



ECOLE NORMALE SUPERIEURE

DEPARTEMENT DES SCIENCES NATURELLES

SECTION DE BIOLOGIE-CHIMIE

SYLLABUS DE L'ECUE « EVALUATION DES APPRENTISSAGES »

CLASSE : BACCALAUREAT II

PAR

PROFESSEUR ETIENNE BARAHINDUKA

Etienne Barahinduka est Professeur associé en sciences de l'éducation à l'Ecole Normale Supérieure du Burundi. Il a déjà publié des articles, en collaboration avec des chercheurs de l'Université Paris Descartes, de l'Association Française d'Education Comparée et des Echanges, de l'Université Libre des Pays des Grands Lacs à Goma et de l'Université du Burundi sur les TICE, l'encadrement des recherches, l'éducation inclusive et l'enseignement des langues.

BUJUMBURA, AVRIL 2025

AVANT-PROPOS

A. Identification et description de l'ECUE

ECUE : BC2415-Evaluation des apprentissages-2 crédits

Titulaire : Professeur Etienne Barahinduka

Processus	Paramètres	Description
Elaboration	Titre de l'ECUE	Evaluation des apprentissages
	Objectif général	Concevoir le dispositif d'évaluation des apprentissages et des compétences en utilisant les techniques appropriées
	Objectifs spécifiques	A la fin de l'ECUE, l'étudiant sera capable de : -Définir et d'expliquer les concepts fondamentaux utilisés dans l'évaluation ; -Concevoir un dispositif d'évaluation en se référant à une taxonomie quelconque ; -Analyser et interpréter les résultats obtenus par les apprenants dans chaque évaluation ; -Organiser une remédiation; -Entreprendre l'auto-évaluation de leurs enseignements et évaluations.
	Pré-requis	-
	Organisation de l'ECUE	C M : 20 H, T P : 10 H
	Bref contenu de l'ECUE	Le cours consiste à clarifier les concepts fondamentaux utilisés dans l'évaluation, il présente les types, les fonctions et les formes d'évaluation. Il décrit les types de question, les principes d'élaboration des outils d'évaluation, les modalités de correction, de notation, d'analyse et d'interprétation des résultats de l'apprentissage. Il indique les principes de l'évaluation des compétences et attire l'attention sur quelques difficultés liées à l'évaluation.
	Références	De Landsheere, G. (1992). <i>Evaluation continue et examens</i> .

	bibliographiques	<i>Précis de docimologie</i>. Bruxelles : Edition Labor. Petitjean, B. (1984). Formes et fonctions des différents types d'évaluation. <i>Pratiques : linguistique, littérature, didactique</i>, 44, 5-20. Tanana, M. (2011). <i>Evaluation formative du savoir-faire des apprenants à l'aide d'algorithmes de classification. Application à l'électronique numérique</i>. Institut National des Sciences Appliquées de Rouen. Rouen, France. Tardif, J. (2006). <i>L'évaluation des compétences. Documenter le parcours de développement</i>. Montréal : Chenelière Education.
Méthodologie et supports pédagogiques	Méthodologie	Méthode interactive
	Supports	Syllabus
Modes d'évaluation	Evaluation formative	Travaux dirigés, travaux pratiques
	Evaluation sommative	Examen écrit

B. Objectifs de l'ECUE

1. Objectif général

L'ECUE vise à rendre les étudiants aptes à concevoir le dispositif d'évaluation des apprentissages et des compétences en utilisant les techniques appropriées.

2. Objectifs spécifiques

A la fin de l'ECUE, l'étudiant sera capable de :

- Définir et d'expliquer les concepts fondamentaux utilisés dans l'évaluation ;
- Concevoir un dispositif d'évaluation en se référant à une taxonomie quelconque ;
- Analyser et interpréter les résultats obtenus par les apprenants dans chaque évaluation ;

- Organiser une remédiation;
- Entreprendre l'auto-évaluation de leurs enseignements et évaluations.

C. Brève introduction de l'ECUE

L'ECUE consiste à clarifier les concepts fondamentaux utilisés dans l'évaluation, il présente les types, les fonctions et les formes d'évaluation. Il décrit les types de question, les principes d'élaboration des outils d'évaluation, les modalités de correction, de notation, d'analyse et d'interprétation des résultats de l'apprentissage. Il indique les principes de l'évaluation des compétences et attire l'attention sur quelques difficultés liées à l'évaluation.

D. Démarche méthodologique

La méthodologie est magistrale pour présenter les activités et le cours théorique. Elle est aussi participative, à travers des activités de recherche personnelle des étudiants à travers les lectures des ouvrages spécialisés, les échanges sur les travaux réalisés individuellement ou en groupe. Les supports sont constitués par une présentation power point et un syllabus.

E. Modalités d'évaluation des apprentissages

L'évaluation formative est constituée par des travaux individuels et/ou en groupes sur 40%.
L'évaluation sommative est constituée par un examen écrit sur 60%.

LISTE DES FIGURES

Figure 1. Tendance des notateurs à concentrer leurs appréciations aux échelons du centre de l'échelle.....	39
Figure 2. Notation des compositions sous l'influence d'un préjugé favorable ou défavorable....	42
Figure 3. La distribution des notes au test pour chacune des catégories d'élèves	48

LISTE DES TABLEAUX

Tableau 1. Effet de stéréotype ou d'inertie	40
Tableau 2. Distribution des notes attribuées à une même copie par différents correcteurs.....	44
Tableau 3. Nombre estimé de correcteurs pour obtenir une fidélité inter-correcteurs par domaine	46
Tableau 4. Les résultats obtenus pour l'ensemble des correcteurs par domaine	47

TABLE DES MATIERES

AVANT-PROPOS	i
A. Identification et description de l'ECUE	i
B. Objectifs de l'ECUE.....	ii
C. Brève introduction de l'ECUE.....	iii
D. Démarche méthodologique.....	iii
E. Modalités d'évaluation des apprentissages.....	iii
LISTE DES FIGURES	iv
LISTE DES TABLEAUX	v
TABLE DES MATIERES	vi
INTRODUCTION	1
Chapitre 1. Clarification des concepts fondamentaux utilisés dans l'évaluation	1
1.1. Docimologie, docimastique et doxologie.....	2
1.2. Examen et concours, observation et évaluation continues.....	2
1.3. Notes et scores.....	4
Chapitre 2. Présentation des types, des fonctions et des formes d'évaluation des apprentissages	5
2.1. Types d'évaluation des apprentissages	5
2.2. Fonctions d'évaluation des apprentissages	12
2.3. Formes d'évaluation des apprentissages.....	12
2.3.3. Travail écrit.....	14
Chapitre 3. Description des types de questions d'évaluation des apprentissages.....	18
3.1. Questions fermées	18
3.2. Questions ouvertes	20
Chapitre 4. Les principes d'élaboration des outils d'évaluation des apprentissages	22
4.1. La planification	22
4.2. La collecte des données	22
4.3. L'interprétation	23
4.4. Jugement et décision	23
4.5. Communication.....	23

4.6. Evaluation du processus	23
4.7. La rétroaction.....	23
Chapitre 5. Les principes de l'évaluation des compétences	24
5.1. Principe 1 : Rendre compte d'une progression dans une trajectoire de développement	24
5.2. Principes 2 : Accorder la priorité aux compétences et non aux ressources	25
5.3. Principes 3 et 4 : Déterminer les ressources mobilisables et mobilisées, combinables et combinées	25
5.4. Principe 5 : Circonscrire les situations de déploiement des compétences.....	26
5.5. Principe 6 : Documenter la trajectoire de développement à partir de preuves partagées.....	26
5.6. Principe 7 : Rendre compte de l'autonomisation dans le déploiement des compétences	27
5.7. Principe 8 : Recourir à des critères multiples dans chaque évaluation	27
5.8. Principe 9 : Intégrer les différences individuelles dans l'appréciation	27
Chapitre 6. Les difficultés liées à l'évaluation des apprentissages	28
6.1. L'histoire de la problématique	28
6.2. Les trois sources « d'erreur ».....	31
6.3. Quelques exemples de biais mis en évidence par la docimologie critique	37
6.4. En guise de conclusion	49
Références bibliographiques	51

INTRODUCTION

L'ECUE est constitué par sept chapitres : la clarification des concepts fondamentaux utilisés dans l'évaluation ; la présentation des types, des fonctions et des formes d'évaluation des apprentissages ; la description des types de questions d'évaluation des apprentissages ; les principes d'élaboration des outils d'évaluation des apprentissages ; les principes de l'évaluation des compétences et les difficultés liées à l'évaluation des apprentissages.

Le chapitre 1 définit la notion de docimologie et le terme évaluation, différencie les termes examen, concours, test, examen interne, examen externe, note et score.

Chapitre 1. Clarification des concepts fondamentaux utilisés dans l'évaluation

Objectif général

Le chapitre 1 vise à rendre les étudiants aptes à définir et à expliquer les concepts fondamentaux utilisés en évaluation des apprentissages.

Objectifs spécifiques

Au terme du chapitre 1, l'étudiant sera capable de :

- définir la notion de docimologie ;
- différencier les termes examen, concours et test ;
- définir le terme évaluation ;
- différencier l'examen interne et de examen externe ;
- différencier note et score.

Avant d'entrer en profondeur du sujet, il importe à l'étudiant de s'approprier de certains concepts souvent utilisés dans le domaine de l'évaluation. Sans être exhaustif, nous pouvons citer les notions de docimologie, examen, concours, test, examen interne, examen externe, évaluation, note et score.

1.1. Docimologie, docimastique et doxologie

1.1.1. Docimologie

La docimologie est une science qui a pour objet l'étude systématique des examens, en particulier des systèmes de notation, et du comportement des examinateurs et des examinés (Dupeyroux, 1950).

1.1.2. La docimastique

La docimastique est la technique des examens. Au début, la docimologie a revêtu un caractère négatif en critiquant les modes de notation et en montrant expérimentalement le manque de fidélité des examens. Par la suite, elle est entrée dans une phase constructive en essayant de proposer des méthodes et des techniques de mesure plus objectives ou, au moins, plus rigoureuses, et en mettant au point les moyens de rendre les notes comparables, de façon à assurer plus de justice scolaire (Mouele, 2017).

1.1.3. La doxologie

La doxologie est l'étude systématique du rôle que l'évaluation joue dans l'éducation scolaire. Guillaumin (1968) lui assigne, notamment, les objectifs suivants : étude des effets inhibiteurs ou stimulants des différentes formes d'examens ; étude des réactions émotionnelles des élèves et de là, des réactions intellectuelles aux jugements du maître ; étude des effets de l'opinion du maître concernant les élèves sur son enseignement et sur l'apprentissage scolaire ; étude des processus mis en jeu et des effets obtenus par l'automatisation, par l'internotation, par la notation d'équipe et par l'absence de notation (Tozzi, 2004).

1.2. Examen et concours, observation et évaluation continues

1.2.1. Examen

L'examen et le concours sont difficilement dissociables de l'idée d'épreuve, devenue d'ailleurs leur synonyme, après avoir désigné plus généralement la souffrance, le malheur, le danger qui relèvent le courage et la résistance. Dans l'examen, l'admission est déterminée par une note, généralement fixée à 50%, que le candidat doit atteindre ou dépasser, tandis que dans le concours, le nombre de places proposées est fixé d'avance. La présence de la menace, du danger, celui du refus, est, dans les deux cas, indéniable (Lopez et Laveault, 2008).

Dans la notion de concours et d'examen, nous percevons une charge agressive totalement absente des concepts d'observation et d'évaluations continues. La sérénité, la bienveillance, l'indulgence aussi, imprègnent le maître qui suit, avec une sympathie d'où toute sévérité n'est

pas exclue, le long cheminement de ses élèves vers l'équilibre du moment et la pleine accession à l'état d'adulte, dans l'avenir (Morrisette, 2009). Les examens marquent les fins d'étapes ; les concours ouvrent les portes aux élus : ce sont des points dans la durée, des événements dans le processus de l'éducation.

L'apprentissage qui précède ces événements consiste fondamentalement en une succession continue de comportements et de feedbacks, c'est-à-dire d'informations éclairant sur leur validité, leur pertinence. Il ne nous appartient pas d'engager ici une longue discussion des modalités et des effets du renforcement des conduites. L'important, c'est que, sans lui, l'apprentissage ne semble pas pouvoir se produire (Barroso, 2017).

L'évaluation, au sens restreint que nous lui donnons dans le présent chapitre, mérite donc une place importante dans l'enseignement, dont elle fait partie intégrante. Elle a toujours, directement ou indirectement, rapport avec le progrès, en extension ou en qualité de l'apprentissage (Endrizzi et Rey, 2008).

L'évaluation joue trois rôles :

1°). Un rôle pronostique : l'élève est-il pourvu des qualités intellectuelles et caractérielles, et des connaissances nécessaires pour aborder une matière nouvelle ou un cycle d'études supérieures ? Est-il là où il doit se trouver ? Répondre à ces questions équivaut à prédire le succès dans l'étape qui va commencer (Lopez et Allal, 2011).

2°). Un rôle de jaugeage : contrôle des acquisitions ; évaluation du progrès, cas où l'on compare l'élève à lui-même et situation de l'élève à un moment donné : (dans sa classe ou son groupe de travail ?; dans l'ensemble des classes parallèles d'une même école ?; dans des ensembles plus vastes : ville, canton, province, pays ?). Il ne s'agit pas nécessairement de procéder à un examen ou à un concours, mais de faire le point, de déterminer la position relative (Mukulu et Cibangu, 2014).

3°). Un rôle diagnostique. Pourquoi un apprentissage parfait ne s'est-il pas produit ? Quelles matières ou techniques l'étudiant domine-t-il insuffisamment, quels sont les processus mentaux en cause ? (Issaieva et Crahay, 2010).

Les instruments nécessaires à l'évaluation, à l'examen ou au concours ne diffèrent pas toujours, mais bien la façon de les utiliser. Aussi, pour éviter des précisions verbales fastidieuses emploierons-nous, dans les pages qui suivent, le mot examen dans deux sens différents, que le contexte éclairera toujours : examen proprement dit et, plus généralement, toute procédure pédagogique ayant une mesure d'apprentissages ou de connaissances pour objet (Jouquan, 2002).

Au sens strict, l'examen interne dans une branche est construit et noté par le maître qui l'a enseignée, et subi par les élèves qui ont reçu cet enseignement, dans le cadre de la classe ou de l'école. Au sens plus large, on qualifie d'internes les examens organisés indépendamment dans chaque école, qu'il existe ou non une coordination ou une unification par branche et par niveaux et sections (Barroso, 2017).

Par examens externes, on désigne les épreuves organisées et notées par des jurys indépendants des écoles, à l'échelon local, régional ou national. Les plus connues de ces épreuves sont celles du baccalauréat français ; citons encore, en Belgique, les examens cantonaux, en fin d'études primaires, et, en Angleterre, « *l'Eleven + Examination* » à l'entrée dans l'enseignement secondaire (Yerly, 2017).

1.3. Notes et scores

La distinction entre notes et scores rendrait, pensons-nous, un grand service en éducation. Dans une dictée, l'élève peut commettre un certain nombre de fautes dont il ne nous appartient guère d'évaluer l'existence : elles sont, ou ne sont pas. Toutefois, le nombre de fautes relevées n'a, en soi, aucune signification éducative : faire cinq fautes dans la dictée de Mérimée témoigne d'une extraordinaire connaissance des arcanes de l'orthographe ; en d'autres conditions, le même résultat annonce une faiblesse grave. Une information relativiste est donc aussi nécessaire à un test ou à toute forme d'évaluation par compte ou décompte de points, selon les règles fixés : les nombres de fautes en dictées, les résultats à un test standardisé, etc. (Yerly et Laveault, 2020).

Par note, il faut entendre une appréciation synthétique traduisant une performance dans le domaine de l'éducation. La note peut être objective ou subjective, mais elle est toujours relative. Attribuer la note A à un élève dont la performance se situe à tel niveau dans un étalonnage national relève de la première catégorie ; marquer sa composition « d'un très bien, bien, insuffisant, médiocre, etc. » relève de la seconde (Dierendonck, Burton et Wanlin, 2009).

Le chapitre 2 présente les types d'évaluation des apprentissages, les fonctions d'évaluation des apprentissages et les formes d'évaluation des apprentissages.

Chapitre 2. Présentation des types, des fonctions et des formes d'évaluation des apprentissages

Objectif général

Le chapitre 2 vise à rendre les étudiants aptes à présenter les types, les fonctions et les formes d'évaluation des apprentissages.

Objectifs spécifiques

A la fin du chapitre 2, les étudiants seront capable de :

- présenter les types d'évaluation des apprentissages ;
- présenter les fonctions d'évaluation des apprentissages ;
- présenter les formes d'évaluation des apprentissages.

2.1. Types d'évaluation des apprentissages

2.1.1. L'évaluation prédictive

Appelée aussi évaluation diagnostic-pronostic, elle est la procédure qui consiste à déterminer les capacités requises pour débiter un apprentissage ou pour regrouper les élèves par niveau. Evaluation ponctuelle, externe, elle intervient souvent au moment de l'orientation des élèves vers les différentes filières et relève alors de la fonction sociale de l'évaluation. Elle peut être aussi liée à la fonction pédagogique quand il s'agira par exemple de déterminer les acquis des élèves en début d'année ou à l'initiale d'une séquence d'apprentissage par des moyens divers : tests, questionnaires (Guimard, Cosneyfroy et Florin, 2007).

2.1.2. L'évaluation sommative

Evaluation finale, externe, elle intervient sous forme de bilan en fin d'apprentissage avec une visée de contrôle, de certification. C'est une forme d'évaluation qui est à référence normative critériée, ce qui signifie qu'elle utilise des barèmes (notes, moyennes) faisant apparaître des différences individuelles (Hameline, 1979). Son but est d'attribuer une place dans un groupe, de classer, par rapport à une population de référence afin de permettre une prise de décision (accès à la classe supérieure, attestation d'une maîtrise de capacités, orientation vers une filière, etc.).

Elle se manifeste par des épreuves répétitives de vérification des connaissances ou des savoirs-faire (évaluation dite « continue » avec ou non des unités capitalisables) et surtout par la voie bien connue des examens ou des concours (évaluation dite « ponctuelle ») (Hameline, 1982a).

Qu'elle intervienne en fin de trimestre, en fin d'année ou en fin de cursus, l'évaluation sommative a donc essentiellement une fonction de sélection, de tri, de différenciation. Elle s'appuie principalement sur la notation des productions écrites des élèves, notation que l'on sait imprécise, peu fiable et dont la validité a souvent été remise en cause. Elle reste néanmoins dans la pédagogie traditionnelle la pratique dominante qui prend la forme du devoir-sanction mensuel ou bimensuel auquel on attribue une note, justifiée par des appréciations (Lopez et Laveault, 2008).

Que lui reproche-t-on?

- de mesurer, dès le début des apprentissages, la totalité des objectifs (évaluation globale) ;
- de ne pas délimiter les objectifs à atteindre au moyen de travaux intermédiaires au cours desquels sont précisées les procédures opérationnelles ainsi que les conditions de réussite (évaluation synthétique) ;
- de faire abstraction des démarches d'apprentissage du groupe ou des individus (évaluation extérieure) ;
- de ne pas expliciter les contrats de travail et les outils d'évaluation (évaluation implicite) ;
- de se limiter à comparer les performances d'un individu à celle des autres à partir d'un produit fini (évaluation normative) (Bercier-Larivière et Forgette-Giroux, 1999).

L'évaluation sommative produit des effets pervers que tout enseignant connaît et que rappellent Viallet et Maisonneuve (1980) :

- Les techniques (examens, tests, etc.) utilisées sont souvent peu valides en ce sens qu'on ne sait pas toujours ce que /on mesure et que les mesures ne sont ni représentatives, ni stables, ni toujours justifiées ;
- On analyse rarement les résultats des examens, soit par manque de temps, soit par manque de moyens, et ils sont donc rarement améliorés ;
- On a peu d'occasion de vérifier ce que les élèves conservent d'un enseignement après l'examen, en particulier si les capacités évaluées perdurent et servent dans la vie professionnelle ou pour la suite des études ;

-Il arrive qu'on accorde de fait une valeur absolue à des mesures toutes relatives, ce qui amène des décisions arbitraires et automatiques (par exemple, on dira qu'à 60 % et plus de « bonnes réponses », un élève aura réussi, sinon devra recommencer ; il faudra avoir au moins 10 sur 20 pour passer en classe supérieure) ;

-L'échec ou la réussite d'un élève dépend dans la plupart des cas de sa position dans la distribution des notes, plutôt que de sa capacité à agir ou à créer. Les épreuves d'évaluation induisent des comportements détournés et non souhaitables de la part des élèves. Par exemple, ces derniers travailleront uniquement en fonction de la note, en cherchant à deviner les questions d'une épreuve ou d'un examen. Ils étudieront plus le professeur que la matière, etc. ;

-L'anxiété provoquée chez certains par les examens et plus généralement par les épreuves d'évaluation, est un facteur qui diminue la fiabilité des résultats. L'évaluation sommative ne permet d'identifier les difficultés d'apprentissage et les parties de cours non assimilées qu'en fin d'une période d'enseignement; et il est alors trop tard pour apporter des corrections ;

-L'élève contrôlé seulement à la fin d'une période d'enseignement, n'a guère pu développer sa capacité d'auto-évaluation, n'ayant pas été préparé à cet exercice. Cela entraîne souvent une divergence entre le jugement qu'il porte sur lui-même et son résultat à l'examen. De là vient l'opinion largement répandue de l'arbitraire des examens et du système scolaire.

2.1.3. Evaluation formative

A la différence de l'évaluation sommative dont le but est de dresser un constat des acquis et qui s'accommode fort bien d'un mode de travail de type transmissif, l'évaluation formative, inexistante ou presque dans les pratiques traditionnelles, fait système avec les modes de travail pédagogique se référant à une théorie « constructiviste » de l'apprentissage (pédagogie différenciée, travail en projet, etc.) (Halima, 2022).

Cette dernière postule que tout individu apprenant, dès l'instant où il est mis en demeure de résoudre un problème, pour une part assimile (incorporation du donné de l'expérience dans les Figures de conduite qu'il possède), pour une autre s'accommode (modification des Figures connus pour appréhender des objets nouveaux). Encore faut-il :

- 1) fournir aux élèves des outils qui consolident les conduites d'accommodation ;
- 2) les mettre dans des situations pédagogiques qui favorisent ces conduites.

On comprend que l'évaluation formative occupe une place importante dans le travail en projet (volonté commune d'agir sur la motivation des élèves, de leur permettre de donner un sens à leurs activités, de faciliter leur prise de conscience des acquis et des erreurs, ou leur degré d'implication dans le travail, etc. (Saiah, 2016).

Selon Allal (1981), en modifiant la tâche ou la situation d'apprentissage, on cherchera, dans une perspective cognitiviste, à créer un décalage optimal entre la « structure du sujet » (c'est-à-dire les représentations et les procédures déjà élaborées par l'élève) et la « structure de la tâche » (c'est-à-dire les exigences externes qui impliquent l'élaboration de représentations et de procédures d'un ordre supérieur). Le décalage entre l'élève et la tâche sera optimal lorsque les informations fournies par la tâche peuvent être assimilées et traitées par l'élève, mais font surgir en même temps des contradictions ou des conflits qui suscitent un dépassement (restructuration) de son mode de traitement actuel (Traoré et Diarra, 2024).

Rendre ainsi, l'élève acteur de son apprentissage implique - c'est là l'enjeu essentiel de l'évaluation formative :

- 1) qu'il soit mis en position de réaliser des tâches complexes (par exemple, écrire une nouvelle, une nouvelle historique, une pièce de théâtre, un dossier documentaire, etc.) (Petitjean, 1984) ;
- 2) qu'il s'approprie les critères de réalisation de la tâche. Cette seconde exigence n'est possible qu'à condition que l'enseignant veille à expliciter: les consignes de réalisation des tâches ; les caractéristiques de l'objet à produire ; les critères de réussite et d'erreur ; les modalités d'implication dans le travail ; la signification que les élèves prêtent à la tâche (représentation, difficultés rencontrées, intérêts pris, etc.) ; et les critères d'évaluation (Scallon, 2004).

Sachant que ceux-ci se partagent entre des critères afférant aux objets de savoirs à maîtriser et des critères relevant des attitudes et des comportements dans le travail (voir les questionnaires élaborés). C'est en ce sens que l'évaluation formative est interne au processus d'apprentissage, qu'elle est continue, plutôt analytique et centrée plus sur apprenant que sur le produit fini (Baribeau, 2015).

Ce décentrement qu'opère l'évaluation formative n'est pas sans effets sur le statut de l'enseignant et sur ses stratégies pédagogiques :

- a) il accepte un partage du savoir en permettant aux élèves de manipuler les critères d'évaluation - ils ne sont plus la seule propriété de l'enseignant- voire de les construire progressivement ;

b) il diversifie les situations de travail et de communication : situation « conseil coopératif » au cours duquel l'enseignant et les élèves négocient le projet, en font une évaluation prévisionnelle et engagent des activités ; situation de production (individuelle ou en groupe) au cours de laquelle les élèves se partagent le travail et le réalisent ; situation « cours » dans laquelle l'enseignant transmet les savoirs organisateurs ; situation d'évaluation collective ou individuelle à fonction socialisatrice, informative, réflexive, régulatrice ;

c) il diversifie les types de tâches. Celles-ci se partagent globalement en deux types : les problèmes et les exercices (Vigner, 1984).

Les problèmes : de forme traditionnelle comme la rédaction ou la dissertation ou de forme renouvelée comme l'écriture d'un texte fictionnel ou la réalisation d'un dossier documentaire, les problèmes sont des activités synthétiques, complexes, divergentes, exigeant la mobilisation de connaissances et de capacités multiples. Si l'on prend l'exemple de l'écriture d'un conte, sa réalisation nécessite la maîtrise simultanée d'opérations se situant à différents niveaux : au niveau linguistique (orthographe, syntaxe, vocabulaire, etc.) ; au niveau textuel (relations interphrastiques et articulation entre superstructures textuelles (récit, argumentation, etc.) et leur manifestation au niveau micro-textuel (temps, thématisation, etc.) ; au niveau discursif (contenu sémantique du texte et choix organisationnel dans les informations disponibles (scripts de routine, arrangements inédits, ellipses, etc.) ; au niveau sémiotique (détermination du texte par son appartenance générique (conte merveilleux, conte d'avertissement, conte satirique, parodie, etc.) ; au niveau pragmatique (surdétermination du texte par les éléments de la situation de communication : place du narrateur, ses intentions communicatives, les images du récepteur, etc.) (De Ketele et Gerard, 2005).

Tes exercices : quelle que soit leur forme (exercice de répétition, de substitution, de production, de conceptualisation, de systématisation, etc.) l'exercice est une activité ponctuelle et analytique qui a pour fonction d'être convergente (application d'une règle, vérification d'un savoir, entraînement à un mécanisme, etc.). Insérés à l'intérieur d'une macro-stratégie d'enseignement commune à la classe (par exemple, écrire l'éditorial d'un journal), ils sont collectifs ou individuels et visent : à favoriser l'appropriation d'un éventail large de micro-objectifs ; et à affermir les acquisitions.

C'est ainsi - avant de faire écrire l'éditorial - qu'on entraînera les élèves à la maîtrise de l'argumentation par l'intermédiaire d'exercices divers-: repérage, dans un texte argumentatif, des

connecteurs utilisés et analyse de leur valeur sémantique (opposition, conséquence, hypothèse, addition, etc.) ; rédaction d'un paragraphe en utilisant différentes stratégies cognitives (association, accumulation, comparaison, opposition, etc.) ; exercices à trous où il s'agit de retrouver les connecteurs effacés, etc. (Howe et Ménard, 1993).

d) il diversifie l'ordre des apprentissages. Rompant le Figure dominant leçon/application/vérification, il placera les moments d'objectivation aussi bien en amont de l'activité de production (lecture d'un texte d'auteur qui contient un procédé) qu'au milieu de son effectuation (lecture d'un texte écrit par les élèves) qu'en son aval (évaluation intermédiaire ou finale) (Morrissette, 2010).

C'est en ce sens que l'évaluation formative est continue et interactive par l'intermédiaire d'un réseau de communication diversifié (relations professeur/ classe, professeur/élève, élève/élèves, etc.).

Ayant proposé aux élèves un matériel pour un travail individuel ou en groupe, le maître circule dans la classe. En observant différents élèves, il cherche à identifier ceux qui n'arrivent pas à avancer dans les tâches proposées. A travers des échanges avec ces élèves, l'enseignant essaye de formuler des hypothèses diagnostiques et de modifier la situation en conséquence, soit en posant des questions ou en offrant des suggestions, soit en proposant un autre matériel soit en réorganisant les activités proposées, etc. Sachant toutefois que le maître ne peut être présent partout lorsque les élèves travaillent individuellement ou en groupe, il faut envisager des moyens complémentaires pour assurer l'intégration de l'évaluation formative aux activités pédagogiques. Un moyen serait de développer des formes de collaboration et d'interaction entre élèves où le rôle d'observateur animateur joué par le maître peut être assumé, du moins en partie, par des élèves (Morrissette et Maheux, 2007).

e) il établit des relations entre ce que maîtrisent les élèves, les caractéristiques de la tâche et la fixation d'objectifs ultérieurs en proposant, quand il est nécessaire, des activités de remédiation : exercices de correction et d'amélioration, questionnaires sur les caractéristiques de la tâche et les difficultés de sa réalisation (Allal et Lopez, 2005).

C'est pourquoi l'évaluation formative est ponctuelle et rétroactive. Ainsi si l'on estime que les difficultés rencontrées par l'élève sont dues à un manque de maîtrise de prérequis, on proposera des exercices de rattrapage. De plus pour chaque objectif non atteint, on organisera des activités de remédiation (exercices individuels, travaux en groupes, lectures complémentaires), etc. Pour

que ces adaptations de l'enseignement ne se limitent pas à des variations d'ordre purement quantitatif (répétition de la leçon initiale, davantage d'exercices d'un même type, etc.) on cherchera à diversifier les modalités de présentation et la nature des tâches proposées à l'élève.

Ce n'est qu'à ce prix que l'évaluation devient partie intégrante de l'apprentissage au sein duquel elle introduit diverses régulations : l'objectivation des critères de réussite rend les élèves capables de mettre à distance leurs productions ; la nomination des procédés, des outils et des tâches facilite le transfert des acquis dans d'autres situations d'apprentissage et dans les situations non scolaires d'utilisation ; l'objectivation des critères de tâches permet aux élèves de « planifier » leur action, d'agir sur la maîtrise du temps et des moyens (informations, supports, etc.) ; et l'objectivation des critères, enfin, permet de réduire le décalage inévitable entre les activités parasites et les activités réalisées (Aylwin, 1995).

On le voit, l'évaluation formative a un triple intérêt :

- 1) effectuée en cours d'apprentissage, dans la mesure où elle ne se limite pas à enregistrer les résultats, dans la mesure aussi où elle repose sur des critères non arbitraires et explicites, elle facilite les processus d'acquisition (appropriation des mécanismes mis en jeu par la tâche et des propriétés du sujet traité).
- 2) effectuée en cours d'apprentissage, elle permet aux élèves de s'initier progressivement à l'auto-évaluation (autonomie conquise dans la réalisation des tâches et dans le contrôle de cette réalisation).
- 3) effectuée en cours d'apprentissage, elle conduit à un recueil d'informations qui permettent aux enseignants de moduler leur pratique en faisant de l'évaluation même un objet d'observation et d'analyse de leur démarche d'enseignement (Deaudelin, Desjardins, Dezutter, Thomas, Corriveau, Lavoie, Bousadra et Hébert, 2007).

Pour résumer, l'évaluation formative assure tout à la fois une fonction de renforcement (est renforçatrice, l'évaluation qui valorise la réponse de l'élève et qui augmente aussi la probabilité d'une réponse du même ordre par la suite), une fonction de correction (l'élève qui a compris ses erreurs peut y remédier plus aisément) et, enfin, une fonction de régulation de l'acte pédagogique (ajustement permanent, entre les objectifs poursuivis et les stratégies utilisées pour les atteindre). En théorie, l'évaluation formative précède donc l'évaluation sommative qu'elle prépare. Or, il faut bien constater que dans les pratiques pédagogiques dominantes, l'évaluation sommative tient lieu d'évaluation formative avec toutes les conséquences que l'on connaît (Hameline, 1982b).

Il reste que, bien que minoritaires, des expérimentations ont été tentées. Elles prouvent qu'il est possible d'engager les processus de transformation sans attendre d'hypothétiques bouleversements (Leselbaum, 1982).

2.2. Fonctions d'évaluation des apprentissages

Il existe trois fonctions de l'évaluation : formative, sommative et pronostique : formative (soutenir les apprentissages; réguler l'apprentissage et l'enseignement) ; sommative (reconnaître officiellement les apprentissages; certifier) ; pronostique (orienter l'avenir de la formation d'une personne ; sélectionner ; et orienter). La finalité générale de toute forme d'évaluation est de réguler les processus de formation (Allal, 2008).

C'est réguler les conditions d'apprentissage, c'est-à-dire varier les méthodes d'enseignement-apprentissage, encadrer efficacement l'étudiant, offrir des rétroactions, etc. La régulation apportée par l'évaluation *formative* consiste en l'adaptation des activités d'enseignement et d'apprentissage afin de favoriser la progression des élèves vers les objectifs de formation (Allal, 2007, p. 141).

2.3. Formes d'évaluation des apprentissages

Selon Nizet et Leroux (2015), les formes d'évaluation des apprentissages sont les suivantes : l'examen et le test, les exercices, le travail écrit, la situation authentique, les activités collaboratives et la communication.

2.3.1. Examen et test

A. Définitions opérationnelles

Activité demandant aux étudiant(e)s la restitution d'éléments d'apprentissage ou leur reproduction, dans des conditions restreintes de temps et d'espace (Fontaine et Loye, 2017).

B. Exemples de méthodes

Examen ou jeu-questionnaire (VF, QCM, association, développement court ou long) (Yerly, 2017).

C. Avantages

Les examens et les tests sont rapides à corriger ou la correction automatisée ; offrent plusieurs types de questions ; permettent d'évaluer un grand nombre d'étudiant(e)s ; l'interprétation objective des réponses (VF, QCM, association) ; utiles pour une évaluation diagnostique des prérequis et des erreurs fréquentes ; le QCM est utile en évaluation formative ; et la possibilité d'offrir une rétroaction individuelle automatisée pour les questionnaires numériques (Forgette-Giroux, Simon et Bercier-Larivière, 1996).

D. Limites

Les limites des examens et des tests sont les suivantes : la construction parfois longue et complexe ; n'évaluent pas des compétences complexes ; évaluent des apprentissages en surface ; font peu appel à l'analyse, au jugement et au sens critique ; encouragent les étudiants (es) à apprendre « par coeur » ; mesurent les apprentissages de façon isolée les uns des autres ; doivent être ajustés, modifiés ou changés d'année en année pour contrer le plagiat ; et sont un type d'évaluation sommative anxiogène pour les étudiant(e)s, car il n'y a pas de chance de se reprendre (Charlin, Bordage et Von Der Vleuten, 2003).

2.3.2. Exercice

A. Définitions opérationnelles

L'exercice est une activité structurée et adaptée permettant aux étudiant(e)s de consolider leurs apprentissages ; et demande de restituer, d'appliquer ou d'analyser ce qui a été appris dans une séquence d'apprentissage. Se réalise individuellement ou en équipe (Laurier, 2014).

B. Exemples de méthodes

Les exercices sont constitués par la résolution de problèmes, les fiches de lecture, de laboratoire, des travaux pratiques, des cartes conceptuelles, la mise en situation, l'expérimentation, la simulation, etc. (Fontaine, et Loye, 2017).

C. Avantages

Les exercices mettent les étudiant(e)s en action ; permettent de diagnostiquer des erreurs d'apprentissages ; mesurent des apprentissages plus complexes ; favorisent l'engagement des

étudiant(e)s ; sont utiles en évaluation formative ; et permettent la formulation d'une rétroaction lors de la correction (Laurier, 2014).

D. Limites

Les exercices peuvent nécessiter beaucoup d'organisation et de matériel (ex. : laboratoires, collecte de données, terrain) ; peuvent occasionner une surcharge de travail ; sont parfois longs à corriger ; nécessitent une grille de vérification ou de correction avec des critères précis ; répétitifs: peuvent encourager les étudiant(e)s à « reproduire une recette » sans nécessairement la comprendre ou l'analyser (Fontaine, Savoie-Zajc et Cadieux, 2013).

2.3.3. Travail écrit

A. Définitions opérationnelles

Un travail écrit est une activité de rédaction dans une perspective de réflexion, de création, d'élaboration, ou de production ; et se réalise individuellement ou en équipe, dans des conditions de temps ou d'espaces gérées par les apprenant(e)s, excepté le moment du dépôt (De Ketele, 1993).

B. Exemples de méthodes

Le travail écrit est utilisé lors de la revue de littérature, du résumé, du rapport de stages, du portfolio, d'un article, d'une dissertation, d'une analyse critique, d'une synthèse, d'un journal de bord, d'une lettre d'opinion, d'un essai, d'un mémoire, d'une thèse, etc. (Berthiaume, David et David, 2010).

C. Avantages

Le travail écrit est rapide à construire ; permet l'autonomie chez les étudiant(e)s ; et mesure des tâches de niveaux taxonomiques complexes (organisation) (Howe et Ménard, 1994).

D. Limites

Le travail écrit nécessite un temps long de correction ; est sujet à l'interprétation subjective des réponses ; nécessite la construction d'une grille d'évaluation critériée ; est sujet aux biais de correction (effet de halo, position du travail dans la pile, effet Pygmalion) ; et est une tâche de rédaction plus complexe à réaliser pour une personne allophone ou en situation de handicap (Papi, Gérin-Lajoie et Hébert, 2020).

2.3.4. Situation authentique

A. Définitions opérationnelles

Une situation authentique est une activité complexe exposant l'étudiant(e) à des attentes professionnelles préparatoires ou terminales dans laquelle il ou elle doit démontrer la mobilisation de ses ressources. Les performances sont offertes en contexte réel ou en contexte stimulé (Kozanitis, 2021).

B. Exemples de méthodes

Une situation authentique est évaluée lors des stages, des examens cliniques, des approches par projet ou par problème, des études de cas, de créations d'une oeuvre, projet Web, des entrevues, des simulations, des expérimentations, des démonstrations, des performances, des journaux de bord, de recherches, etc. (Guérin-Lajoie, Hébert et Papi, 2024).

C. Avantages

Une situation authentique permet à l'étudiant(e) de démontrer ses savoirs, savoir-faire et savoir-être ; mesure des apprentissages de niveaux taxonomiques complexes ; encourage le jugement et l'innovation ; simule le contexte d'une situation de travail ; favorise la consultation, la rétroaction et l'amélioration en cours de projet ; la motivation et l'engagement des étudiant(e)s ; et l'intégrité académique (Meunier et Michaud, 2017).

D. Limites

Une situation authentique est longue à construire et longue à corriger ; nécessite des consignes claires et précises ; des ressources matérielles, humaines et financières ; demande un encadrement individualisé ; est un projet d'ampleur pour les étudiant(e)s ; nécessite la construction d'une grille d'évaluation rigoureuse et basée sur les objectifs d'apprentissage et une interprétation subjective est possible de la performance par le milieu (intervenant(e)s, patient(e)s, client(e)s, collaborateur (trice) s (Côté, 2017).

2.3.5. Activité collaborative

A. Définitions opérationnelles

Une activité collaborative est une activité d'interactions et d'interdépendance mutuelle visant un apprentissage ou une production collective ; et peut aussi concerner une démarche d'appréciation (Jaillet, 2005).

B. Exemples de méthodes

Les activités collaboratives sont organisées lors des séminaires, des plénières, des groupes d'experts, des discussions, des jeux de rôle, des débats, des projets de recherche, etc. Leurs outils d'évaluation : des portfolios d'équipe, des co-évaluations, des évaluations par les pairs, etc. (Baudrit, 2007).

C. Avantages

Les activités collaboratives suscitent l'engagement des étudiant(e)s par la collaboration ; favorisent l'interdépendance positive ; encouragent le jugement et l'innovation ; permettent de mesurer des compétences transversales (collaboration, communication, jugement critique) ; et favorisent l'intégrité académique (Pépin, 2019).

D. Limites

Les activités collaboratives nécessitent une planification structurée de l'activité ; parfois la gestion de conflits ; sont sujettes aux biais de correction, notamment lorsque les étudiant(e)s évaluent leurs pairs ; à l'autocongratulation lorsque les étudiant(e)s s'évaluent eux-mêmes ; et rendent difficile la distinction de la participation individuelle dans un projet collaboratif (Holman, 2005).

2.3.6. Communication

A. Définitions opérationnelles

La communication est une activité de communication individuelle ou en équipe qui permet de rendre publique une production verbale ou écrite témoignant des apprentissages. Elle peut se réaliser avec ou sans obligation d'interaction (Lefebvre et Deaudelin, 2001).

B. Exemples de méthodes

L'évaluation de la communication est possible lors des exposés à oral, de la présentation par affiche, du forum, du montage multimédia, d'un reportage photo, d'une entrevue, etc. (Nolin, 2013).

C. Avantages

L'évaluation de la communication est rapide à construire ; favorise l'intégrité académique ; mesure des apprentissages de niveaux taxonomiques complexes ; son temps de correction est peu élevé : s'évalue pendant la prestation ; et permet d'évaluer (selon les objectifs du cours) la qualité de l'expression orale ou écrite de l'étudiant(e) (Dionne, 2005).

D. Limites

L'évaluation de la communication occasionne une possible interprétation subjective de la prestation, nécessite la construction d'une grille d'évaluation critériée ; est sujette aux biais de correction (ex : biais de Halo, influence de la forme sur l'évaluation du contenu) ; à l'autocongratulation (autoévaluation ou évaluation par les pairs) ; nécessite l'apprentissage ou l'utilisation d'outils numériques parfois moins familiers ; exige la communication en direct: engendre un stress supplémentaire pour certaines personnes (Martin, Lefrançois, Guichard, Tapp et Arsenault, 2016).

Le chapitre 3 décrit les questions fermées et les questions ouvertes.

Chapitre 3. Description des types de questions d'évaluation des apprentissages

Objectif général

Le chapitre 3 vise à rendre les étudiants aptes à décrire les types de questions d'évaluation des apprentissages.

Objectifs spécifiques

A la fin du chapitre 3, les étudiants seront capables de :

-décrire les questions fermées ;

-décrire les questions ouvertes.

Nous disposons d'un grand choix de types de question que nous pouvons utiliser dans les évaluations. Nous pouvons distinguer : les « questions ouvertes » où l'apprenant répond avec ses propres mots et les « questions fermées » où la (ou les) réponse(s) se trouve(nt) parmi une liste de choix proposée à l'apprenant (Tanana, 2011).

Les questions fermées sont plus faciles à implémenter et sont corrigées d'une manière objective, elles sont surtout utilisées dans des situations où l'on veut tester les connaissances de l'apprenant sur un sujet particulier. Par contre, comme elles sont créées avec des « mots » de l'enseignant, il peut y avoir des problèmes d'interprétation de la part de l'apprenant (Kabore, 2023).

Les questions ouvertes servent plus à mettre l'apprenant dans une situation où il pourra s'exprimer avec son propre style, faire preuve de prise de décision personnelle, de créativité et de communication. Elles sont beaucoup plus subjectives et donc plus difficiles à corriger (Potvin, 2010).

Par la suite, nous présentons une description des différents types de question utilisés dans les évaluations.

3.1. Questions fermées

L'inconvénient de ce type de questions est qu'elles ne permettent à l'apprenant, ni de s'exprimer ni de rédiger avec son propre texte. Nous distinguons :

3.1.1. Les QCM (Question à Choix Multiples)

Un QCM est composé d'un énoncé et de plusieurs réponses possibles. C'est la forme d'exercices la plus simple. L'apprenant répond en choisissant (cochant) une ou plusieurs réponses (Bouvy et Warnier, 2016).

Il existe plusieurs variantes de QCM : vrai/faux, trouver la seule bonne réponse, trouver la meilleure réponse, trouver la seule réponse fausse, trouver toutes les bonnes réponses, etc. (Lepage et Romainville, 2009). Leur correction est simple et rapide, elle peut être effectuée de façon automatique.

Les QCM ont été intensivement utilisés dans les évaluations pédagogiques car ils sont faciles à implémenter et permettent un traitement rapide des réponses. Par ailleurs, il leur est souvent reproché d'être surtout adapté à l'évaluation des connaissances, d'inciter au « bachotage » et de ne pas solliciter les capacités d'expression des apprenants (Lieury, 1996).

3.1.2. Question de type association ou appariement

Ces questions proposent à l'apprenant deux listes de choix où il devra faire correspondre ou associer des éléments de la première liste avec un ou plusieurs éléments de la deuxième liste. Nous pouvons avoir des associations graphiques à réaliser, comme par exemple déplacer un élément de la première liste dans un emplacement qui correspondrait à un élément de la deuxième liste, ou relier par un trait graphique un objet de la première liste avec un autre objet de la deuxième liste, etc. Nous pouvons aussi avoir des associations textuelles, les deux listes étant référencées par des indices alphanumériques et il suffit d'avoir une troisième liste avec les associations possibles. La complexité de la question est augmentée si l'appariement entre les deux listes n'est pas bijective, c'est-à-dire, si un élément de la première liste peut être associée à plusieurs éléments de la seconde, et réciproquement (Charlin, Bordage et Von Der Vleuten, 2003).

3.1.3. Question de type réarrangement

Ces questions proposent à l'apprenant de mettre dans un certain ordre une liste d'éléments qui font partie de l'énoncé. Celui-ci contient des propositions dans un ordre quelconque. L'apprenant doit alors les « réarranger » dans le bon ordre (Loye et Fontaine, 2018).

3.1.4. Questions semi-ouvertes

Ce type de questions est un compromis entre le type précédent et le type suivant, il ne permet pas à l'apprenant de s'exprimer en son propre langage, mais, celui-ci a la possibilité de répondre de

plusieurs façons différentes (Beauset, Seha et Duroisin, 2024). La réponse à ce genre de question doit être brève et précise, et pour laquelle on donne des orientations. Nous distinguons :

A. Question à réponse courte

L'apprenant répond à la question par un ou plusieurs mots. L'enseignant pose des questions claires et précises qui imposent des réponses brèves et spécifiques. Généralement, les réponses correspondent à des mots ou de courtes expressions (Gerard, 2006).

B. Question de type « carte conceptuelle »

Les cartes conceptuelles sont des graphes qui permettent de représenter les concepts d'un domaine de connaissances donné et les liens entre ces concepts. Elles ont été introduites à l'origine par Novak et Gowin (1984).

Ce formalisme a été utilisé par Delorme (2005) pour le développement d'un outil générique pour l'évaluation des connaissances de l'apprenant pour les hypermédias adaptatifs. Cet outil permet de générer un rapport complet sur la production de l'apprenant, permettant ainsi de savoir si l'apprenant a compris ou non une notion du cours et éventuellement les difficultés rencontrées afin d'y remédier (Gérin-Lajoie, Hébert et Papi, 2024).

Delorme (2005) définit 2 types de construction des cartes conceptuelles par les apprenants :

- La construction complète d'une carte conceptuelle : l'apprenant devra choisir les concepts et les liens qui doivent y figurer, ainsi que la structure de la carte ;
 - La construction partielle d'une carte conceptuelle, dont la structure, les concepts et les liens ont été fixés. L'apprenant doit simplement placer les concepts et/ou liens au bon endroit sur la carte.
- L'enseignant commence par modéliser les notions qu'il souhaite aborder à l'aide d'une carte conceptuelle. La carte de l'enseignant est utilisée comme référence pour évaluer les cartes des apprenants. L'évaluation est basée sur les différences existant entre les deux cartes.

D'autres travaux de recherche ont aussi utilisé les cartes conceptuelles pour représenter les conceptions d'un apprenant, nous pouvons citer entre autres Tribollet, Langlois et Jacquet (2000) et Pudenko, Basque et Legros (2003).

3.2. Questions ouvertes

Les questions ouvertes sont plus compliquées à mettre en oeuvre et à corriger automatiquement, sachant que c'est l'apprenant qui répond librement à la question : en tenant compte des données du problème à résoudre ; en respectant la méthode de résolution ; en interagissant avec le système pour modifier les paramètres du problème ; etc. (Fagnant et Goffin, 2017).

Le principal avantage de ce type de question est la possibilité d'évaluer des objectifs pédagogiques de niveau cognitif élevé. Il invite les apprenants à s'exprimer avec leur propre langage et leur style. Par contre, la correction de telles questions est subjective et plus longue à réaliser. Pour le bon déroulement de ce type d'évaluation, il faut respecter certaines règles ou procédures pour l'élaboration de telles questions. L'énoncé doit être clair et précis, et contenir toutes les informations pour la résolution du problème (Ouedraogo, 2023).

Le chapitre 4 indique les principes d'élaboration des outils d'évaluation des apprentissages: la planification, la collecte des données, l'interprétation, le jugement et la décision, la communication, l'évaluation du processus et la rétroaction.

Chapitre 4. Les principes d'élaboration des outils d'évaluation des apprentissages

Objectif général

Le chapitre 4 vise à rendre les étudiants aptes à indiquer les principes d'élaboration des outils d'évaluation des apprentissages.

Objectifs spécifiques

A la fin du chapitre 4, l'étudiant sera capable de :

- indiquer comment se fait la planification des outils d'évaluation des apprentissages ;
- indiquer comment préparer la collecte des données d'évaluation des apprentissages ;
- indiquer la nécessité de l'interprétation des données d'évaluation des apprentissages ;
- indiquer en quoi consiste l'étape de jugement des données d'évaluation des apprentissages,
- indiquer comment se fait la communication des résultats d'évaluation des apprentissages ;
- indiquer comment se fait l'évaluation du processus d'évaluation des apprentissages ;
- indiquer comment se fait la rétroaction des résultats d'évaluation des apprentissages.

4.1. La planification

La planification de l'évaluation va de pair avec la planification des situations d'enseignement-apprentissage. L'évaluation doit effectivement être en concordance avec ce que les étudiants font et apprennent (Allal, 2008 ; Prégent, Bernard et Kozantis, 2009 ; Dupuis et Dubois, 2024).

4.2. La collecte des données

La collecte de données qui vont servir à l'évaluation doit être préparée par le choix de la méthode ou de la stratégie d'évaluation appropriée. La collecte à proprement parler consiste à faire faire le travail par les étudiants. Les méthodes d'évaluation choisies doivent être variées, valides (elles vérifient ce qu'elles souhaitent vérifier et permettent aux étudiants de démontrer leurs compétences) et fiables

(donneraient les mêmes résultats si les données étaient collectées par quelqu'un d'autre, à un autre moment) (Raudonyté et Foimapafisi, 2021).

4.3. L'interprétation

L'interprétation permet de donner un sens aux données recueillies afin de porter un jugement de qualité et de prendre des décisions appropriées. Les grilles d'évaluation déterminent clairement les critères et les indicateurs des différents niveaux de respect de ces critères. Elles facilitent l'interprétation des éléments d'évaluation (Jouquan, 2002).

4.4. Jugement et décision

L'étape du jugement et de la décision consiste à mettre un résultat sur une production et sur l'ensemble du trimestre. Le jugement doit être accompagné de commentaires qui facilitent sa compréhension (Hayem, 2024).

4.5. Communication

La communication des résultats à l'établissement est aussi accompagnée par la remise des résultats aux étudiants eux-mêmes (Talbot et Raîche, 2017).

4.6. Evaluation du processus

Après le processus d'évaluation, il convient de se questionner quant aux outils d'évaluation et d'interprétation des résultats et aux résultats obtenus par les étudiants. On peut, notamment, se poser les questions suivantes : La stratégie globale d'évaluation a-t-elle permis d'obtenir un portrait des étudiants par rapport à tous les objectifs ou compétences à évaluer? L'évaluation vérifiait-elle ce qu'elle souhaitait vérifier? La qualité des travaux correspondait-elle à ce qui était anticipé? Les grilles d'évaluation étaient-elles efficaces? Les étudiants ont-ils pu s'approprier clairement tout le processus? Les éléments d'évaluation ont-ils favorisé la réussite? Etc. (Forgette-Giroux, Simon et Bercier-Larivière, 1996).

4.7. La rétroaction

La rétroaction est au centre pour éviter de la considérer uniquement dans l'étape de la communication des résultats. Les rétroactions doivent être présentes sous la forme d'un dialogue tout au long du processus. Elles proviennent des étudiants et de l'enseignant (Nduwingoma, Ntwari et Ntahonkiriye, 2020).

Le chapitre 5 présente les principes à la base de l'évaluation des compétences.

Chapitre 5. Les principes de l'évaluation des compétences

Objectif général

Le chapitre 5 vise à rendre les étudiants aptes à indiquer les principes de l'évaluation des compétences.

Objectifs spécifiques

A la fin du chapitre 5, l'étudiant sera capable de :

- rendre compte d'une progression dans une trajectoire de développement ;
- accorder la priorité aux compétences et non aux ressources ;
- déterminer les ressources mobilisables et mobilisées, combinables et combinées ;
- circonscrire les situations de déploiement des compétences ;
- documenter la trajectoire de développement à partir des preuves partagées ;
- rendre compte de l'autonomisation dans le déploiement des compétences ;
- recourir à des critères multiples dans chaque évaluation ;
- intégrer les différences individuelles dans l'appréciation.

5.1. Principe 1 : Rendre compte d'une progression dans une trajectoire de développement

Le développement des compétences qui se déroule sur plusieurs années nécessite de mettre en oeuvre des situations d'apprentissage qui provoquent le développement de compétences ; de mettre en oeuvre des situations d'évaluation en cohérence avec les situations d'apprentissage ; et de rendre compte à l'étudiant de ses progrès, de son évolution et de ses apprentissages relativement au développement de chaque compétence (Tardif, 2006 ; Fontaine et Loye, 2017).

Le développement des compétences ne permet pas de proposer des évaluations sporadiques ; et de produire une information sur un état donné. Il permet de situer les évaluations sur une trajectoire de développement d'apprentissage relatif à la compétence visée. Concrètement, il s'agit de fournir une information rigoureuse sur le degré de maîtrise atteint sur la trajectoire de développement de la compétence visée ; les progrès accomplis ; et les apprentissages à privilégier (Malo et Monfette, 2023).

Ceci impose aux enseignants de maîtriser : le modèle cognitif qui préside aux choix des situations d'apprentissage et d'évaluation en déterminant les étapes de développement d'une compétence ; les obstacles au développement d'une compétence; en différenciant des trajectoires de développement ; et en déterminant des étapes de développement mutuellement exclusives (Bouchemaa, 2022). Ceci exige aux enseignants la maîtrise des apprentissages à réaliser dans le développement de ces compétences.

5.2. Principes 2 : Accorder la priorité aux compétences et non aux ressources

L'enseignant se préoccupe de la capacité de l'élève ou de l'étudiant de mobiliser et de combiner efficacement les ressources dans un contexte précis (Perrenoud, 1998). Il faut éviter d'évaluer les ressources comme entités séparées des compétences elles-mêmes au risque de mettre l'accent uniquement sur les ressources cognitives.

Ce qui a trait aux ressources doit être enchâssé d'une manière synergique dans les compétences. Une attention particulière doit être accordée à l'efficacité des mobilisations et des combinaisons des ressources (Talbot, 2015).

5.3. Principes 3 et 4 : Déterminer les ressources mobilisables et mobilisées, combinables et combinées

L'objectif est de mettre en place les stratégies et les moyens appropriés pour être en mesure de rendre compte : des ressources mobilisées et combinées par l'apprenant ; des ressources dont dispose l'apprenant et qu'il n'a pas mobilisées, ni combinées dans ce contexte - ce sont les ressources mobilisables et combinables (Poumay et Georges, 2010) ; et des ressources qui auraient dû être mobilisées dans ces contextes et qui ont été ignorées (Michaud et Roy, 2022). Soit parce qu'elles ne font pas encore partie de son bagage personnel ; soit par méconnaissance, parce qu'il ne les maîtrise pas à un degré suffisant ; et soit par représentation erronée ou lacunaire de la tâche à accomplir (Daigle, 2013), en vue de l'efficacité des combinaisons.

Les stages constituent l'occasion d'évaluation des ressources mobilisées et combinées par l'élève. La pratique en stage nécessite le recours à de nombreuses ressources de l'ordre : de la connaissance (psychologie, pédagogie, didactique) ; des relations interpersonnelles ; de l'éthique (Tardif, 2006, p. 109).

La carte conceptuelle est le moment propice à l'enseignant de demander à l'apprenant de produire des Figures par rapport aux ressources mobilisées et combinées par les compétences (Ouellet, 2008).

L'intérêt de la mise en exergue des ressources mobilisées et combinées permet d'évaluer l'attention accordée aux ressources qui n'ont pas été mobilisées et qui auraient dû l'être permet à l'enseignant de sélectionner les apprentissages qu'il faut réaliser par la suite sur le plan des ressources (Pronovost, 2018).

La carte conceptuelle permet de demander à l'apprenant de dévoiler l'ensemble des ressources, internes et externes, qu'il aurait pu éventuellement considérer dans une situation de même nature (Chaumont, 2015). L'attention accordée aux ressources mobilisables et combinables permet à l'enseignant d'estimer systématiquement le degré de transférabilité des ressources d'un étudiant face à diverses problématiques (Martin, 2012).

5.4. Principe 5 : Circonscrire les situations de déploiement des compétences

Les évaluations doivent rendre compte aux apprenants de l'évolution de leurs compétences relativement à la situation, au moyen des situations d'apprentissage allant des problématiques simples aux plus complexes ; et des situations d'évaluation en cohérence avec les situations d'apprentissage allant dans ce cas des problématiques les plus simples aux plus complexes (Poumay et Georges, 2010).

5.5. Principe 6 : Documenter la trajectoire de développement à partir de preuves partagées

Plus on privilégie le recours à des situations complexes, plus on augmente l'importance de documenter rigoureusement à partir de preuves les conclusions évaluatives (Tardif, 2003).

- Souhait : les enseignants souhaiteraient que les preuves sur lesquelles reposent leurs conclusions évaluatives soient irréfutables et que, par conséquent, leur jugement ne puisse pas être mis en doute (Tardif, 2006). Ceci est rendu possible grâce au travail collégial des enseignants. La collégialité assure un plus haut degré de validité des preuves à partir desquelles on infère un niveau de développement de la compétence ou un niveau de maîtrise des ressources mobilisables et combinables (St-Onge, 2018). Pour favoriser la validité la plus élevée possible, plus d'un enseignant apprécie la production d'un élève ou d'un étudiant.

5.6. Principe 7 : Rendre compte de l'autonomisation dans le déploiement des compétences

Le niveau de compétence est fonction du niveau d'autonomie de l'apprenant, de sa capacité plus ou moins grande à mobiliser et combiner les ressources internes et externes (Grimand, 2007).

Les moyens pour développer l'autonomie de l'apprenant sont les suivants :

- Développer ses stratégies cognitives, ses stratégies d'apprentissage, apprendre à y recourir de plus en plus fréquemment et de façon de plus en plus appropriée ;

- Développer ses stratégies métacognitives, se donner les moyens de se réguler ;

- Développer ses stratégies coopératives sans quoi l'apprenant risque de devoir abandonner la réalisation de nombreuses tâches complexes qui nécessairement ne peuvent être traitées sans la constitution d'équipes coopératives (Poumay et Georges, 2010).

5.7. Principe 8 : Recourir à des critères multiples dans chaque évaluation

Les critères d'évaluation de la compétence portent sur : le développement de la compétence elle-même ; les ressources cognitives ; les ressources relationnelles, motrices et sociales ; et l'autonomisation progressive dans le déploiement de la compétence (Nguyen et Blais, 2007).

5.8. Principe 9 : Intégrer les différences individuelles dans l'appréciation

Il est crucial que les critères retenus pour apprécier le développement de la compétence ne contraignent pas les apprenants à devenir des copies conformes les uns par rapport aux autres (Poumay et Georges, 2010). Il importe de se référer au modèle cognitif qui préside à l'acquisition de la compétence visée. Ce modèle a entre autres l'avantage de : déterminer les étapes de développement d'une compétence ; et de différencier des trajectoires de développement.

Le chapitre 6 indique l'historique de la problématique des difficultés liées à l'évaluation des apprentissages, les sources d'erreurs et les exemples de biais mis en évidence par la docimologie critique.

Chapitre 6. Les difficultés liées à l'évaluation des apprentissages

Objectif général

Le chapitre 6 vise à rendre les étudiants aptes à attirer l'attention sur les difficultés liées à l'évaluation des apprentissages.

Objectifs spécifiques

A la fin du chapitre 6, l'étudiant sera capable de :

- indiquer l'historique de la problématique des difficultés liées à l'évaluation des apprentissages ;
- indiquer les sources d'erreurs de l'évaluation des apprentissages ;
- indiquer les exemples de biais mis en évidence par la docimologie critique.

6.1. L'histoire de la problématique

L'école est un lieu dans lequel l'évaluation est omniprésente. Il semble même parfois à certains que l'élève fréquente davantage l'école pour récolter des notes que pour apprendre véritablement quelque chose (Gerard, 2013).

Avec la massification et la démocratisation de l'accès à tous les niveaux scolaires, il faut pouvoir comptabiliser échecs et succès à travers un système de notation de façon à rendre un « verdict » en fin d'année. La notation est ainsi une réponse à la division du travail et à l'unicité de l'élève et du temps: il doit être possible de prendre, à un moment donné, une décision relative à chaque élève, ce qui implique la prise en compte d'informations provenant de sources multiples (Fontaine et Loye, 2017).

Si donc l'école a vu se systématiser et se professionnaliser « l'art de bien enseigner » à travers la didactique, elle a connu le même développement de « l'art de bien évaluer » à travers la docimologie. Dans la suite du texte, nous envisagerons principalement les travers de l'évaluation classique des élèves, nous en pointerons les limites, en suivant les chemins de la docimologie critique. Le présent chapitre ne constitue qu'une introduction critique à la docimologie pratique, la problématique étant d'être conscient et attentif aux difficultés rencontrées lors de l'évaluation (Pilet, Allard et Horoks, 2019).

Les développements de la docimologie et de la mise en cause des notes scolaires remontent au début du vingtième siècle. Dès les années 1910, les Etats-Unis firent confiance aux QCM dans

les tests de sélection, par souci d'objectivité et en réponse à la difficulté de noter. A partir de 1922, en France, Henri Piéron s'attaque aux problèmes posés par la subjectivité de la note. Dès 1929, il attire l'attention en ces termes : « *C'est un principe général que, pour être reçu à un examen, il faut avoir la moyenne, [...] dès lors, [...] pour un grand nombre de candidats, ce sera [...] le hasard qui décidera de leur admission ou de leur recalage. En effet, on sait que [...] c'est dans la région moyenne qu'ils se massent [...]* » (Piéron, 1963, p. 9).

Aux USA, en Angleterre, et même en Belgique, diverses expériences mettent en évidence le manque de fiabilité des notes scolaires (Leclerc, Nicaise et Demeuse, 2004). Piéron (1963, p. 13). Ils citent notamment les travaux menés à ce sujet, en 1931, par Andréa Jadoulle, la célèbre psychopédagogue du Laboratoire de pédagogie d'Angleur. En France, Laugier et Weinberg étudient ce même phénomène dès 1927.

C'est en 1931 qu'une impulsion déterminante sera donnée par la subside, par la Carnegie Corporation de New York, d'une recherche pilotée par l'International Institute of Education du Teacher's College de l'Université Columbia, et fonctionnant via des commissions nationales : américaine, anglaise, écossaise, finlandaise, française, suisse et allemande (cette dernière étant arrêtée par la nazification de 1933).

La commission française utilisa des copies notées lors du fameux baccalauréat. En 1934 fut publié, par la commission française, le rapport « Etudes docimologiques » (Laugier, Piéron, Pieron, Toulouse et Weinberg, 1934). Le baccalauréat offrait une situation exceptionnelle puisque les mêmes questions sont posées à de très nombreux étudiants, durant de véritables examens, et sont collectées et corrigées par de nombreux correcteurs sélectionnés. A travers cette situation, complètement externe, l'équipe française met en évidence de nombreux biais de notation. On se doute que la guerre interrompit ce processus de recherche sur le continent européen (Suchaut, 2008).

6.1.1. Les examens internes

En Belgique, quel que soit le niveau d'enseignement considéré, les examens sont généralement administrés par les enseignants qui ont donné les cours. C'est eux qui créent les questions et qui corrigent les copies. Cette façon de faire a des avantages comme celui de « coller » à la matière qui a effectivement été enseignée, ou celui d'une familiarité des élèves au type de questions. Il a le désavantage de laisser libre cours aux différences (de sévérité par exemple) inter correcteurs ou inter établissements, ce qui pose le problème éthique de l'équité et de l'égalité de traitement,

tout spécialement quand l'examen est « sanctionnant » et quand le professeur sait de qui il corrige la copie (Lopez et Laveault, 2008).

6.1.2. Les examens externes à correction subjective

Dans un souci d'égalité de traitement, la France, depuis Napoléon, pratique le baccalauréat, examen (le même pour tous les étudiants d'une même « Académie ») conçu et corrigé par des enseignants n'ayant pas participé à l'encadrement des candidats évalués dont les copies sont rendues anonymes. On devine les précautions à prendre par les formateurs pour respecter « le programme » et par les concepteurs des questions pour « éviter les fuites ». Ces examens restent toutefois subjectifs quant à la correction (Dierendonck, Burton et Wanlin, 2009).

6.1.3. Les examens externes à correction objective

Poussant encore plus loin le souci d'« equity » et de « unbiased evaluation », les Américains ont non seulement conçu, à la charnière du secondaire et de l'enseignement supérieur, des examens (par exemple le Scholastic Aptitude Test ou SAT) qui sont les mêmes pour tous, mais dont la correction est objective (ce sont des QCM) (Lopez et Laveault, 2008).

D'où l'expression « objective tests », par un élargissement de sens légèrement abusif. Le terme « docimologie » apparaît quant à lui en 1929 sous la plume d'Henry Piéron et est popularisé par celui-ci dans son ouvrage intitulé « Examens et docimologie », en 1963. Ce mot puise sa racine dans le grec (examiner, épreuve). A ses débuts, la docimologie est surtout critique ou négative: elle met en évidence les problèmes, sans les résoudre, du moins de manière pratique, au niveau où le problème se pose, c'est-à-dire au niveau des enseignants chargés de procéder à l'évaluation. Progressivement cependant, les chercheurs s'attachent à proposer des solutions qui permettent de limiter au mieux le caractère subjectif de la notation (Yerly, 2017).

Comme nous allons le voir par la suite, ce caractère subjectif n'est pas à imputer uniquement au maître chargé de noter l'élève, il relève de mécanismes souvent complexes et qui mettent en jeu enseignants, élèves et système éducatif (Dierendonck, Burton et Wanlin, 2009).

Les méthodes employées pour étudier les biais de notation reposent sur différentes procédures, selon le type de biais à mettre en évidence. On peut ainsi utiliser, de manière expérimentale, les approches suivantes:

(a) Une même série de copies est corrigée plusieurs fois par le même correcteur, à des moments différents, sans que celui-ci s'en rende compte, ce qui permet de mesurer la stabilité intra-correcteurs;

- (b) Une même série de copies est corrigée par plusieurs correcteurs différents, ce qui permet de mesurer la concordance inter-correcteurs;
- (c) Une même copie est placée dans un ensemble de copies dans des positions différentes (précédée de copies meilleures ou plus faibles), ce qui permet de mesurer l'effet de contraste, ou de séquence;
- (d) Une même copie est placée dans un ensemble de copies dont les valeurs sont plus ou moins dispersées largement (tantôt parmi des copies ayant toutes reçu la même note lors d'une évaluation préalable, tantôt parmi des copies très variées en qualité); etc.
- (e) Une même copie est corrigée par plusieurs groupes de correcteurs auxquels on fournit des informations complémentaires différentes sur l'élève, ses notes antérieures, etc. (Tessaro, 2015).

6.2. Les trois sources « d'erreur »

Si l'on cherche à identifier les sources d'erreurs ou de biais qui entachent la notation par les enseignants, pour conserver une approche comparable à celle adoptée par la théorie classique des tests, les chercheurs identifient au moins trois sources principales : les enseignants, le système scolaire et les élèves. Nous allons aborder successivement ces trois sources. Nous montrerons combien il importe d'envisager un problème qui semble assez singulier, donner une note à un élève particulier, à travers un ensemble beaucoup plus large de déterminants et d'interactions (Perrenoud, 1998).

6.2.1. Le système scolaire

La première source de biais que l'on doit prendre en considération est l'influence du contexte scolaire sur les procédures d'évaluation en général. Ainsi, la classe dans laquelle se trouve l'élève peut être déterminante. Les conclusions tirées notamment par Grisay sur « l'effet Posthumus » au début des années quatre-vingt – et de nombreuses fois vérifiées et commentées par la suite (Grisay, 1984 ; Crahay, 1996 ; Demeuse, 2002) – offrent un très large aperçu sur les interférences que peut avoir le contexte d'une « classe particulière » sur l'évaluation des élèves qui la composent : avec les mêmes performances, et toute autre chose restant égale par ailleurs, un élève est jugé par son professeur comme un « bon élève » dans une classe alors qu'il peut se voir contraint de doubler son année scolaire dans une autre (Matracci, 2020)!

Tout dépend non pas des performances particulières de l'élève dans l'absolu mais bien de ses performances par rapport à celles de ses condisciples. Le hasard du microcosme de la classe dans laquelle se trouve les élèves est donc primordiale puisque le professeur est fréquemment poussé à

établir des différences de performance entre eux, et ce même si ceux-ci sont très proches, c'est-à-dire à adopter une attitude plus normative que critériée. De ce fait, leurs résultats en fin d'année sont souvent distribués selon une courbe de Gauss (quelques élèves « faibles », quelques élèves « forts », et la grande majorité dans la « moyenne ») (Perrenoud, 1995; Crahay, 1996 ; Merle, 1998).

Cependant si la classe dans laquelle est placé un élève plutôt que dans une autre et son influence sur l'évaluation finale est importante, l'établissement scolaire l'est parfois tout autant. Alors que de nombreux travaux – notamment américains – et leurs diverses interprétations avaient pu laisser sous-entendre dans la foulée des Golden sixties, que l'école a peu d'impact sur les résultats des élèves, le « school can make a difference » est actuellement reconnu, aussi bien au niveau des résultats effectifs des élèves que de l'évaluation de ceux-ci (Chanudet, 2019).

Il semble toutefois que l'effet-classe que nous venons d'évoquer soit bien plus important que l'effet-établissement (Bressoux, 1994, 1995), même si l'influence de ce dernier est pourtant indéniable. Cette différence entre établissements peut résulter d'une stratégie délibérée destinée à pratiquer une sélection par écrémage et/ou à médiatiser une certaine image de l'excellence dont la « fameuse réputation » de certaines écoles d'élite est la concrétisation (Duru-Bellat et Mingat, 1993 ; Merle, 1998).

Et quand bien même la complexité du contexte de l'école et de la classe ne suffirait pas à laisser entrevoir toute l'incertitude, la non-objectivité dont est déjà empli à ce stade l'acte évaluatif de l'enseignant, des biais bien plus élémentaires peuvent démontrer que la note est encore parfois influencée par d'autres déterminants totalement extérieurs au travail de l'élève en soi : depuis longtemps par exemple la docimologie a pu montrer qu'une même copie est notée différemment par l'enseignant selon son ordre de correction, selon qu'elle se trouve parmi les premières ou parmi les dernières feuilles de la pile que l'enseignant doit corriger (Bonniol, 1965) ; et dans le même ordre d'idée, qu'une même copie sera jugée différemment selon qu'elle suit une copie jugée « très bonne » ou « très faible » par l'enseignant (Bonniol et Piolat, 1971). Bon nombre d'enseignants n'hésitent d'ailleurs pas à déclarer qu'ils établissent leur barème de notation seulement après avoir lu plusieurs copies, et que celui-ci peut être appliqué différemment selon les élèves et selon la succession des résultats, copie après copie (Chevallard, 1991 ; Nicaise, 2001).

6.2.2. Les particularités intrinsèques aux élèves

A côté de ce contexte de scolarisation, des particularités intrinsèques à l'élève peuvent également influencer subjectivement le jugement professoral : dans la foulée des premières sociologies dénonciatrices des années soixante et septante (Bourdieu et Passeron, 1964, 1970 ; Baudelot et Establet, 1971 ; Boudon, 1973), on pensera avant toute autre chose à l'origine sociale des élèves et aux stéréotypies diverses qui peuvent y être associées : ainsi, certains correcteurs ont parfois tendance à attribuer de meilleures notes aux enfants issus des milieux les plus favorisés (Pourtois, Bonacina, Delbecq et Segard, 1978) alors que dans d'autres circonstances, ce sont justement les élèves issus des milieux défavorisés qui sont « surcotés », notamment pour des raisons de « paternalisme bienveillant » et de correction – si minime soit-elle volontaire des inégalités sociales et scolaires (Dardenne, 1999 ; Nicaise, 2001). Le même type de conclusions a également pu être tiré dans ce sens avec des déterminants comme l'apparence physique ou le genre sexuel des élèves : parfois les élèves jugés « plus beaux » ou « plus proches des idéaux médiatiques » par les enseignants peuvent être mieux cotés (Leyens et Yzerbit, 1997 ; Merle, 1998) et, dans le même ordre d'idée, il semble que les filles ont à souffrir plus fréquemment que les garçons quant à leur notation, notamment suite aux phénomènes de « menace stéréotypique » et de reproduction des inégalités sociales en matière de genre sexuel (Dardenne, 1999). Par contre, il peut apparaître également dans d'autres circonstances que les mêmes filles sont habituellement mieux évaluées parce qu'elles font preuve d'un « meilleur comportement » en classe et maîtrisent mieux les règles et exigences du « métier d'élève » – soit donc qu'elles sont plus proches que leurs pendants masculins d'un comportement idéalement attendu par le professeur (Felouzis, 1993; Duru-Bellat, 1995).

Par-là, on peut constater que l'évaluation des enseignants porte tout autant sur les « savoir être » que sur les plus classiques savoirs et savoir-faire (Bourdieu et Passeron, 1970 ; Crahay, 2000), et que le même incitant peut influencer celle-ci dans des sens totalement opposés. Comme le complexe « scolarité », le complexe « élève » est très influent et la relation ne cesse de se complexifier encore un peu plus lorsque entrent en jeu des caractéristiques qui sont simultanément dépendantes des deux : pensons par exemple au niveau scolaire de l'élève (les résultats de ses évaluations précédentes, notamment ceux présentés dans son bulletin scolaire) et au statut de sa classe (la réputation, l'image véhiculée par la classe qu'il fréquente, notamment celle présentée lors des conseils de classe). Par un comportement proche d'un classique

phénomène de réduction de la dissonance cognitive (Festinger, 1957), un enseignant peut parfois être tenté de minimiser la différence qui apparaîtrait entre sa propre attente quant aux résultats de l'élève et les résultats effectivement obtenus. L'habituel « premier de classe » peut se voir ainsi tiré vers le haut alors que les performances réalisées ne correspondent pas à ce que l'enseignant attendait de lui comme à l'accoutumée (Caverni, Fabre et Noizet, 1975, Merle, 1998).

Le complexe « élève » peut être vu comme une mise en commun d'une multitude d'effets d'attente qui se rapportent tous, de près ou de loin, à la thèse devenue classique aujourd'hui de « Pygmalion à l'école » de Rosenthal et Jacobson (Rosenthal et Jacobson, 1969). Selon ce modèle, certains déterminants (origine sociale, sexe, niveau scolaire, etc.) amènent l'enseignant à développer rapidement des attentes diverses vis-à-vis de ses élèves (notamment sur leurs résultats scolaires futurs) et à différencier peu à peu, son propre comportement - verbal comme non-verbal, conscient comme inconscient. Il a tendance à orienter ses élèves vers le résultat scolaire attendu : qu'elle soit positive ou négative, la prédiction peut alors se révéler fortement créatrice d'effets (Good, 1987).

6.2.3. Les enseignants

Comme nous l'avons souligné précédemment, l'existence et l'interférence de nombreux biais dans la procédure évaluative peut encore être exacerbée puisque, dans ce domaine tout du moins, la liberté du maître au sein de sa classe est presque totale : il administre les épreuves comme il l'entend, il choisit la matière et le sujet des interrogations, il choisit leur forme, leur moment, leur durée, leur importance sur la note finale, il choisit les critères et les normes qui détermineront son jugement professoral, etc. puis c'est lui qui applique le modèle d'évaluation aux productions de ses élèves qui s'avèrent, d'une certaine manière, le reflet de son propre travail d'enseignant (Boyer, 2012).

La quête de l'objectivité de la note est donc semée d'embûches majeures. Dépassant par-là la vision mécaniste du maître simple notateur dans l'absolu, il faut appréhender que l'acte évaluatif passe également au travers de nombreux « filtres interactifs » avant que la note finale puisse être arrêtée. Donc, outre les biais classiques que nous avons présentés, le fait de noter un élève est également une action proprement rationnelle qui trouve ses fondements – pour reprendre le raisonnement de Weber (1971) – à la fois dans les intérêts et les valeurs propres de l'enseignant. Ainsi, si le jugement professoral peut être dépendant de l'origine sociale des élèves, de leur âge, de leur sexe ou du type d'établissement fréquenté, il l'est tout autant – comme le souligne très

judicieusement Merle (1996) – d’un ensemble quotidien « d’arrangements » et de « bricolage » des notes, intentionnels ou non. Dès lors, la compréhension des actes évaluatifs nécessite d’une vision ultra-systémique (Perrenoud, 1995, 1998), mais se double de l’exigence d’une approche ultra-individuelle et biographique de chaque sujet évaluateur particulier.

A. Les arrangements internes

Le premier type de « bricolages », de modifications plus ou moins licite des procédures d’évaluation, est destiné directement à la classe in vivo et aux élèves qui la composent. Il peut servir à entretenir un bon climat de travail, à encourager les élèves qui éprouvent des difficultés ou qui ont des problèmes d’ordre extrascolaire (dans ce cas, les notes sont revues « à la hausse »), à restaurer l’autorité concrète ou symbolique du maître en sanctionnant certains comportements (les notes sont alors revues « à la baisse »), à sauvegarder une moyenne de points habituelle, à amener un élève vers une orientation future plutôt qu’une autre, à céder aux éventuelles « pressions » diverses des élèves, etc. Habituellement, ces comportements ne « sortent » pas de la classe, ils ne sont pas délibérément cachés par le maître mais celui-ci s’en vante rarement car ils font partie de sa propre « cuisine interne », de ses procédures personnelles (Merle, 1996). L’apposition d’une note relève donc bien également de la transaction, et constitue un moment particulier – mais essentiel – d’un processus beaucoup plus large, celui d’une véritable « négociation didactique » entre l’enseignant et ses élèves (Chevallard, 1991).

B. Les arrangements externes

Les arrangements dits externes prennent la même forme que les précédents, mais ils sont destinés à la direction de l’école, à l’administration, aux collègues, aux parents d’élèves, bref, à toute personne qui ne participe pas directement au quotidien de la classe, mais qui interagit néanmoins avec elle. Il s’agit souvent pour l’enseignant de présenter une image de sa classe qui satisfasse au mieux ces personnes extérieures : qu’adviendrait-il si trop d’élèves étaient en échec ? Qu’adviendrait-il si tous avaient des résultats exceptionnels ? La réputation et le « statut » prêté à l’enseignant pourrait être mis à mal et il en serait de même pour celui de l’établissement (Morales-Perlaza et Morrissette, 2020).

Ce type d’arrangements est évidemment lié très étroitement au précédent (Merle, 1996) : si par exemple une interrogation écrite est particulièrement mal exécutée par l’ensemble de la classe et que le professeur décide de ne pas en tenir compte, la finalité d’un tel acte est double car celui-ci

a des conséquences internes, mais aussi externes à la classe considérée comme le seul groupe d'élèves (Tessaro, 2015).

Au travers de ces interactions, une évaluation trop sévère apparaît très vite comme injuste, mais une évaluation trop généralement favorable s'apparente à du laxisme et celui-ci nuit inévitablement à la réputation de l'enseignant et de l'établissement (Perrenoud, 1998). Un difficile équilibre doit donc s'établir. Ainsi, l'enseignant tente souvent de ne pas s'écarter de ses moyennes et des distributions habituelles des notes (Grisay, 1984 ; Crahay, 1996). Il montre ainsi à quiconque qu'il « tient » sa classe (Chevallard, 1991). Sur l'importance de ces arrangements externes – soit pour l'essentiel « intéressés » –, Grisay montre que de nombreux enseignants, une minorité il est vrai, avouent prendre en considération des éléments « illégitimes » très diversifiés pour établir leur décision finale – de réussite ou de doublement – pour certains de leurs élèves. Ainsi, les instituteurs peuvent être influencés par les insistances répétées des parents, par le risque que ceux-ci retirent leur enfant de l'établissement si celui-ci venait à doubler, par le fait que la réussite – ou le doublement – d'un élève provoque l'ouverture ou la fermeture d'une classe à la future rentrée scolaire, par le fait que le prochain enseignant de l'élève sera prêt à l'aider et le soutenir plus qu'à l'accoutumée, par le fait également que l'école accorde ou non une grande importance à sa « réputation », etc. (Grisay, 1986, 1991). Il est évident qu'ici encore les enseignants qui usent de ces pratiques ont de grandes réticences à les dévoiler expressément : « faire part de sa « cuisine » évaluative et partager ses doutes nécessitent de dévoiler les limites de son propre jugement et de se mettre en cause professionnellement » (Merle, 1996, p. 86).

C. Les arrangements pour soi

Ce troisième type « d'arrangements évaluatifs » est fréquemment ignoré dans de nombreuses études et ceci principalement à cause de la difficulté de les appréhender et de les regrouper au sein de types-idéaux exploitables puisqu'ils dépendent directement de l'histoire et de la personnalité même du sujet-correcteur. Ils sont pris à l'égard de soi-même et peuvent dépendre d'une foule de représentations personnelles, chacune plus difficilement saisissable que l'autre : « l'idéal pédagogique » de l'enseignant, sa conception générale de l'éducation, son propre parcours scolaire, son origine sociale, ses engagements politiques et associatifs particuliers, etc. Les normes de justice scolaire auxquelles peut adhérer, parfois avec force, l'enseignant sont également déterminantes : l'égalité des chances, l'égalité de traitement ou l'égalité de résultat, le besoin de l'élève, sa contribution et son mérite, le refus de doublement, etc. (Barrère, 2000 ;

Nicaise, 2001). Il en est de même avec les conceptions générales sur les écoliers : ce sont d'abord des élèves à scolariser ou des enfants à éduquer et à socialiser que l'on a en face de soi ? (Dubet et Martuccelli, 1996). Tous ces facteurs, tous ces sens donnés aux pratiques par le sujet selon ses représentations interviennent évidemment de façon conjointe (Baribeau, 2015).

6.2.4. En bref

Ces différentes pratiques montrent donc bien que l'évaluation des élèves, avant d'être une simple apposition d'une note que l'on croit encore parfois être « vraie », « relève de processus et de procédures au croisement des contraintes sociales et des biographies des élèves et des maîtres » (Merle, 1996, p. 306). Ainsi, le processus d'évaluation est dépendant d'un triple rapport entre le professeur et ses élèves, le professeur et ses contraintes externes, et le professeur et son passé, son intériorité, lorsqu'il s'engage personnellement dans son travail.

La procédure évaluative n'est donc pas un acte identiquement posé par chaque enseignant. L'objectivité n'y est pas une norme. Ses règles et ses critères, même les plus généraux, ne sont d'ailleurs que trop rarement définis et arrêtés au préalable, comme ce peut être le cas d'une procédure juridique par exemple.

6.3. Quelques exemples de biais mis en évidence par la docimologie critique

6.3.1. La distribution forcée

On attribue assez généralement à Posthumus, enseignant hollandais en poste en Indonésie durant la seconde guerre mondiale et interné dans un camp japonais durant celle-ci, la paternité d'une loi formulée de la manière suivante par De Landsheere (1992, p. 242): « Un enseignant tend à ajuster le niveau de son enseignement et ses appréciations des performances des élèves de façon à conserver, d'année en année, approximativement la même distribution (gaussienne) de notes ». Cette « loi de Posthumus » indique que la distribution des notes résulte d'une sorte de prototype, communément admis: il existe peu d'élèves exceptionnels (très faibles ou très brillants), mais beaucoup d'élèves relativement moyens. Lorsque l'on se place dans la situation d'une épreuve interne, c'est assez souvent ce type de distribution qui est mise en évidence, alors même que des épreuves externes, appliquées aux mêmes élèves, indiquent des niveaux moyens très variables et des dispersions différentes, d'une classe à l'autre (Leclerc, Nicaise et Demeuse, 2004).

De Landsheere (1992, p. 36), explique ce phénomène de la manière suivante : « Un professeur qui enseigne de façon non individualisée dans une classe où les élèves ne sont pas spécialement sélectionnés donne normalement à son cours un degré de difficulté adapté à la majorité du

groupe. Si l'ajustement est correct, il y aura donc beaucoup de résultats moyens, peu de très bons et peu de très mauvais. La distribution de ces résultats s'approchera de la courbe gaussienne. Cette distribution, dite normale, est à l'image de beaucoup de qualités humaines, telles qu'elles se répartissent dans des groupes nombreux, pris au hasard. » (Dachet, 2022).

Le même auteur dénonce ce qu'il appelle le dangereux mythe de la courbe de Gauss : « Dans les sciences humaines, la courbe en cloche de Gauss joue un rôle considérable, parce qu'elle est l'image même de la répartition de bien des aptitudes et des qualités : les individus moyens abondent, mais les génies et les idiots, les géants et les nains sont rares. Comme les tests mesurent souvent des aptitudes, des traits de personnalité ou des performances de vastes populations, et servent à classer les individus en les comparant les uns aux autres, il est naturel que ces épreuves soient étalonnées selon la répartition gaussienne : en gros, 70 % de moyens, 13 % de bons, 13 % de médiocres, 2 % d'excellents, 2 % de très mauvais. » (Gilles, 2002).

Mais, l'école n'a pas, en principe, pour visée première, la sélection. Il convient donc de s'interroger sur la fatalité de la répartition gaussienne des résultats, d'autant que le traitement réservé à chaque élève dépend, dans ce modèle, de sa position relative initiale dans le groupe d'apprentissage (Crahay, 1996, 2000).

De l'intériorisation de cette distribution forcée découlent deux phénomènes particuliers:

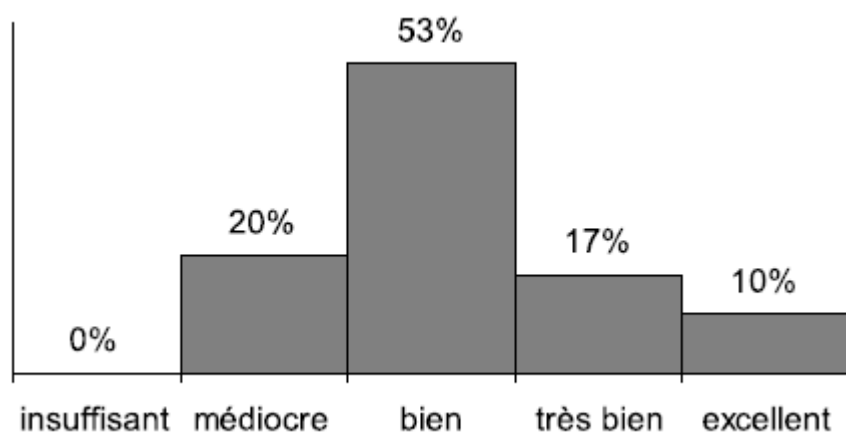
A. L'effet de tendance centrale

On observe fréquemment que les notateurs de performances concentrent leurs appréciations sur les échelons du centre de l'échelle. On peut y trouver deux grands types d'explications. La première est qu'ils ont la courbe de Gauss en tête, et se figurent donc que le plus grand nombre doit se trouver au centre de l'échelle. Certains juges vont même jusqu'à modifier certaines des notes pour que la courbe soit plus « parfaite ». La deuxième explication est la prudence (ou la lâcheté) puisqu'en donnant une note centrale, le correcteur ne peut jamais être aux « antipodes » de la note vraiment méritée par la performance (Detroz et Romainville, 2017).

Rot et Butas (1959) rapportent que Gjorgjevski a invité 5 professeurs d'une même branche de l'enseignement secondaire à noter indépendamment les uns des autres 100 copies de leur discipline sur une échelle à 5 degrés (1 = Insuffisant; 2= Médiocre; 3 = Bien, 4 = Très bien ; 5 = Excellent). Il a ensuite extrait 15 copies qui avaient toutes reçu la note « Bien » par les 5 correcteurs. Elles ont été confiées, pour nouvelle correction, à 4 autres professeurs, qui ont à

nouveau distribué les 15 copies à travers les 5 catégories de notes (De Ketele, 1993), comme l'indique la figure 1 suivante.

Figure 1. Tendance des notateurs à concentrer leurs appréciations aux échelons du centre de l'échelle



B. La surprenante stabilité des taux de réussite et d'échec d'année en année.

Certains enseignants sont fiers d'une telle stabilité, preuve pour eux que « l'ordre des choses » (la courbe de Gauss et un score de passage toujours fixé au même endroit) est « respecté ». Hutmacher (1993), à Genève, a développé une version de cette théorie adaptée à l'enseignement primaire, ce qu'il appelle l'hypothèse socio-arithmétique selon laquelle les maîtres ont dans la tête le nombre « normal » d'échecs (redoublements) par classe : 2 élèves, et font en sorte que ce résultat soit observé. Ce qui débouche sur la conséquence paradoxale que plus la classe est petite (10 élèves par exemple), plus le taux d'échecs est élevé (20% pour 10 élèves, 10% pour 20 élèves) (Pelletier, 2012).

6.3.2. Les biais résultant de l'interaction entre le correcteur et l'étudiant ou la copie évaluée

Dans le cas des évaluations internes, menées par l'enseignant lui-même, un certain nombre d'effets indésirables peuvent résulter de la connaissance que l'enseignant a de l'élève et de l'idée qu'il se fait de ses compétences, a priori.

A. Effet de stéréotype ou d'inertie

Le premier type de « parasitage » de la note peut résulter d'une sorte d'effet d'inertie: le correcteur a tendance à attribuer à un étudiant une note comparable à celles que celui-ci a acquises auparavant. La connaissance des résultats antérieurs d'un élève – même inconnu - tend à influencer l'évaluateur. On assiste à une sorte d'imitation par contagion, etc. Par stéréotypie, on

entend une immuabilité plus ou moins accusée qui s'installe dans le jugement porté sur l'élève, comme le précise De Landsheere (1992).

Caverni, Fabre et Noizet (1975) ont mené l'étude suivante. A des professeurs de sciences de l'enseignement secondaire, ils ont demandé de noter (sur 20) chacun les 4 mêmes copies, accompagnées de « 5 notes censées avoir été obtenues précédemment par l'auteur de la copie ». Chaque série de 5 notes avait deux caractéristiques : sa moyenne (élevée = 13/20 ou faible = 7/20) et sa dispersion, exprimée par la Marge de Variation (MV), c'est-à-dire l'écart entre les notes extrêmes (MV forte = 10 ; MV faible = 2) (De Bosscher, Fournier et Carton, 2017).

De Landsheere (1992) commente : la moyenne exprimait le niveau moyen de l'élève, tandis que la dispersion exprimait la régularité ou l'irrégularité de ses performances. Un autre descripteur aurait (encore) pu être utilisé : la succession des notes peut marquer un progrès (ce qui était le cas ici pour toutes les copies) ou, au contraire, une régression (Rigutto, 2021).

Le tableau 1 ci-dessous indique le résultat obtenu pour chacune des 4 copies (a, b, c et d) dans deux situations particulières : l'information sur les résultats préalables faisait apparaître une moyenne forte (13/20) et une marge de variation faible (2 points), dans le premier cas, et l'inverse (moyenne faible et marge de variation forte), dans le second cas. Comme on peut le constater, la seconde situation est plus défavorable que la première (De Landsheere, 1992, p. 48).

Tableau 1. Effet de stéréotype ou d'inertie

Copies	a	b	C	D	Moyenne (sur les 4 copies)
Moyenne forte	12	8.5	15.25	3	9.69
Marge de variation faible					
Moyenne faible	9.75	6.5	11.75	2.75	7.79
Marge de variation forte					

On aurait tort de croire que la stéréotypie influence uniquement les évaluations à base subjective accusée. Elle atteint des exercices aussi « objectifs » que la dictée orthographique. L'expérience suivante en témoigne. Un professeur de langue maternelle fait régulièrement des dictées. Bientôt, il connaît les élèves qui réussissent habituellement le mieux et le moins bien cet exercice. Si l'on

détermine la fréquence des fautes « oubliées », non perçues par le correcteur, on constate que les oublis en faveur des bons élèves sont significativement plus élevés que pour les élèves faibles. Dans le premier cas, le maître s'attend à ne pas rencontrer d'erreurs ; dans le second, il les guette.

Caverni, Fabre et Noizet (1975) notent qu'il est probable que les premiers indices recueillis, qu'ils soient positifs ou négatifs, vont guider le recueil des indices, l'évaluateur cherchant davantage des indices susceptibles de confirmer ses premières inférences que des indices susceptibles de les remettre en question. De Landsheere (1992), en affirmant qu'il semble que s'il doit faire des fautes, l'élève a intérêt à les faire dans la seconde moitié de son examen. C'est ce qu'une expérience rapportée par Noizet et Caverni (1975) confirme.

B. Effet de halo

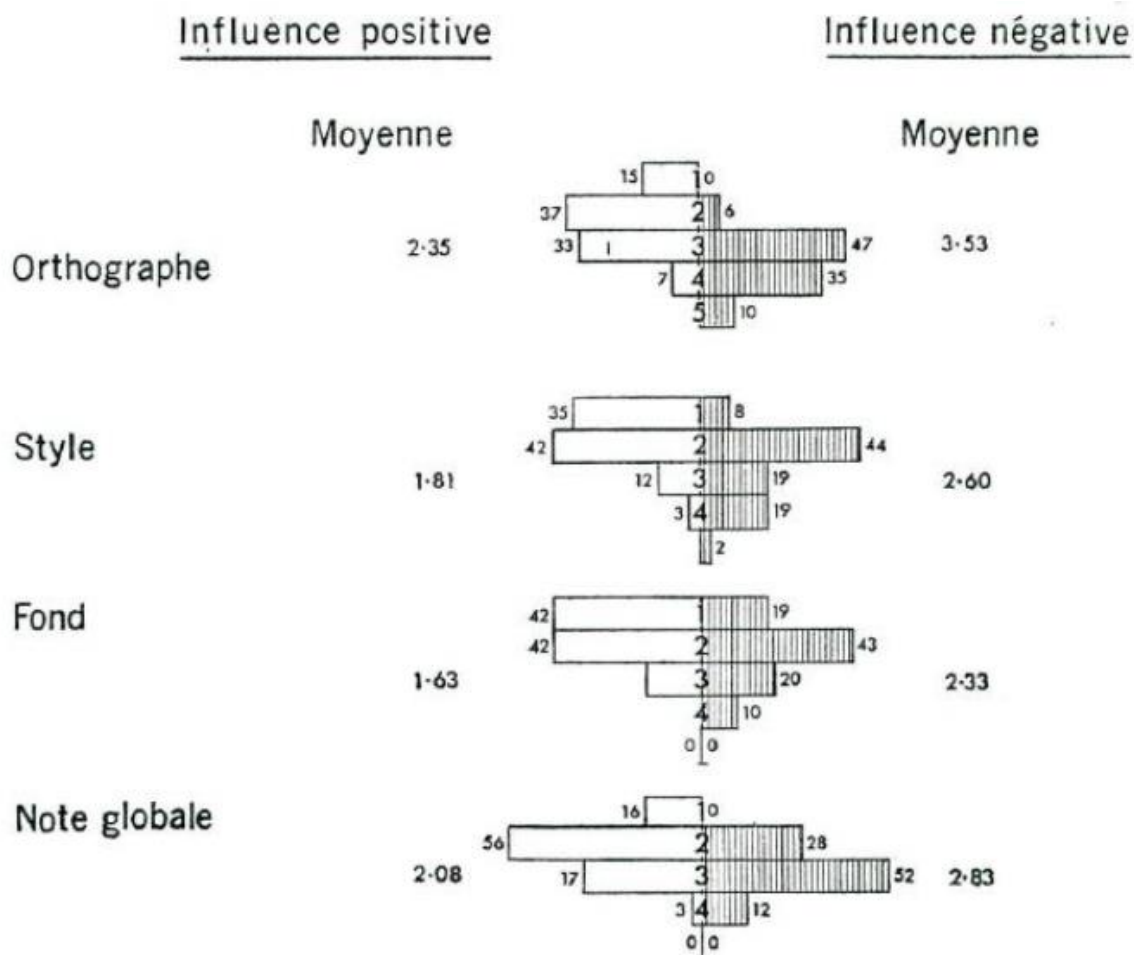
Un autre type de « parasitage » de la note résulte de l'influence de celle-ci par des aspects non pertinents. Dans ce cas, par exemple, la note est influencée (contaminée) par des caractéristiques de l'étudiant comme son aspect physique, sa présentation vestimentaire, sa prononciation ou son accent, etc. (El Youssifi, Boumahdi et Yousfi, 2022).

L'effet de halo présente un caractère affectif accusé. Souvent, on surestime les réponses d'un élève de belle allure, au regard franc, à la diction agréable. Soit pour des raisons de lisibilité, soit pour des raisons nettement affectives, l'écriture peut aussi influencer le correcteur (De Landsheere, 1992). Dans cet ordre d'idée, il a montré que la mauvaise qualité de l'écriture fait baisser le score.

Il a fait l'expérience suivante. Deux rédactions dactylographiées ont été soumises à 2 groupes de 46 instituteurs de 4^{ème} année primaire. Au groupe 1, il dit « Le travail 1 est l'oeuvre d'un élève moyen qui aime lire des bandes dessinées; son père et sa mère sont employés. Le travail 2 a été fait par un enfant doué ; son père est rédacteur d'un quotidien connu (Laboureur, 2023).

Pour le groupe 2, les commentaires ont été inversés. Trois aspects (orthographe, Style, Fond) devaient être jugés indépendamment, en plus d'une « note Globale », chaque fois sur une échelle à 5 niveaux (1 = TB ; 5 = insuffisant), tel que l'indique la figure 2 (Lafontaine et Monseur, 2009).

Figure 2. Notation des compositions sous l'influence d'un préjugé favorable ou défavorable



Comme le précise De Landsheere (1971), dont nous reproduisons la figure inspirée des résultats de Weiss, « Pour les quatre aspects considérés, les notes attribuées au travail pour lequel on a créé un préjugé favorable ont été significativement supérieures aux autres. Pour l'orthographe, qui semblait le plus devoir échapper à l'effet oedipien de la prédiction, on observe qu'au travail de l'élève présenté comme doué, 16% des correcteurs accordent la note très bien et aucun la note insuffisant; si le même élève est présenté comme moyen, les correcteurs n'accordent aucun très bien, mais 11% notent insuffisant ». On note ainsi que des contre-performances lors de tests peuvent être révélatrices d'évaluation surfaites lors de situations non standardisées (Elbazini, 2024).

6.3.3. Effets de contraste entre copies ou entre étudiants

Plusieurs effets parasites peuvent être identifiés comme relevant de l'interaction entre copies successives. On les qualifie d'effets de contraste ou de séquence : la copie qui suit une copie brillante risque d'être désavantagée et inversement (Raulin, 2017).

De Landsheere (1992) décrit ce phénomène de la manière suivante: « Les élèves rompus aux examens ont depuis longtemps découvert l'importance des contrastes : passer immédiatement après un candidat brillant se révèle défavorable ; succéder à plus faible que soi peut être avantageux, à condition que la médiocrité des réponses que l'interrogateur vient d'obtenir ne l'ait pas mis de trop méchante humeur. »

De manière à mettre le phénomène en évidence, Bonniol (1965) a présenté une série de devoirs à corriger par deux groupes de 9 correcteurs. Ce sont les mêmes devoirs, mais ils sont présentés dans l'ordre inverse dans les deux groupes. Il observe que les différences (importantes) entre les deux groupes « sont plutôt imputables aux deux ordres de correction qu'aux différences de critères dont les examinateurs font état ».

A partir de cette observation, il a décidé d'introduire systématiquement après le premier tiers et après le deuxième tiers de la séquence des copies initiales des copies aux propriétés (valeur de la note) connues : faibles ou très bonnes. Bonniol appelle ces copies des ancrs. Il définit le concept d'ancre de la manière suivante: « un stimulus privilégié qui joue comme un stimulus de référence, soit parce qu'il est présent plus fréquemment que les autres, soit parce qu'il est situé dans une position particulière, soit parce qu'il est signalé d'une manière ou d'une autre à l'attention du sujet ».

Il appelle Ancre Haute une copie meilleure que les autres, et Ancre Basse une copie moins bonne que les autres. Pour lui est une Ancre Lourde la succession de trois ancrs du même type (Bonora, 1996).

Pour De Landsheere (1992), on émet deux hypothèses : l'introduction des ancrs exercera des effets de contraste, se traduisant par des déplacements dans l'échelle d'évaluation par surestimation ou sous-estimation des travaux succédant à l'ancre dans la série, et par modification de l'étendue de l'échelle utilisée » Il rapporte que dans sa série d'expériences, Gjorgjevski a extrait de ses 100 copies : 12 jugées « Très bien » (échelon 4) ; 12 jugées « Médiocre » (échelon 2). Dans chacun de ces groupes de 12, il a glissé 3 copies jugées « Bonnes » (échelon 3). Dans le premier groupe, les trois copies ont vu leur moyenne passer de 3 à 2,4 et

dans l'autre groupe, de 3 à 3,87. Ce qui confirme les deux hypothèses signalées par De Landsheere.

6.3.4. L'instabilité d'un même correcteur

Les effets de séquence, de contraste, etc. mais aussi des variations internes au correcteur (fatigue ou distraction momentanée, hasard) font qu'un même correcteur peut, à des moments différents, donner des notes différentes à une même copie. Avec quelles conséquences pour les candidats, se demandera-t-on dans une perspective pratique ? (Kaufmann, 1983).

Il a demandé à 14 historiens de noter une deuxième fois 15 compositions 12 à 19 mois après les avoir notées une première fois. Toute trace de correction avait été effacée. Les professeurs accordaient non seulement des points, mais indiquaient la réussite globale ou l'échec. Dans 92 cas sur 210, soit près de la moitié des cas, le verdict a été différent d'une fois à l'autre (Plane, 2002).

6.3.5. Les différences entre correcteurs

On peut mettre en évidence, en faisant corriger la même copie par plusieurs correcteurs qualifiés, des différences parfois fort importantes entre les notes attribuées à celle-ci. Ce type d'études a été mené très tôt (Martin, 2002).

Dans une expérience, rapportée par Piéron (1963, p. 123), une même composition française a été jugée par 76 professeurs de français. Voici la distribution de leurs notes (NP = Nombre de correcteurs attribuant une note donnée), dans le tableau 2 :

Tableau 2. Distribution des notes attribuées à une même copie par différents correcteurs

Note	0-1	2-3	4-5	6-7	8-9	10-11	12-13
NP	1	6	20	34	10	3	2

De manière à corriger ce phénomène, Laugier et Weinberg ont appelé valeur « vraie » la moyenne d'un nombre assez grand de notations indépendantes, pensant qu'en multipliant les correcteurs, on compensera leurs fluctuations (Piéron, 1963, p.22). Ils ont cherché à déterminer le nombre minimum d'examineurs compétents auxquels il faudrait faire appel pour obtenir la notation méritant confiance. Dans ce but, ils ont utilisé la formule de Spearman-Brown qui a été présentée dans la partie relative à l'accroissement de la fidélité des tests en fonction de la longueur, selon la théorie classique (Pelnard-Considère et Reuchlin, 1969).

Pour rappel, la formule de Spearman-Brown, qui a été décrite de manière générale pour tout allongement quelconque d'un test par un coefficient m , peut s'écrire de la manière suivante :

$$\rho_{mm} = \frac{m\rho_{11}}{1 + (m-1)\rho_{11}}$$

Où : ρ_{11} est la fidélité du test initial ; ρ_{mm} est la fidélité du test de longueur modifiée.

Au départ de cette formule, les auteurs vont considérer la fidélité inter-correcteurs comme la fidélité originale ρ_{11} . Cette fidélité inter-correcteurs est établie sur la base de la corrélation des notes remises par deux correcteurs confrontés aux mêmes copies. S'il existe plus de deux correcteurs, on calculera la corrélation moyenne au départ de toutes les corrélations calculables entre les notes transmises par chaque paire de correcteurs. La valeur ρ_{mm} sera la fidélité inter-correcteurs qui résulterait de la multiplication du nombre de correcteurs par m .

Ainsi, si 4 correcteurs fournissent une fidélité inter-correcteurs moyenne de 0,870, on obtiendra respectivement les fidélités inter-correcteurs suivantes: 0,953 pour 12 correcteurs (soit $m= 3$) ; 0,964 pour 16 correcteurs (soit $m= 4$) ; 0,982 pour 32 correcteurs (soit $m= 8$).

Dans le cas du doublement du nombre de correcteurs (soit 8 correcteurs et $m= 2$), la formule, dans le premier cas, devient en effet, après substitution :

$$\rho_{mm} = \frac{2 * 0,87}{1 + (2-1)0,87} = 0,93$$

On peut aussi se poser le problème inverse: de combien de correcteurs devrait-on disposer pour obtenir une fidélité inter-correcteur donnée. La réponse s'obtient à partir d'une simple transformation de la formule de Spearman-Brown:

$$m = \frac{\rho_{mm}(1 - \rho_{11})}{\rho_{11}(1 - \rho_{mm})}$$

Ainsi, si l'on désire une fidélité inter-correcteurs d'au moins 0,90, alors que la fidélité inter-correcteurs moyenne de départ, établie sur 4 correcteurs est de 0,87, on devra multiplier le nombre de correcteurs par 1,34, ce qui impliquerait 6 correcteurs (en fait, la valeur calculée indique 5,4, mais il faut bien envisager le recours à des correcteurs entiers !). La formule, appliquée dans le cas d'une fidélité inter-correcteurs de 0,99 s'écrit de la manière suivante :

$$m = \frac{0,99 * (1 - 0,87)}{0,87 * (1 - 0,99)} = 14,8$$

Dans ce cas particulier, il faudra donc avoir recours à $14,8 * 4$ correcteurs, soit environ 60 correcteurs (en fait, 59,2 d'après le calcul).

Se basant sur cette formule, Piéron (1963) rapporte dans le tableau 3 les résultats estimés par Laugier et Weinberg pour ce qui concerne les épreuves du baccalauréat: « Recherchant un coefficient élevé de fidélité (0,99), et se fondant sur les moyennes des indices de corrélation obtenus pour chaque catégorie d'épreuves, ils ont trouvé que ce nombre minimum était le suivant :

Tableau 3. Nombre estimé de correcteurs pour obtenir une fidélité inter-correcteurs par domaine

Domaine	Nombre estimé de correcteurs pour obtenir une fidélité inter-correcteurs de 0,99
Composition française	78
Version latine	19
Anglais	28
Mathématique	13
Dissertation philosophique	127
Physique	16

D'autres résultats, des mêmes auteurs, sont rapportés par Agazzi (1967): Pour les 6 mêmes domaines, 6 correcteurs ont chaque fois noté les examens de 0 à 20, une note inférieure à 10 signifiant l'échec. Le tableau 4 suivant présente les résultats obtenus pour l'ensemble des 6 correcteurs. La première colonne indique le nombre de copies refusées par les 6 correcteurs, la dernière, le nombre de copies acceptées par ces 6 correcteurs et la colonne du centre, le nombre de copies pour lesquelles on enregistre au moins une note discordante (au moins un refus et 5 notes suffisantes ou l'inverse).

Tableau 4. Les résultats obtenus pour l'ensemble des correcteurs par domaine

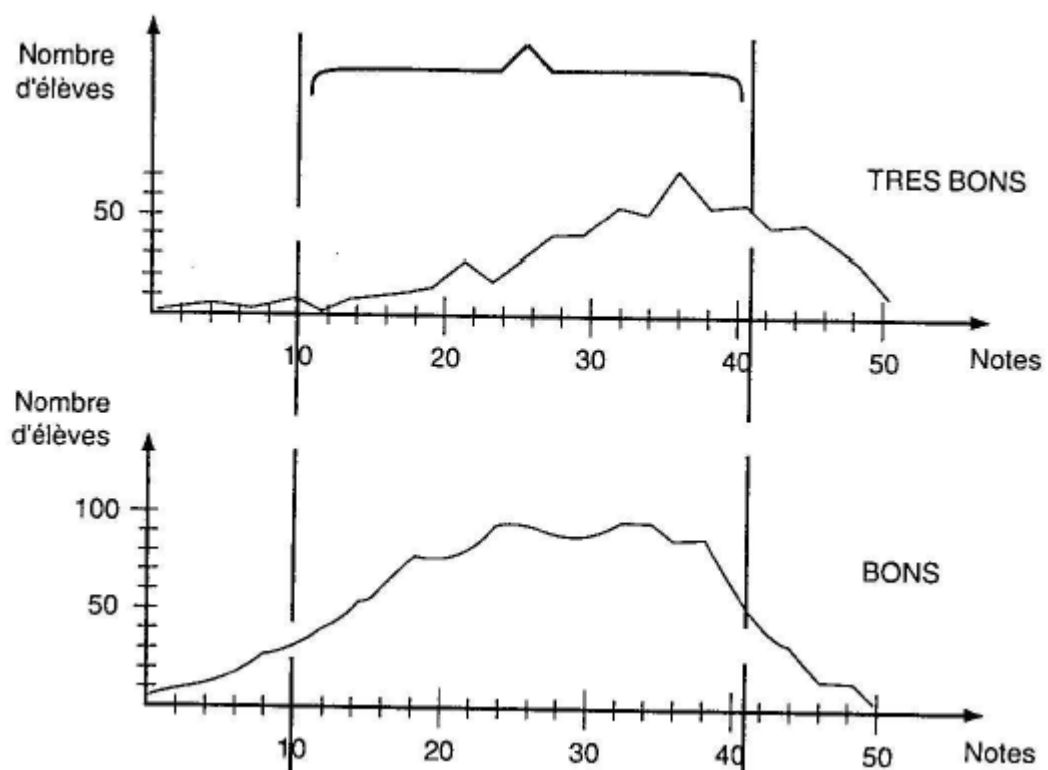
	6 notes insuffisantes	Au moins un avis discordant	6 notes suffisantes
Version latine	40%	50%	10%
Composition française	21%	70%	9%
Anglais	37%	47%	16%
Mathématique	44%	36%	20%
Philosophie	9%	81%	10%
Physique	37%	50%	13%

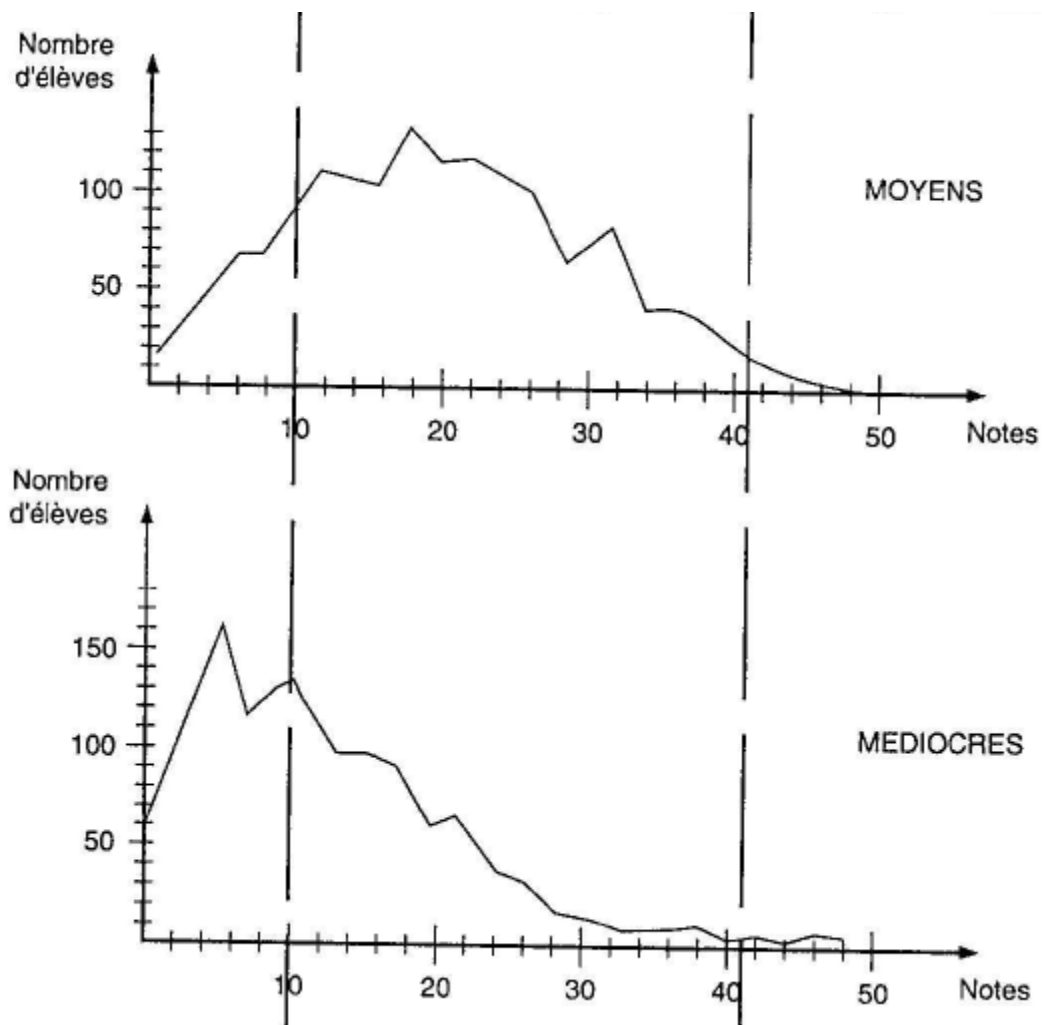
Piéron (1963) et De Landsheere (1971) ne manquent pas de présenter d'autres exemples encore. Ce type de résultats a, entre autre, provoqué la remise en cause des notes chiffrées. Malheureusement, d'autres expériences ont aussi montré les limites du système d'appréciations globales du type "Très bien, Bien, Satisfaisant, Faible, Insuffisant", comme nous allons le voir. On n'a pas manqué de penser que des expressions verbales (d'ailleurs en nombre plus limité que 21 notes possibles) augmenteraient la concordance inter-correcteurs. D'où l'adoption par des systèmes scolaires entiers (la Communauté française de Belgique, par exemple) de ce type d'échelles (Duru-Bellat, 1986).

Il avertit cependant du danger : l'instituteur, certainement, connaît mieux que personne les points du programme qui sont acquis ou non par chacun de ses élèves. Ce qu'il ignore, c'est la gravité qui s'attache à chaque faiblesse, à chaque lacune, lorsqu'on la considère non plus au sein d'une classe qui peut être « forte » ou « faible », mais par rapport à l'ensemble du pays. De là, les divergences d'appréciation mises en lumière par l'enquête (Lazher et Aroua, 2023).

Ces divergences sont illustrées par les 4 courbes ci-dessous. Elles sont issues d'une enquête nationale française, menée en 1958, ici sur le calcul au « cours moyen 2^{ème} année », ce que l'on appelle la 5^{ème} primaire en Belgique. Les instituteurs avaient été invités à attribuer à chaque élève un des 4 adjectifs suivants pour caractériser son niveau en calcul : très bon, bon, moyen, médiocre. De cette manière, 654 élèves furent jugés très bons, 1303 bons, 1551 moyens, 1300 médiocres. La catégorie « moyen » est celle qui rassembla le plus grand nombre d'élèves. Ces élèves ont par ailleurs subi un test de calcul noté « objectivement » de 0 à 50. La figure 3 ci-dessous reproduit, pour chacune des « catégories d'élèves » résultant de l'avis des maîtres, la distribution des notes au test.

Figure 3. La distribution des notes au test pour chacune des catégories d'élèves





Ces quatre distributions se recouvrent largement : dans la zone de notes qui va de 10 à 40, le même niveau de performance au test peut malheureusement correspondre à n'importe laquelle des 4 notes verbales globales (Cambon, 1968).

6.4. En guise de conclusion

Nous l'avons précisé avant d'entamer ce chapitre, la docimologie critique a permis d'attirer l'attention des correcteurs et des enseignants sur la nécessité d'un soin particulier quand aux différents parasitages possibles de la notation. Les expériences nombreuses et déjà anciennes qui sont mentionnées, et qui ne constituent qu'un petit échantillon de ce qui a pu être réalisé entre les années 1920 et les années 1970, ne doivent pas conduire au rejet absolu de la notation subjective. C'est impossible. Ces résultats ne doivent sans doute pas plus conduire au rejet de toute forme d'évaluation en dehors de l'usage de questions à réponse fermée (vrai/faux, QCM), mais faire réfléchir à des formules efficaces - la méthode d'examen doit rester praticable dans des

conditions normales - et justes. Ce second critère est essentiel car il s'agit d'apprécier les compétences de sujets humains. Les examens décident de plus en plus, quand il ne s'agit pas de concours, du sort de personnes. Il convient d'y être attentif. Et si cette attention n'est pas nécessairement spontanée chez tous les examinateurs, le risque est grand de voir intervenir de plus en plus d'autres acteurs dans la sphère scolaire à l'occasion de l'évaluation. On constate en effet que la judiciarisation (le fait de recourir de préférence à des solutions judiciaires pour régler des litiges, plutôt qu'à l'accord amiable ou à la médiation) et la juridiciarisation (la propension à privilégier le recours aux tribunaux pour trancher des litiges qui pourraient être réglés par d'autres voies, la médiation, l'accord amiable) sont deux menaces importantes qui pèsent sur la liberté énorme, pour ne pas dire totale, qui avait prévalu jusqu'il y a peu dans le domaine de l'évaluation scolaire. Pour s'en convaincre, il suffit de se reporter au dossier d'information préparé pour la journée organisée conjointement par l'Association pour le Développement des Méthodologies d'Evaluation en Education et l'Association internationale de Pédagogie universitaire, le 4 décembre 2001 à Liège (Leclercq, Nicaise et Demeuse, 2004). On peut notamment consulter l'article de De Landsheere sur la responsabilité civile découlant de l'enseignement dispensé » et qui est paru dans le Journal des procès du 30 décembre 1988. Ce dernier attire l'attention sur l'importance du phénomène de judiciarisation aux Etats-Unis, il y a plus de quinze ans, et peut préfigurer ce qui pourrait se produire de plus en plus chez nous, si on n'apporte pas un soin particulier aux évaluations.

Références bibliographiques

Agazzi, A. (1967). *Les aspects pédagogiques des examens*. Strasbourg: Conseil de l'Europe.

Allal, L. (1981). *Stratégies d'évaluation formative : conceptions psychopédagogiques et modalités d'application*. Paris : Peter Lang.

Allal, L. (2007). Évaluation : lien entre enseignement et apprentissage. Dans V. Dupriez et G.

Lachapelle (Eds.), *Enseigner* (pp. 139-145). Paris : PUF.

Allal, L. (2008). Évaluation des apprentissages. Dans A. Henriot-Van Zanten (Ed.), *Dictionnaire de l'éducation* (pp. 311-314). Paris : PUF.

Allal, L. et Lopez, L.M. (2005). L'évaluation formative de l'apprentissage : revue de publication en langue française. *Synthèse des publications*, 3, 265-290.

Aylwin, U. (1995). Apologie de l'évaluation formative. *Pédagogie collégiale*, 8(3), 24-32.

Baribeau, A. (2015). *Analyse des pratiques d'évaluation d'enseignants du secondaire 4 et 5 dans des décisions sommatives de certification des apprentissages des élèves*. Thèse de doctorat en sciences de l'éducation. Université du Québec à Trois-Rivières. Trois-Rivières, Canada.

Barrère, A. (2000). Sociologie du travail enseignant. *L'Année sociologique*, 50(2), 469-492.

Barroso, J. (2017). L'évaluation des apprentissages et la régulation des politiques éducatives. *Journal international de recherche en éducation et formation*, 3(2), 9-19.

Baudelot, C., Establet, R. (1971). *L'école capitaliste en France*. Paris: Maspero.

Baudrit, A. (2007). Apprentissage coopératif/ Apprentissage collaboratif : d'un comparatisme conventionnel à un comparatisme critique. *Les sciences de l'éducation*, 40(1), 115-136.

Beauset, R., Seha, M. et Duroisin, N. (2024). Etude exploratoire des conceptions, des perceptions et des pratiques déclarées d'enseignants francophones belges et français en géométrie. *Grand N*, 113, 75-104.

Bercier-Larivière, M. et Forgette-Giroux, R. (1999). L'évaluation des apprentissages scolaires : une question de justesse. *Revue canadienne de l'éducation*, 24(2), 169-182.

- Berthiaume, D., David, J. et David, T. (2010). Réduire la subjectivité lors de l'évaluation des apprentissages à l'aide d'une grille critériée : repères théoriques et application à un enseignement interdisciplinaire. *Revue internationale de pédagogie de l'enseignement supérieur*, 27(2), 1-13.
- Bonora, D. (1996). Les modalités de l'évaluation. *Revue internationale d'éducation de Sèvres*, 11, 1-13.
- Bonniol, J.-J. (1965). Les divergences de notation tenant aux effets d'ordre de la correction. *Cahiers de Psychologie*, 8, 181-188.
- Bonniol, J.-J., Piolat, M. (1971). Comparaison des effets d'ancrage obtenus dans une tâche d'évaluation. Expérience de multi-correction en mathématique et en anglais. *Actes du 17^{ème} Congrès international de psychologie appliquée*, 8, 1179-1189.
- Bouchemaa, S. (2022). L'évaluation des compétences en classe de langue : quels principes ? Quelles utilisations ? *Revue des sciences humaines*, 33(1), 891-899.
- Boudon, R. (1973). *L'inégalité des chances*. Paris: Armand Colin.
- Bourdieu, P. et Passeron, J.C. (1964). *Les héritiers*. Paris : Minuit.
- Bourdieu, P., Passeron, J.-C. (1970). *La reproduction. Eléments pour une théorie du système d'enseignement*. Paris: Les éditions de Minuit.
- Bouvy, T. et Warnier, L. (2016). *Evaluer les acquis des étudiants à l'aide des QCM*. Louvain-la-Neuve : Université Catholique de Louvain.
- Boyer, P. (2012). *La compétence morphographique d'élèves de première secondaire. L'effet du biais d'évaluation de son efficacité personnelle sur la performance*. Thèse de doctorat en sciences de l'éducation. Université du Québec à Montréal. Montréal, Canada.
- Bressoux, P. (1994). Note de synthèse : Les recherches sur les effets-écoles et les effets maîtres. *Revue française de pédagogie*, 108, 91-137.
- Bressoux, P. (1995). Les effets du contexte scolaire sur les acquisitions des élèves : effet école et effets-classes en lecture. *Revue française de sociologie*, 36, 273-294.
- Cambon, J. (1968). Normes pédagogiques et problèmes d'orientation scolaire. Etude effectuée à un pallier d'orientation : la classe de troisième. *Bulletin de l'Institut national d'étude du travail et d'orientation professionnelle*, 24, 3-187.

Caverni, J.-P., Fabre, J.-M., Noizet, G. (1975). Dépendance des évaluations scolaires par rapport à des évaluations antérieures : études en situation simulée. *Le Travail Humain*, 38(2), 213-222.

Chanudet, M. (2019). *Etude des pratiques évaluatives des enseignants dans le cadre d'un enseignement centré sur la résolution de problèmes en mathématiques*. Thèse de doctorat en sciences de l'éducation. Université de Genève. Genève, Suisse.

Charlin, B., Bordage, G. et Von Der Vleuten, C. (2003). L'évaluation du raisonnement clinique. *Pédagogie médicale*, 4(1), 42-52.

Chaumont, M. (2015). *Documenter le jugement professionnel d'enseignantes et d'enseignants lors de l'évaluation certificative de compétences qui se développent dans plus d'un cours au collégial*. Mémoire de master en sciences de l'éducation. Université de Sherbrooke. Sherbrooke, Canada.

Chevallard, Y. (1991). *Vers une analyse didactique des faits d'évaluation*. Bruxelles: De Boeck Université.

Côté, F. (2017). L'évaluation des apprentissages au collégial. Un réseau de concepts pour guider les pratiques évaluatives. *Pédagogie collégiale*, 30(4), 3-9.

Crahay, M. (1996). *Peut-on lutter contre l'échec scolaire ?* Bruxelles: De Boeck Université.

Crahay, M. (2000). *L'école peut-elle être juste et efficace ? De l'égalité des chances à l'égalité des acquis*. Bruxelles: De Boeck Université.

Dachet, D. (2022). *Questions d'actualité en docimologie : des biais cognitifs aux paramètres de l'évaluation*. Liège : Université de Liège.

Daigle, M. (2013). *Évaluation de l'implantation d'un portfolio comme outil pédagogique de développement et d'évaluation des compétences en soins critiques d'étudiantes en sciences infirmières*. Mémoire de master en sciences infirmières. Université du Québec à Rimouski. Rimouski, Canada.

Dardenne, B. (1999). *Psychologie sociale*. Liège: Université de Liège.

Deaudelin, C., Desjardins, J., Dezutter, O., Thomas, L., Corriveau, A., Lavoie, J., Bousadra, F. et Hébert, M. (2007). L'évaluation formative en contexte de renouveau pédagogique au primaire : analyse de pratiques au service de la réussite. *Nouveaux cahiers de la recherche en éducation*, 10(1), 27-45.

De Bosscher, S., Fournier, V. et Carton, L. (2017). Notation en français et en mathématiques en cycle 3. Quelle influence du sexe de l'élève et de l'enseignant-e ? *Spiral-E*, 59, 23-35.

De Ketele, J.M. (1993). L'évaluation conjugquée en paradigmes. *Revue française de pédagogie*, 103, 59-80.

De Ketele, J.-M. et Gerard, F.-M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences. *Mesure et évaluation en éducation*, 28(3), 1-26.

De Landsheere, G. (1971). *Evaluation continue et examens. Précis de Docimologie*. Bruxelles: Editions Labor et Paris: Fernand Nathan.

De Landsheere, G. (1992). *Evaluation continue et examens. Précis de docimologie*. Bruxelles : Edition Labor.

Delorme, F. (2005). *Evaluation et modélisation automatiques des connaissances des apprenants à l'aide de cartes conceptuelles*. Thèse de doctorat en informatique. Université nationale des sciences appliquées. Rouen, France.

Demeuse, M (2002). *Analyse critique des fondements de l'attribution des moyens destinés à la politique de discrimination positive en matière d'enseignement en Communauté française de Belgique*. Thèse doctorale en sciences de l'éducation. Université de Liège. Liège, Belgique.

Detroz, P. et Romainville, M. (2017). *Comprendre et améliorer les pratiques d'évaluation des acquis des étudiants dans l'enseignement supérieur*. Namur : Université de Namur.

Dierendonck, C., Burton, R. et Wanlin, P. (2009). L'évaluation des apprentissages scolaires au Luxembourg. *Mesure et évaluation en éducation*, 32(3), 101-123.

Dionne, E. (2005). Enjeux en évaluation des apprentissages dans le cadre de programmes d'études formulés par compétences : l'exemple de science et technologie. *Mesure et évaluation en éducation*, 28(2), 49-66.

Dubet, F., Martuccelli, D. (1996). *A l'école. Sociologie de l'expérience scolaire*. Paris: Seuil.

- Dupeyroux, P. (1950). La docimologie. *Bulletin du groupe d'études de psychologie de l'Université de Paris*, 4(3), 93-94.
- Dupuis, M.E. et Dubois, M.J. (2024). *La planification de l'évaluation d'une situation complexe*. Paris : Agence française de développement.
- Duru-Bellat, M. (1986). Notation et orientation : quelle cohérence, quelle conséquence ? *Revue française de pédagogie*, 77, 23-38.
- Duru-Bellat, M. (1995). Note de synthèse : Filles et garçons à l'école, approches sociologiques et psycho-sociales. *Revue française de pédagogie*, 110, 75-109.
- Duru-Bellat, M., Mingat, A. (1993). *Pour une approche analytique du système éducatif*. Paris: Presses Universitaires de France.
- Elbazini, A. (2024). La pratique évaluative entre la subjectivité et l'objectivité des enseignants-évaluateurs : cas du cycle secondaire. *Revue internationale du chercheur*, 5(1), 294-323.
- El Youssifi, M., Boumahdi, A. et Yousfi, B. (2022). Evaluation des effets des biais de la correction sur l'objectivité du jugement évaluatif en enseignement supérieur. *Revue marocaine de l'évaluation et de la recherche en éducation*, 8, 258-278.
- Endrizzi, L. et Rey, O. (2008). L'évaluation au cœur des apprentissages. *Veille scientifique et technologique*, 38, 1-19.
- Fagnant, A. et Goffin, C. (2017). Les conceptions des futurs enseignants du secondaire en matière d'évaluation : entre un accord de principe et une vision limitée de l'évaluation formative. *Mesure et évaluation en éducation*, 40(1), 1-32.
- Felouzis, G. (1993). Interactions en classe et réussite scolaire. Une analyse des différences filles-garçons. *Revue française de sociologie*, 34, 199-222.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.
- Fontaine, S. et Loye, N. (2017). L'évaluation des apprentissages : une démarche rigoureuse. *Pédagogie médicale*, 18(1), 189-198.
- Fontaine, S., Savoie-Zajc, L. et Cadieux, A. (2013). L'impact des CAP sur le développement de la compétence des enseignants en évaluation des apprentissages. *Education et francophonie*, 41(2), 10-34.

- Forgette-Giroux, R., Simon, M. et Bercier-Larivière, M. (1996). Les pratiques d'évaluation des apprentissages en salle de classe : perceptions des enseignantes et des enseignants. *Revue canadienne de l'éducation*, 21(4), 384-395.
- Gerard, F.M. (2006). L'évaluation des acquis des élèves dans le cadre de la réforme éducative en Algérie. Dans *Réforme de l'éducation et innovation pédagogique en Algérie* (pp. 85-124). Alger : Programme d'appui de l'UNESCO à la réforme du système éducatif (PARE).
- Gerard, F.M. (2013). L'évaluation au service de la régulation des apprentissages : enjeux, nécessités et difficultés. *Revue française de linguistique appliquée*, 18(1), 75-92.
- Gilles, J.L. (2002). *Construction et gestion qualité d'outils d'évaluation*. Liège : Université de Liège.
- Good, T. (1987). Two decades of research on teacher expectations: Findings and future directions. *Journal of Teacher Education*, 24, 32-47.
- Grimand, A. (2007). L'appropriation des outils de gestion et ses effets sur les dynamiques organisationnelles : le cas du déploiement d'un référentiel des emplois et des compétences. *Revue management et avenir*, 54, 237-257.
- Grisay, A. (1984). Les mirages de l'évaluation scolaire. Rendements en français, notes et échecs à l'école primaire ? *Revue de la Direction Générale de l'Organisation des Etudes*, 19(5), 29-42.
- Grisay, A. (1991). Que peut-on prescrire en matière d'éducation-bilan ? Bruxelles: De Boeck Université.
- Guérin-Lajoie, S., Hébert, M.H. et Papi, C. (2024). L'évaluation en situation authentique. *Le réseau de l'Université du Québec*, 13(1), 1-2.
- Guillaumin, J. (1968). L'aspect interpersonnel de la notation : de la docimologie à la doxologie pédagogique. *Bulletin de la société de Binet et Simon*, 86, 250-275.
- Guimard, P., Cosneyfroy, O. et Florin, A. (2007). Evaluation des comportements et des compétences scolaires par les enseignants et prédiction des performances et des parcours à l'école élémentaire et au collège. *L'orientation scolaire et professionnelle*, 36(2), 1-26.
- Halima, B. (2022). *Le rôle et l'utilité de l'évaluation formative dans l'enseignement-apprentissage du français langue étrangère. Cas des apprenants de la 2^{ème} année secondaire*. Mémoire de master en didactique et ingénierie de la formation. Université Ibn Khaldoun-Tiaret. Alger, Algérie.

Hameline, D. (1979). *Les objectifs pédagogiques en formation initiale et en formation continue*. Paris : ESF.

Hameline, D. (1982a). *Formes nouvelles d'évaluation*. Paris : Institut National de Recherche Pédagogique.

Hameline, D. (1982b). L'évaluation formative dans un enseignement différencié. *Revue française de pédagogie*, 60(1), 73-78.

Hayem, M. (2024). *Optimisation de l'évaluation continue dans l'enseignement du français langue étrangère au cycle primaire : enjeux et orientations futures*. Mémoire de master en didactique et langues appliquées. Université 8 mai 1945 Guelma. Alger, Algérie.

Holman, J. (2005). *Une exploration de l'interaction sociale en ligne lors de la réalisation d'activités d'apprentissage collaboratif dans deux espaces interactifs : un site internet et des wikis*. Mémoire de master en technologie éducative. Université Laval. Laval, Canada.

Howe, R. et Ménard, L. (1993). Un nouveau paradigme en évaluation des apprentissages. *Pédagogie collégiale*, 6(3), 36-40.

Howe, R. et Ménard, L. (1994). Croyances et pratiques en évaluation des apprentissages. *Pédagogie collégiale*, 7(3), 21-27.

Hutmacher, W. (1993). *Quand la réalité résiste à la lutte contre l'échec scolaire. Analyse du redoublement dans l'enseignement primaire genevois*. Genève: Service de la Recherche sociologique.

Issaieva, E. et Crahay, M. (2010). Conceptions de l'évaluation scolaire des élèves et des enseignants : validation d'échelles et étude de leurs relations. *Mesure et évaluation en éducation*, 33(1), 31-61.

Laboureur, L. (2023). Existe-t-il un biais de l'évaluation lié à son origine social dans le cadre d'une évaluation des performances à visée sélective et/ou d'une évaluation des performances à visée formative ? Mémoire de master en sciences de l'éducation. Université de Liège. Liège, Belgique.

Lafontaine, D. et Monseur, C. (2009). Les évaluations des performances en mathématiques sont-elles influencées par le sexe de l'élève ? *Mesure et évaluation en éducation*, 32(2), 71-98.

Lazher, K. et Aroua, N. (2023). Problèmes d'évaluation et de continuité d'apprentissage. Vision docimologique de l'abandon scolaire. *Revue des sciences islamiques et civilisation*, 8(1), 415-436.

Leclercq, D., Nicaise, J. et Demeuse, M. (2004). Docimologie critique : des difficultés de noter des copies et d'attribuer des notes aux élèves. Dans M. Demeuse (Dir.), *Introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation* (pp. 273-292). Liège : Les Editions de l'Université de Liège.

Lefebvre, S. et Deaudelin, C. (2001). Les interactions et la performance à l'écrit d'élèves du primaire dans une situation d'apprentissage avec les pairs soutenu par ordinateur. *Revue des sciences de l'éducation*, 27(3), 621-648.

Loye, N. et Fontaine, S. (2018). S'instrumenter pour évaluer. *Pédagogie médicale*, 19, 95-106.

Jaillet, A. (2005). Peut-on repérer les effets de l'apprentissage collaboratif à distance ? *Distances et savoirs*, 3(1), 49-66.

Jouquan, J. (2002). L'évaluation des apprentissages des étudiants en formation médicale initiale. *Revue internationale francophone d'éducation médicale*, 3(1), 38-52.

Kabore, A. (2023). *L'évaluation en situation d'effectif pléthorique dans le domaine des lettres à l'Université Joseph Ki-Zerbo : approche par le logiciel AMC*. Ouagadougou : Editions Francophones Universitaires d'Afrique.

Kaufmann, J. (1983). Une étude des différences de sévérité entre professeurs en seconde. *L'orientation scolaire et professionnelle*, 12(2), 111-129.

Kozanitis, A. (2021). Evaluation des apprentissages en formation à distance : les situations authentiques à la rescousse. *Revue internationale des technologies en pédagogie universitaire*, 18(1), 276-290.

- Laugier, H., Piéron, H., Piéron, H., Toulouse, E., Weinberg, D. (1934). *Etudes docimologiques sur le perfectionnement des examens et concours*. Paris: Conservatoire national des arts et métiers.
- Laurier, M. (2014). La politique québécoise d'évaluation des apprentissages et les pratiques évaluatives. *Education et francophonie*, 42(3), 31-49.
- Leclerc, D., Nicaise, J. et Demeuse, M. (2004). Docimologie critique : des difficultés de noter des copies et d'attribuer des notes aux élèves. Dans M. Demeuse (Dir.), *Introduction aux théories et méthodes de la mesure en sciences psychologiques et en sciences de l'éducation* (pp. 273-292). Liège : Les Editions de l'Université de Liège.
- Lepage, P. et Romainville, M. (2009). *Le tutorat en communauté française de Belgique. Inventaire des pratiques, éléments d'évaluation et recommandations*. Namur : Centre de didactique supérieure de l'académie universitaire.
- Leselbaum, N. (1982). *Autonomie et autoévaluation*. Paris: Edition Economica.
- Leyens, J.-Ph., Yzerbit, V. (1997). *Psychologie sociale*. Liège: Mardaga.
- Lieury, A. (1996). Mémoire encyclopédique et devenir scolaire : Etude longitudinale d'une cohorte sur les quatre années du collège français. *Psychologie et Psychométrie*, 17(3), 33-44.
- Lopez, L.M. et Allal, L. (2011). Le jugement professionnel en évaluation : un acte cognitif et une pratique sociale située. *Revue suisse des sciences de l'éducation*, 30(3), 465-482.
- Lopez, L. M. et Laveault, D. (2008). L'évaluation des apprentissages en contexte scolaire : développements, enjeux et controverses. *Mesure et évaluation en éducation*, 31(3), 5-34.
- Malo, A. et Monfette, O. (2023). Entre principes théoriques et usage des outils d'évaluation des compétences professionnelles en stage. Prise en compte du point de vue de stagiaires et de formateurs. *Journal international de recherche en éducation et formation*, 9(3), 97-119.
- Martin, J. (2002). Aux origines de la science des examens (1920-1940). *Histoire de l'éducation*, 94, 1-20.
- Martin, N. (2012). *Conception d'un portfolio pour documenter le développement des compétences de l'élève au collégial*. Mémoire de master en sciences de l'éducation. Université de Sherbrooke. Sherbrooke, Canada.

- Martin, E., Lefrançois, C., Guichard, A., Tapp, D. et Arsenault, L. (2016). Processus de construction d'une grille critériée pour l'évaluation de productions écrites complexes à l'Université. *Revue internationale de pédagogie de l'enseignement supérieur*, 32(2), 1-23.
- Matracci, A. (2020). Le questionnement didactique pour aborder l'évaluation des apprentissages en créativité. *Pédagogie collégiale*, 34(1), 1-5.
- Merle, P. (1996). *L'évaluation des élèves. Enquête sur le jugement professoral*. Paris: Presses Universitaires de France.
- Merle, P. (1998). *Sociologie de l'évaluation scolaire*. Paris: Presses Universitaires de France.
- Meunier, H. et Michaud, N. (2017). Une évaluation authentique en enseignement supérieur. Le cas d'un cours en évaluation des apprentissages. *Pédagogie collégiale*, 31(1), 34-39.
- Michaud, N. et Roy, M. (2022). Former et évaluer des compétences en enseignement supérieur : un dispositif d'évaluation novateur. *Revue internationale de pédagogie de l'enseignement supérieur*, 38(3), 1-17.
- Morales-Perlaza, A. et Morrisette, J. (2020). Vers la consolidation d'une sociologie de l'évaluation scolaire : une revue de la littérature narrative en français. *Mesure et évaluation en éducation*, 43(3), 39-65.
- Morrisette, J. (2009). Une resocialisation de l'évaluation des apprentissages. *Revue canadienne des jeunes chercheuses et chercheurs en éducation*, 2(1), 1-8.
- Morrisette, J. (2010). Un panorama de la recherche sur l'évaluation formative des apprentissages. *Mesure et évaluation en éducation*, 33(3), 1-27.
- Morrisette, J. et Maheux, G. (2007). Evolution de la fonction de l'évaluation formative des apprentissages à travers le discours ministériel québécois entre 1981 et 2002. *Revue des sciences de l'éducation*, 33(3), 727-748.
- Mouele, P. (2017). *L'intégration des pratiques évaluatives des apprentissages : analyse des besoins de formation continue chez les enseignants gabonais du primaire (cas de l'évaluation*

sommative). Mémoire de master en administration et évaluation en éducation. Université Laval. Laval, Canada.

Mukulu, P.K. et Cibangu, J.K. (2014). Attitude des élèves de 5^{ème} secondaire face à l'évaluation en Histoire à Kinshasa Ngaliema. *La recherche en éducation*, 11, 73-90.

Nduwingoma, P., Ntwari, I. et Ntahonkiriye, M. (2020). *Evaluation des apprentissages et analyse des pratiques de classe au Burundi*. Bujumbura : Institut de pédagogie appliquée de l'Université du Burundi.

Nguyen, D.Q. et Blais, J.G. (2007). Approche par objectifs ou approche par compétences ? Repères conceptuels et implications pour les activités d'enseignement, d'apprentissage et d'évaluation au cours de la formation clinique. *Pédagogie médicale*, 8, 232-251.

Nicaise, J. (2001). *Pratiques, sens et sens pratique au coeur des évolutions institutionnelles. Les instituteurs de sixième primaire et le jugement professoral*. Mémoire de licence en sciences de l'éducation. Université de Liège. Liège, Belgique.

Nizet, I. et Leroux, J. (2015). La construction de savoirs conceptuels en évaluation : enjeux pour le développement professionnel en contexte de formation continue. *Journal international en éducation et formation*, 2(1), 15-29.

Nolin, R. (2013). *Pratiques déclarées d'enseignement et d'évaluation de l'oral d'enseignants du primaire au Québec*. Mémoire de master en sciences de l'éducation. Université du Québec à Montréal. Montréal, Canada.

Novak, J.D. et Gowin, D.R. (1984). *Learning how to learn*. New York: Cambridge Press.

Ouedraogo, A. (2023). L'utilisation des questions fermées en évaluation des apprentissages par les enseignants de Koudougou (Burkina Faso). *Recherches et regards d'Afrique*, 5(2), 10-26.

Ouellet, S. (2008). Les compétences préalables à la construction de l'identité professionnelle chez les futurs enseignants en adaptation scolaire. Dans E.C. Molina et C. Gervais (Dir.), *Les stages en formation à l'enseignement. Pratiques et perspectives théoriques* (pp. 83-101). Québec : Presses de l'Université du Québec.

Papi, C., Gérin-Lajoie, S. et Hébert, M.H. (2020). Se rapprocher de l'évaluation à distance : dix pistes de réponse. *Journal international de recherche en éducation et formation*, 1, 201-206.

- Pelletier, P. (2012). *Le pouvoir discriminant des épreuves écrites lors d'un examen d'entrée classant*. Mémoire de master en sciences de l'éducation. Université François Rabelais. Tours, France.
- Pelnard-Considère, J. et Reuchlin, M. (1969). Travaux docimologiques sur les examens en Faculté de médecine. *Bulletin de l'Institut national d'étude du travail et d'orientation professionnelle (BINOP)*, 25(4), 241-256.
- Pépin, D. (2019). *Intégrer l'approche collaborative favorisant le développement d'un travail personnel des étudiantes et étudiants des cours de projet de design en techniques de design intérieur*. Mémoire de master en sciences de l'éducation. Université de Sherbrooke. Sherbrooke, Canada.
- Perrenoud, P. (1995). *La fabrication de l'excellence scolaire : du curriculum aux pratiques d'évaluation*. Genève: Droz.
- Perrenoud, P. (1998). *L'évaluation des élèves. De la fabrication de l'excellence à la régulation des apprentissages. Entre deux logiques*. Bruxelles: De Boeck Université.
- Petitjean, B. (1984). Formes et fonction des différents types d'évaluation. *Pratiques : linguistique, littérature, didactique*, 44, 5-20.
- Piéron, H. (1963). *Examens et docimologie*. Paris: Presses universitaires de France.
- Pilet, J., Allard, C. et Horoks, J. (2019). Une entrée par l'évaluation des apprentissages pour analyser les interactions entre l'enseignant ou l'enseignante et les élèves dans les moments de mise en commun. *Education et francophonie*, 47(3), 121-139.
- Plane, S. (2002). Les effets d'un instrument d'écriture à l'épreuve de la recherche. Deux ou trois choses que l'on sait ou que l'on ne sait pas sur le traitement de texte. *Repères*, 26(1), 163-186.
- Poumay, M. et Georges, F. (2010). *L'évaluation des compétences. Principes et illustrations*. Liège : Université de Liège.
- Pourtois, J-P., Bonacina, R., Delbecq, A., Segard, M. (1978). Le niveau d'expectation de l'examineur est-il influencé par l'appartenance sociale de l'enfant ? *Revue française de pédagogie*, 44, 34-37.

- Potvin, C. (2010). *L'évaluation de compétences à l'aide de logiciels de conception d'évaluation informatisée en technique de l'informatique au collégial*. Mémoire de master en sciences de l'éducation. Université de Sherbrooke. Sherbrooke, Canada.
- Prégent, R., Bernard, H. et Kozantis, A. (2009). *Enseigner à l'université dans une approche-programme. Guide à l'intention des nouveaux professeurs et chargés de cours*. Montréal : Presses internationales Polytechnique.
- Pronovost, L. (2018). *Conception d'un outil pour établir le seuil de réussite dans l'évaluation des compétences d'un cours dans un programme d'études collégiales*. Mémoire de master en sciences de l'éducation. Université de Sherbrooke. Sherbrooke, Canada.
- Pudelko, B., Basque, J. et Legros, D. (2003). Vers une méthode d'évaluation des cartes conceptuelles fondée sur l'analyse en systèmes. *EIAH*, 555-558.
- Raudonyté, L. et Foimapafisi, T. (2021). *Utilisation des données d'évaluation des apprentissages en Guinée*. Paris : Institut international de planification de l'éducation.
- Raulin, D. (2017). Objets et pratiques de l'évaluation scolaire. *Revue française de pédagogie*, 200, 69-79.
- Rigutto, L. (2021). *Les jugements scolaires des professeurs de sciences en Fédération Wallonie-Bruxelles concordent-ils avec les résultats aux épreuves externes certificatives ?* Mémoire de master en sciences de l'éducation. Université de Liège. Liège, Belgique.
- Rosenthal, R. A., Jacobson, L. (1971). *Pygmalion à l'école*. Tournai: Casterman.
- Rot, N., Butas, Z. (1959). Les distributions des notes scolaires comparées aux distributions des résultats obtenus aux tests de connaissances. *Le travail humain*, 12, 1-2.
- Saiah, A.A. (2016). *Evaluation formative et pédagogie différenciée. Cas des élèves de 2^{ème} année moyenne*. Mémoire de master en didactique du français langue étrangère. Université Abdelhamid Ibn Badis-Mostaganem. Alger, Algérie.
- Scallon, G. (2004). L'évaluation des compétences et l'importance du jugement. *Pédagogie collégiale*, 18(1), 14-20.

St-Onge, C. (2018). Enjeux et défis de l'évaluation longitudinale : quelques pistes de réflexion préalables à son implantation. *Pédagogie médicale*, 19, 137-142.

Suchaut, B. (2008). *La loterie des notes au bac. Un réexamen de l'arbitraire de la notation des élèves*. Bourgogne : Université de Bourgogne.

Talbot, N. (2015). *Elaboration d'une échelle d'un niveau d'adéquation perçue des pratiques d'évaluation des apprentissages en salle de classe à l'approche par compétences selon les étudiants*. Thèse de doctorat en sciences de l'éducation. Université du Québec à Montréal. Montréal, Canada.

Talbot, N. et Raïche, G. (2017). Validation du cadre de référence PIEA des pratiques d'évaluation des apprentissages en classe dans une approche par compétences selon la perception d'étudiants du collégial. *Revue des sciences de l'éducation*, 43(3), 90-123.

Tanana, M. (2011). *Evaluation formative du savoir-faire des apprenants à l'aide d'algorithmes de classification. Application à l'électronique numérique*. Institut National des Sciences Appliquées de Rouen. Rouen, France.

Tardif, J. (2003). Développer un programme par compétences : de l'intention à la mise en œuvre. *Pédagogie collégiale*, 16(3), 36-44.

Tardif, J. (2006). *L'évaluation des compétences. Documenter le parcours de développement*. Montréal : Chenelière Education.

Tessaro, W. (2015). Les épreuves externes comme soutien au développement professionnel des enseignants dans leurs pratiques d'évaluation certificative : les apports d'une recherche collaborative. *Journal international de recherche en éducation et formation*, 2(1), 49-65.

Tozzi, M. (2004). Débat scolaire : les enjeux anthropologiques d'une didactisation. *Tréma*, 23, 1-10.

Traoré, I.S. et Diarra, S.N. (2024). L'évaluation formative dans le fondamental au Mali : les soutiens collectif et individuel aux apprentissages des élèves. *Recherches et regards d'Afrique*, 115-147.

Tribollet, B., Langlois, F. et L. Jacquet, L. (2000). Protocoles d'emploi des cartes conceptuelles au lycée et en formation des maîtres. *TREMA, IUFM de Montpellier*, (18), 61, 2000.

Viallet, F. et Maisonneuve, P. (1981). *80 fiches d'évaluation pour la formation et l'enseignement*. Paris : Les Editions d'organisation.

Vigner, G. (1984). *L'exercice dans la classe de français*. Paris : Hachette.

Weber, M. (1971). *Economie et société*. Paris: Plon.

Yerly, G. (2017). Evaluation des apprentissages en classe et évaluation à grande échelle : quels sont les effets des épreuves externes sur les pratiques évaluatives des enseignants ? *Mesure et évaluation en éducation*, 40(1), 33-60.

Yerly, G. et Laveault, D. (2020). Evaluer les apprentissages en contexte de pandémie : aller au-delà de la notation pour soutenir la réussite de tous les élèves. *Formation et profession*, 28(4), 1-12.