018.1550840>
- Gibbons, R. E., Xu, X., Villafañe, S. M., & Raker, J. R. (2018). Testing a reciprocal causation model between anxiety, enjoyment and academic performance in postsecondary organic chemistry. *Educational Psychology, 38*(6), 838–856.
- Gordon, S. P., & Reese, M. (1997). High-stakes testing: Worth the price? *Journal of School Leadership, 7*(4), 345–368.
- Gresy, B., & Georges, P. (2013). Rapport sur l'égalité entre les filles et les garçons dans les modes d'accueil de la petite enfance, Inspection générale des affaires sociales [missionnée par le Ministère des Droits des femmes]. *Téléchargeable Gratuitement Sur Le Site Du Ministère Des Droits Des Femmes*.
- Guyon, N., & Huillery, E. (2016). Biased aspirations and social inequality at school: Evidence from french teenagers. *The Economic Journal*.
- Harlen, W., Crick, R. D., Broadfoot, P., Daugherty, R., Gardner, J., James, M., & Stobart, G. (2002). *A systematic review of the impact of summative assessment and tests on students' motivation for learning*.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children* (pp. xxiii, 268). Paul H Brookes Publishing.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Taylor & Francis.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.
- Hayek, A.-S., Toma, C., Guidotti, S., Oberlé, D., & Butera, F. (2017). Grades degrade group coordination: Deteriorated interactions and performance in a cooperative motor task. *European Journal of Psychology of Education, 32*(1), 97–112.
- Hayek, A.-S., Toma, C., Oberlé, D., & Butera, F. (2015). Grading hampers cooperative information sharing in group problem solving. *Social Psychology*.
- Hill, K. T., & Eaton, W. O. (1977). The interaction of test anxiety and success-failure experiences in determining children's arithmetic performance. *Developmental Psychology, 13*(3), 205.
- Hill, K. T., & Wigfield, A. (1984). Test anxiety: A major educational problem and what can be done about it. *The Elementary School Journal, 85*(1), 105–126.
- Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology, 49*(1), 4–14. <https://doi.org/10.1037/a0027238>
- Hoff, K., & Pandey, P. (2004). *Belief Systems and Durable Inequalities: An Experimental Investigation of Indian Caste*. The World Bank. <https://doi.org/10.1596/1813-9450-3351>



- Jones, E. E., & Berglas, S. (1978). Control of attributions about the self through self-handicapping strategies: The appeal of alcohol and the role of underachievement. *Personality and Social Psychology Bulletin*, 4(2), 200–206.
- Jussim, L., & Harber, K. D. (2005). Teacher Expectations and Self-Fulfilling Prophecies: Knowns and Unknowns, Resolved and Unresolved Controversies. *Personality and Social Psychology Review*, 9(2), 131–155. https://doi.org/10.1207/s15327957pspr0902_3
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968.
- Kellaghan, T., Madaus, G. F., & Raczek, A. E. (1996). *The use of external examinations to improve student motivation*. American Educational Research Association.
- Klute, M., Apthorp, H., Harlacher, J., & Reale, M. (2017). Formative assessment and elementary school student academic achievement: A review of the evidence. *Regional Educational Laboratory Central: Retrieved from: Ttp://Ies. Ed. Gov/Ncee/Edlabs*.
- Laveault, D. (2007). De la «régulation» au «réglage»: Élaboration d'un modèle d'autoévaluation des apprentissages. *Régulation Des Apprentissages En Situation Scolaire et En Formation*, 207–234.
- Lavy, V., & Sand, E. (2015). *On The Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers' Stereotypical Biases* (Working Paper No. 20909; Working Paper Series). National Bureau of Economic Research. <https://doi.org/10.3386/w20909>
- Lepareur, C., Gandit, M., & Grangeat, M. (2017). Évaluation formative et démarche d'investigation en mathématiques: Une étude de cas. *Éducation et Didactique*, 11(11–3), 101–120.
- Leyens, J.-P., Désert, M., Croizet, J.-C., & Darcis, C. (2000). Stereotype Threat: Are Lower Status and History of Stigmatization Preconditions of Stereotype Threat? *Personality and Social Psychology Bulletin*, 26(10), 1189–1199. <https://doi.org/10.1177/0146167200262002>
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Prentice-Hall, Inc.
- Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology*, 73(1), 91.
- Loganathan, P., Talib, C. A., Zawadzki, R., Aliyu, F., Ali, M., & Ibrahim, M. A. (2020). The Impact of Formative Assessment on High Achievers and Low Achievers—A Systematic Review. *PROCEEDING UMSURABAYA*.
- Lummis, M., & Stevenson, H. W. (1990). Gender differences in beliefs and achievement: A cross-cultural study. *Developmental Psychology*, 26(2), 254–263. <https://doi.org/10.1037/0012-1649.26.2.254>
- Ma, X., & Xu, J. (2004). The causal ordering of mathematics anxiety and mathematics achievement: A longitudinal panel analysis. *Journal of Adolescence*, 27(2), 165–179.
- Maass, A., & Cadinu, M. (2003). Stereotype threat: When minority members underperform. *European Review of Social Psychology*, 14(1), 243–275. <https://doi.org/10.1080/10463280340000072>
- Martinot, D., Bagès, C., & Désert, M. (2012). French children's awareness of gender stereotypes about mathematics and reading: When girls improve their reputation in math. *Sex Roles*, 66(3–4), 210–219.
- Marx, D. M., Stapel, D. A., & Muller, D. (2005). We can do it: The interplay of construal orientation and social comparisons under threat. *Journal of Personality and Social Psychology*, 88(3), 432.



- McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology*, 21(1), 89–101.
- McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology*, 39(1), 83–90.
- Menesses, K. F., & Gresham, F. M. (2009). Relative efficacy of reciprocal and nonreciprocal peer tutoring for students at-risk for academic failure. *School Psychology Quarterly*, 24(4), 266.
- Merton, R. K. (1948). The self-fulfilling prophecy. *The Antioch Review*, 8(2), 193–210.
- Ministère de l'éducation nationale, & Direction de l'évaluation, de la prospective et de la performance. (2020). *Filles et garçons sur le chemin de l'égalité 2020: De l'école à l'enseignement supérieur*.
- Muntoni, F., & Retelsdorf, J. (2018). Gender-specific teacher expectations in reading—The role of teachers' gender stereotypes. *Contemporary Educational Psychology*, 54, 212–220. <https://doi.org/10.1016/j.cedpsych.2018.06.012>
- Muntoni, F., Wagner, J., & Retelsdorf, J. (2020). Beware of Stereotypes: Are Classmates' Stereotypes Associated With Students' Reading Outcomes? *Child Development*, n/a(n/a). <https://doi.org/10.1111/cdev.13359>
- Muzzatti, B., & Agnoli, F. (2007). Gender and mathematics: Attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology*, 43(3), 747–759. <https://doi.org/10.1037/0012-1649.43.3.747>
- Null, L. D. H. (1992). *The effects of learning for mastery on first and second-grade decoding skill and general reading achievement*.
- OECD (2019). *PISA 2018 Results (Volume II)*. <https://doi.org/10.1787/b5fd1b8f-en> (OECD, 2005)
- OECD. (2005). Formative Assessment: Improving Learning in Secondary Classrooms. *Policy Brief*.
- Pansu, P., Régner, I., Max, S., Colé, P., Nezlek, J. B., & Huguet, P. (2016). A burden for the boys: Evidence of stereotype threat in boys' reading performance. *Journal of Experimental Social Psychology*, 65, 26–30. <https://doi.org/10.1016/j.jesp.2016.02.008>
- Papageorge, N. W., Gershenson, S., & Kang, K. M. (2019). Teacher Expectations Matter. *The Review of Economics and Statistics*, 102(2), 234–251. https://doi.org/10.1162/rest_a_00838
- Perry, N. E. (1998). Young children's self-regulated learning and contexts that support it. *Journal of Educational Psychology*, 90(4), 715.
- Piaget, J. (1975). *L'équilibration des structures cognitives. Problème central du développement*. Paris, France: Presses Universitaires de France.
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33.
- Plomin, R., & Stumm, S. von. (2018). The new genetics of intelligence. *Nature Reviews. Genetics*, 19(3), 148. <https://doi.org/10.1038/nrg.2017.104>
- Pressley, M., El-Dinary, P. B., Marks, M. B., Brown, R., & Stein, S. (1992). Good strategy instruction is motivating and interesting. *The Role of Interest in Learning and Development*, 333–358.
- Pulfrey, C., Darnon, C., & Butera, F. (2013). Autonomy and task performance: Explaining the impact of grades on intrinsic motivation. *Journal of Educational Psychology*, 105(1), 39.
- Quinn, J. M. (2018). Differential Identification of Females and Males with Reading Difficulties: A Meta-Analysis. *Reading and Writing*, 31(5), 1039. <https://doi.org/10.1007/s11145-018-9827-8>



- Régner, I., Steele, J. R., Ambady, N., Thinus-Blanc, C., & Huguet, P. (2014). Our future scientists: A review of stereotype threat in girls from early elementary school to middle school. *Revue Internationale de Psychologie Sociale*, 27(3), 13–51.
- Retelsdorf, J., Schwartz, K., & Asbrock, F. (2015). “Michael can’t read!” Teachers’ gender stereotypes and boys’ reading self-concept. *Journal of Educational Psychology*, 107(1), 186–194. <https://doi.org/10.1037/a0037107>
- Reynolds, S., Martin, K., & Groulx, J. (1995). Patterns of understanding. *Educational Assessment*, 3(4), 363–371.
- Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255.
- Rosenthal, R., & Et Jacobson, L. (1968). *Pygmalion in the Classroom: Teaching expectation and pupils intellectual development*. New York.
- Sadker, M., & Sadker, D. (1986). Sexism in the Classroom: From Grade School to Graduate School. *The Phi Delta Kappan*, 67(7), 512–515.
- Salikutluk, Z., & Heyne, S. (2017). Do gender roles and norms affect performance in maths? The impact of adolescents’ and their peers’ gender conceptions on maths grades. *European Sociological Review*, 33(3), 368–381.
- Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist*, 26(3–4), 207–231.
- Schunk, D. H. (1995). Self-efficacy, motivation, and performance. *Journal of Applied Sport Psychology*, 7(2), 112–137.
- Schunk, D. H. (1996). Goal and self-evaluative influences during children’s cognitive skill learning. *American Educational Research Journal*, 33(2), 359–382.
- Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading & Writing Quarterly*, 19(2), 159–172.
- Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology*, 18(3), 337–354.
- Sherman, D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., Tomassetti, S., Nussbaum, A. D., & Cohen, G. L. (2013). Deflecting the trajectory and changing the narrative: How self-affirmation affects academic performance and motivation under identity threat. *Journal of Personality and Social Psychology*, 104(4), 591–618. <https://doi.org/10.1037/a0031495>
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797.
- Stephens, N. M., Hamedani, M. G., & Destin, M. (2014). Closing the social-class achievement gap: A difference-education intervention improves first-generation students’ academic performance and all students’ college transition. *Psychological Science*, 25(4), 943–953.
- Stoet, G., & Geary, D. C. (2012). Can Stereotype Threat Explain the Gender Gap in Mathematics Performance and Achievement? *Review of General Psychology*, 16(1), 93–102. <https://doi.org/10.1037/a0026617>
- Twenge, J. M., & Campbell, W. K. (2001). Age and birth cohort differences in self-esteem: A cross-temporal meta-analysis. *Personality and Social Psychology Review*, 5(4), 321–344.



- Vallée, B., Magoutier, F., Voisin, D., & Montalan, B. (2020). Reducing the effects of the stereotype threat that girls perform less well than boys in mathematics: The efficacy of a mixed debate in a real classroom situation. *Social Psychology of Education, 23*(5), 1327–1341. <https://doi.org/10.1007/s11218-020-09583-x>
- von Soest, T., Wagner, J., Hansen, T., & Gerstorf, D. (2018). Self-esteem across the second half of life: The role of socioeconomic status, physical health, social relationships, and personality factors. *Journal of Personality and Social Psychology, 114*(6), 945.
- Walton, G. M., Murphy, M. C., & Ryan, A. M. (2015). Stereotype threat in organizations: Implications for equity and performance. *Annu. Rev. Organ. Psychol. Organ. Behav., 2*(1), 523–550.
- Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. Wittrock, *Handbook of Research on Teaching* (Pp 315-327).
- Wittenbaum, G. M., Hollingshead, A. B., & Botero, I. C. (2004). From cooperative to motivated information sharing in groups: Moving beyond the hidden profile paradigm. *Communication Monographs, 71*(3), 286–310.
- Yee, D. K., & Eccles, J. S. (1988). Parent perceptions and attributions for children's math achievement. *Sex Roles, 19*(5–6), 317–333.
- Yeh, S. S. (2001). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher, 30*(9), 12–17.
- Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology, 82*(1), 51.
- Zuckerman, M., Kieffer, S. C., & Knee, C. R. (1998). Consequences of self-handicapping: Effects on coping, academic performance, and adjustment. *Journal of Personality and Social Psychology, 74*(6), 1619.