

# La comparaison analogique comme soutien au changement conceptuel sur l'évaluation

## *Analogical comparison as support for conceptual change in assessment*

**Stéphanie Naud** – [stephanie.naud@unige.ch](mailto:stephanie.naud@unige.ch) – <https://orcid.org/0009-0008-2194-6635>

**Céline Girardet** – [celine.girardet@unige.ch](mailto:celine.girardet@unige.ch) – <https://orcid.org/0000-0002-6467-5678>

**Emmanuel Sander** – [emmanuel.sander@unige.ch](mailto:emmanuel.sander@unige.ch) – <https://orcid.org/0000-0003-0044-3437>

Université de Genève – Suisse

**Pour citer cet article** : Naud, S., Girardet, C. et Sander, E. (2024). La comparaison analogique comme soutien au changement conceptuel sur l'évaluation. *Évaluer. Journal international de recherche en éducation et formation*, 10(3), 131-165. <https://doi.org/10.48782/e-jiref-10-3-131>

### Résumé

La conception intuitive de l'évaluation s'apparente à une pratique formelle, ponctuelle et instrumentée permettant d'attribuer une note. L'erreur serait donc à éviter car synonyme d'échec. Toutefois les recherches dans le champ s'inscrivent dans une perspective visant à dépasser cette approche « sanction » de l'évaluation et à faire prévaloir une évaluation comme soutien à l'apprentissage. Le paradigme de comparaison analogique, processus de comparaison guidée, est reconnu comme favorisant une meilleure compréhension des concepts et de leurs relations. Cet article est composé de deux études réalisées en Suisse romande avec des étudiants en formation à l'enseignement secondaire. La première explore leurs conceptions à propos de l'évaluation et de l'erreur. La seconde questionne les effets de la comparaison analogique de textes académiques (tâche contrôle) versus de textes métaphoriques (tâche expérimentale) sur le changement conceptuel des participants à propos de l'évaluation et de l'erreur. Les résultats, issus d'analyses qualitatives et quantitatives, confirment que l'évaluation est prioritairement conçue par analogie avec la mesure et que l'erreur est intuitivement perçue comme un échec. La deuxième étude met en évidence qu'une formation s'appuyant sur la comparaison analogique contribue à soutenir le changement conceptuel des participants. Les effets varient suivant la nature des textes support de la comparaison analogique mais tendent vers le développement d'une conception plus intégrative de l'évaluation. L'interrelation des concepts de l'évaluation et leur influence sur le changement conceptuel sont finalement discutées.

### Mots-clés

Comparaison analogique, conception intuitive, évaluation, erreur, formation pour l'enseignement, changement conceptuel

## **Abstract**

The intuitive conception of assessment aligns with a formal, isolated and instrumented practice aimed at assigning grades. Errors are to be avoided, as they are synonymous with failure. However, research in this field seeks to move beyond this “sanction-focused” approach, to promote assessment as a means of supporting learning. The analogical comparison paradigm, a guided comparison process, is recognized for fostering a deeper understanding of concepts and their relationships. This article presents two studies conducted in French-speaking Switzerland with pre-service secondary school teachers enrolled in initial teacher education. The first study explores their conceptions of assessment and error. The second examines the effects of analogical comparison using academic texts (control task) versus metaphorical texts (experimental task) on participants’ conceptual change regarding assessment and error. Results from qualitative and quantitative analyses confirm that participants primarily conceive assessment by analogy with measurement, and intuitively perceive error as failure. The second study demonstrates that training based on analogical comparison supports participants’ conceptual change. The effects varied depending on the type of texts used but tended towards the development of a more integrative conception of assessment. Finally, the interplay between assessment concepts and their influence on conceptual change are discussed

## **Keywords**

Analogical comparison, intuitive conception, assessment, error, teaching formation, conceptual change.

## 1. Introduction

L'évaluation occupe une place certaine dans la profession enseignante. C'est un terme polysémique auquel sont rattachées des croyances, des théories et tout autant de pratiques. Cela est en partie dû aux multiples fonctions de l'évaluation, source de confusion (Cardinet, 1985 ; Crahay *et al.*, 2019). De surcroît, des conceptions associées, comme l'erreur et la justice scolaire, participent à la conceptualisation de l'évaluation.

Historiquement, celle-ci est pensée en termes de classement, de sélection des élèves (Crahay *et al.*, 2019 ; Perrenoud, 1998). Une autre conception, celle de l'évaluation formative dans le contexte de l'apprentissage des élèves, a émergé dans le cadre des travaux sur la pédagogie de maîtrise (e.g., Bloom *et al.*, 1971). Elle a progressivement été élargie vers une évaluation intégrée à l'enseignement dans une perspective de régulation des apprentissages (Allal et Mottier Lopez, 2005). Ainsi, les recherches tendent à donner à l'évaluation un statut de soutien à l'apprentissage dans une visée inclusive (Cnesco, 2023 ; Fagnant et Goffin, 2017). Dès lors, comment donner plus de place à une conception d'évaluation-soutien par rapport à une évaluation-sanction, et privilégier une conception de l'erreur comme une étape féconde et non avant tout comme un échec à éviter ?

Plusieurs recherches mettent en avant la pertinence de l'usage des métaphores pour accéder aux croyances des enseignants (Bullough et Stokes, 1994 ; Raths, 2001) mais peu d'études ont investigué l'usage des métaphores dans la formation des enseignants (Boraita et Crahay, 2013). Cet article s'intéresse aux conceptions et changement conceptuel de futurs enseignants du secondaire à propos de l'évaluation, et questionne l'effet que pourrait avoir l'usage pédagogique des analogies dans un contexte de formation initiale visant la prédominance de la conception de l'évaluation-soutien plutôt que sanction<sup>1</sup>.

## 2. L'analogie, moteur de la conceptualisation

### 2.1. *Conceptions intuitives et raisonnement analogique*

Les conceptions intuitives, représentations initiales des individus (Lautrey *et al.*, 2008), sont basées sur les croyances personnelles et l'interprétation des expériences quotidiennes. Les pratiques pédagogiques sont influencées, entre autres, par les conceptions des enseignants (Kagan, 1992 ; Patrick et Pintrich, 2001) et l'agir évaluatif ne fait pas exception (Xu et Brown, 2016). Robustes, ces conceptions agissent comme un filtre dans la perception et l'appropriation des connaissances théoriques abordées en formation (Girardet, 2020 ; Sander, 2017). Par ailleurs, les concepts d'un même champ sont interreliés, cela pouvant se modéliser par un réseau sémantique (Chi, 2008 ; Collins et Quillian, 1969 ; Dupuch et Sander, 2007). Avoir accès à ces conceptions et leurs liens permettrait de les rendre visibles et d'agir dessus. Le cursus de formation à l'enseignement devrait permettre « d'explicitier, identifier, comprendre et travailler les croyances initiales des étudiants » (Boraita et Crahay, 2013, p. 118).

Les métaphores, présentes dans le discours, reflètent un raisonnement par analogie et révèlent les conceptions des enseignants mobilisées dans une situation donnée (Bullough,

---

<sup>1</sup> Dans cet article le terme croyance réfère à une caractéristique psychologique de l'enseignant, ancrée dans sa culture (Crahay *et al.*, 2010). Ainsi la conception est la dimension cognitive d'une croyance, tout comme les convictions en seraient la dimension affective (Lafortune et Fennema, 2003).

1991). Le raisonnement par analogie est une forme de raisonnement inductif. Une nouvelle situation est comprise par analogie, c'est-à-dire par rattachement à des concepts déjà connus (Sander, 2018). Dans un contexte de formation, la cible de l'analogie (la notion scolaire, nouvelle) est appréhendée dans les termes de la source (la notion quotidienne, connue). Par exemple, un jeune élève s'appuierait sur ce qu'il sait d'une orange (source) pour mieux comprendre le principe de croute terrestre (cible). Les expériences antérieures sont capitalisées : des analogies successives font naître et enrichissent les concepts. La source de l'analogie est choisie inconsciemment par l'apprenant parmi l'ensemble de ses connaissances en mémoire (Hofstadter et Sander, 2013 ; Sternberg, 1977). L'enseignant peut soutenir le raisonnement par analogie en sélectionnant et proposant aux élèves une métaphore à valeur explicative (Gray et Holyoak, 2021). De l'appariement source/cible découlent des inférences : les propriétés de la notion quotidienne source sont projetées sur la notion scolaire cible (Holyoak et Thagard, 1989 ; Sander, 2000 ; Sternberg, 1977). En outre, certaines inférences sont issues des relations entre les concepts d'un même réseau sémantique (Dupuch et Sander, 2007). Ainsi, un concept hérite des propriétés des concepts qui lui sont superordonnés (e.g., si les animaux respirent, les mammifères respirent, les baleines respirent). Les individus conceptualisent à partir de métaphores, qui, portées par le langage, influencent leur façon de comprendre, d'apprendre, de prendre des décisions et d'agir (Lakoff et Johnson, 1980 ; Thibodeau et Boroditsky, 2011). Le raisonnement par analogie est donc un soutien cognitif (Gentner *et al.*, 2001) avec un potentiel didactique conséquent pouvant initier le changement conceptuel (Nersessian, 1999 ; Vosniadou, 2003).

## **2.2. Changement conceptuel**

### **2.2.1 Résistance et leviers**

Le caractère résistant des conceptions intuitives pose un défi dans des contextes de formation (Kagan, 1992). En effet, leur résilience est un obstacle reconnu au transfert des connaissances (Sanders et McCutcheon, 1986). Un des déclencheurs potentiels de l'évolution des croyances est la réflexion sur les croyances préalables, qui relève d'une démarche métacognitive lorsqu'elle est systématisée. Le degré de nouveauté de ces croyances et leur force de liens agissent comme modérateurs de l'évolution (Girardet, 2017). La contradiction a également été présentée comme facteur déclencheur d'une évolution des croyances (Kagan, 1992), le conflit cognitif catalysant l'évolution des conceptions (Limón, 2001).

### **2.2.2 Modèle de prévalence**

L'approche traditionnelle du changement conceptuel vise un abandon des conceptions intuitives (Nersessian, 1999) alors qu'il est établi que celles-ci perdurent voire prédominent y compris après enseignement, même pour des experts (Lautrey *et al.*, 2008 ; Willame et Snauwaert, 2018). L'idée que conceptions intuitives et nouvelles puissent coexister n'est pas récente (Chi, 1992). Issu de recherches ancrées dans le champ de la psychologie expérimentale et des neurosciences cognitives, le modèle de prévalence conceptuelle explique le changement conceptuel par un renversement de prévalence dans les choix des conceptions des individus (Potvin, 2013 ; Potvin *et al.*, 2015). Il s'agit d'un modèle alternatif aux approches traditionnelles. En effet, dans le modèle de prévalence, l'apprentissage a lieu en situation de compétition entre différentes conceptions. Ainsi, réussir une tâche nécessite une prévalence (Potvin, 2013) et une inhibition (Masson, 2012) de certaines idées. En outre, le conflit cognitif est introduit lorsque l'étudiant est en mesure d'accorder du crédit à une autre conception. Dans un contexte de formation, il s'agirait alors de rendre les étudiants conscients de leurs

conceptions intuitives, de leur présenter d'autres perspectives conceptuelles pour une même notion et de leur apprendre à inhiber les conceptions incompatibles avec une vision intégrative de l'évaluation ayant une finalité éducative.

### 2.2.3 Comparaison analogique

La comparaison entre deux situations favorise la compréhension et l'abstraction (Alfieri *et al.*, 2013 ; Kurtz *et al.*, 2001). C'est le principe du paradigme de comparaison analogique. Il s'agit de présenter à des participants deux scénarios analogues, de s'assurer de leur compréhension de la structure commune des deux scénarios en demandant de spécifier les différences, de décrire les scénarios et d'évaluer leurs similarités. De surcroît, une comparaison intensive favorise une meilleure compréhension de la structure causale commune (Kurtz *et al.*, 2001). Pour les novices comme pour les participants expérimentés, ce processus de comparaison guidée peut être un moyen efficace d'abstraire des principes en vue d'une application ultérieure (Loewenstein *et al.*, 1999). Elle peut donc être un levier dans la formation des enseignants à propos de l'évaluation.

## 3. Réseau conceptuel de l'évaluation

### 3.1. Conceptions intuitives de l'évaluation

Dans la pratique enseignante, la conception intuitive de l'évaluation s'apparente à « une pratique formelle, ponctuelle, instrumentée, débouchant nécessairement sur une note. Tout ce qui s'en écarte n'est pas conçu comme une évaluation » (Mottier Lopez, 2023, p. 18). En l'occurrence, l'évaluation formative « élargie », c'est à dire intégrant des interactions informelles et impliquant l'élève via l'autoévaluation au sens large (Allal et Mottier Lopez, 2005), n'est que rarement perçue par les enseignants comme de l'évaluation en tant que telle (Mottier Lopez, 2023). Néanmoins, l'intérêt d'enseignants et futurs enseignants à propos de l'évaluation formative pour soutenir l'apprentissage des élèves est réel, sans que l'effet sur les pratiques de classe ne soit probant (Fagnant et Goffin, 2017 ; Issaieva *et al.*, 2015).

### 3.2. Fonctions et référence de l'évaluation

Les deux fonctions de l'évaluation les plus mobilisées sont l'évaluation formative et l'évaluation certificative<sup>2</sup> (Mottier Lopez, 2015). La visée essentielle de l'évaluation formative tient en sa contribution au processus de régulation de l'enseignement et de l'apprentissage, cherchant la progression des apprenants vers les objectifs établis (Allal, 2008 ; Mottier Lopez, 2015 ; Perrenoud, 1998). L'évaluation certificative, quant à elle, aboutit à une décision de certification des acquis des élèves et intervient à la fin d'une séquence d'enseignement (Mottier Lopez, 2015)<sup>3</sup>.

Indépendamment de la fonction de l'évaluation, différents cadres de référence permettent d'interpréter les informations recueillies. Avec une référence normative, les productions d'élèves sont comparées entre elles, ce qui engendre un classement (Crahay *et al.*, 2019 ; Roegiers, 2010). Si la référence est critériée, l'interprétation des informations recueillies est faite par rapport aux critères d'évaluation spécifiant les attentes (Mottier Lopez, 2015 ;

<sup>2</sup> Dans cet article l'évaluation certificative sera indifférenciée de l'évaluation sommative.

<sup>3</sup> Une interprétation peut être normative ou autoréférenciée tout en étant également critériée (Mottier Lopez, 2015).

Roegiers, 2010). Dans le cas d'une interprétation autoréférencée, la production de l'élève est comparée à ses réalisations précédentes (Mottier Lopez, 2015).

Dans un contexte de formation à l'enseignement, la conceptualisation de l'évaluation-soutien est privilégiée en valorisant l'évaluation formative et certificative critériée par opposition à l'évaluation normative, sanction. L'objectif est de viser l'équité éducative et de réduire la place du classement normatif.

### ***3.3. Métaphores de l'agir évaluatif***

L'activité évaluative de l'enseignant dépend de plusieurs dimensions dont : la culture de l'évaluation, les compétences évaluatives, les gestes évaluatifs et la posture de l'évaluateur (Jorro, 2000, 2013). Quatre postures sont décrites métaphoriquement par Jorro (2000) et chacune représente une vision différente de la régulation de l'enseignement-apprentissage. Celles du contrôleur et du pisteuse donnent une place dominante à l'enseignant et un rôle passif à l'élève. Ces postures se distinguent par le fait que le contrôleur valorise la performance, l'effort et la réussite alors que le pisteuse encourage l'effort, la persévérance et la relation avec l'élève. Le conseiller et consultant offrent davantage de liberté à l'élève avec l'autoévaluation. Le conseiller adopte une approche constructiviste, considère l'erreur comme formative et s'intéresse aux démarches des élèves. Quant au consultant, il attribue une fonction émancipatrice au savoir et reste en retrait pour permettre l'émergence du sens. Chacune de ces postures a donc un rapport spécifique à l'erreur (Jorro, 1996).

La posture évaluative de l'enseignant peut varier au cours d'une même tâche (Jorro, 2000). Considérant par ailleurs qu'un nombre limité de postures adoptées par l'enseignant risquerait de restreindre les postures d'apprentissage des élèves (Bucheton et Soulé, 2009), cette diversité est valorisée dans le contexte de formation à l'enseignement.

### ***3.4. Conceptions de l'erreur***

La conception de l'erreur et le modèle pédagogique de l'enseignant sont liés (Astolfi, 2015). Ainsi, dans les modèles transmissif et behavioriste, l'erreur a un écho négatif, elle est sanctionnée. Synonyme d'échec, l'objectif est alors de la prévenir pour l'éviter. A contrario, l'approche constructiviste associe l'erreur à une étape dans le parcours d'apprentissage de l'élève. Avec ce statut positif, elle est perçue comme un outil pour apprendre et pour enseigner (Bachelard, 1993 ; Astolfi, 2015 ; Gagné *et al.*, 2023). Dans cette perspective, au lieu d'éviter l'erreur, il s'agit de concevoir des situations didactiques facilitant leur émergence pour les traiter et s'appuyer dessus. En effet, l'erreur est le support d'un diagnostic sur les conceptions intuitives des élèves, susceptibles de faire l'objet de régulation (Sander *et al.*, 2018). La conception positive de l'erreur est liée au paradigme de l'évaluation-soutien à l'enseignement et apprentissage.

## **4. Objectifs de la recherche**

A la lumière de ces apports, la présente recherche se compose de deux études articulées à un dispositif de formation. Celui-ci mobilise la comparaison analogique, dans le but de soutenir la prévalence conceptuelle de l'évaluation-soutien relativement à l'évaluation-sanction.

Une première étude explore les conceptions des étudiants en formation à l'enseignement secondaire à propos de l'évaluation et de l'erreur, avant un dispositif de formation. Au regard de la littérature, l'hypothèse est que les futurs enseignants expriment une conception de

L'évaluation proche d'une fonction certificative ainsi qu'une conception plutôt délétère de l'erreur. Les résultats obtenus pour la première étude constituent les données prétest de la seconde étude.

L'objectif de la seconde étude est de comparer les effets d'une tâche de comparaison analogique sur le changement conceptuel des étudiants. Dans la condition expérimentale, les participants réalisent une comparaison analogique de textes métaphoriques, récits filant une métaphore pour illustrer des pratiques et conceptions de l'évaluation et de l'erreur. La condition contrôle consiste, quant à elle, en une comparaison analogique de textes académiques, extraits d'un chapitre d'ouvrage. L'hypothèse est qu'analyser des textes à dominante métaphorique soutiendrait davantage le changement conceptuel des étudiants à propos de l'évaluation et de l'erreur par rapport à l'analyse de textes académiques.

## **5. Étude 1 (prétest) : recueil des conceptions avant le dispositif de formation**

### ***5.1. Contexte et participants***

Les participants à cette recherche sont les étudiants en dernière année de formation initiale à l'enseignement secondaire, toutes disciplines confondues, inscrits à l'institut universitaire de formation à l'enseignement de Genève. Cet échantillon non aléatoire est composé de 90 personnes dont 46 % de femmes. La moyenne d'âge est de 33 ans (écart type de 7 ans).

### ***5.2. Recueil de données***

Le recueil des données a été réalisé via un questionnaire administré en ligne. Le temps moyen pour y répondre a été de 26 minutes. Chaque item figurait sur une page distincte, sans possibilité de retour aux précédentes. L'ensemble du questionnaire est présenté en annexe (annexe A).

L'ensemble des étudiants a d'abord rédigé une définition de l'évaluation. Puis, une série de 14 affirmations ont été présentées de façon contrebalancée pour éviter que l'ordre de passation exerce une influence sur les réponses des participants. Ces items, issus de la recherche de Crahay *et al.* (2013), invitaient les étudiants à exprimer leur degré d'accord quant aux visées de l'évaluation : certificative critériée (vérification de l'atteinte des objectifs fixés), normative (sélection, classement) ou formative (régulation des apprentissages). Voici un exemple d'affirmation pour chacune des trois visées : « lors des évaluations, les réponses des élèves permettent de cerner les progrès qu'ils ont réalisés depuis le début de l'année » (rattachée à l'évaluation formative) ; « Les évaluations permettent de distinguer les bons et les mauvais élèves » (rattachée à l'évaluation normative) ; « Les évaluations permettent de voir si les élèves maîtrisent le contenu de l'enseignement » (rattachée à l'évaluation certificative critériée). Il y a 5 affirmations associées à la visée certificative, 5 pour la visée formative et 4 pour la visée normative. La réponse consistait en un positionnement sur une échelle de Likert en quatre niveaux, allant de 0 (pas du tout d'accord) à 3 (tout à fait d'accord) pour chacune des propositions. Le score obtenu est donc compris entre 0 et 3.

La tâche suivante présentait les métaphores de postures évaluatives et les gestes évaluatifs associés (Jorro, 2000) en ordre contrebalancé. La Figure 1 précise les formulations proposées aux étudiants. Chacun sélectionnait une ou plusieurs métaphore(s) de posture puis l'associait au(x) moment(s) d'une séquence d'enseignement fictive : mise en situation, production initiale, modules et/ou production finale (Gvozdic *et al.*, sous presse).

En tant que contrôleur-se, j'effectue un jugement sur les performances des élèves en distinguant la réussite de l'insuffisance. J'annote les productions en complément des appréciations.

En tant que pisteur-se, j'anticipe les cheminements et je suis la progression de l'élève. J'encourage l'effort et j'instille le goût du savoir.

En tant que conseiller-ère, j'analyse les tâches, je m'intéresse aux démarches des élèves et j'incite à l'auto-évaluation. Je différencie mes interventions.

En tant que consultant-e, je reste en retrait et j'instaure une communauté de pairs. Je favorise la parole des élèves et m'intéresse à leur rapport au savoir.

**Figure 1 :** Métaphores illustrant les postures évaluatives

La dernière tâche consistait en une formulation de définitions de l'erreur et de l'évaluation formative.

### **5.3. Méthode de codage et d'analyse**

Les définitions (évaluation, erreur et évaluation formative) ont été insérées dans un tableau, puis codées et analysées au regard des conceptions sous-jacentes mobilisées par les étudiants. La construction du codage s'est majoritairement faite a priori, les variables ayant été choisies en s'appuyant sur la littérature traitant ce sujet. Lors du traitement des données, une variable a fait l'objet d'un codage plus affiné. Ainsi, la dimension informative de l'évaluation formative a été déclinée en deux items : informative pour l'enseignant et informative pour l'élève.

Afin d'évaluer la fiabilité des codages, un tiers des réponses (sélectionnées aléatoirement) a été codé en double aveugle<sup>4</sup>. Une concordance de 88 à 100 %, selon les variables, a été obtenue à la suite d'une mesure d'accord inter-juges via le Kappa de Cohen. L'accord inter-juges peut être qualifié d'excellent.

Concernant les définitions de l'évaluation, le codage s'est centré sur la présence ou non des éléments suivants : une référence à la mesure (mesure, échelle, niveau, température, chiffrée, taille, règle, etc.) ; l'évocation d'une note (note, résultat, chiffrée, etc.) ; une référence aux critères d'évaluation (critères, objectifs, repères, attendus, etc.) ; la présence d'une vision binaire (réussi ou échoué, acquis ou non, validé ou non, etc.) ; l'évocation de régulation (régulation, adaptation, réorganisation, différenciation, etc.) et la référence à une épreuve (épreuve, contrôle, bilan, test, formel, etc.). Les non-réponses ont aussi été codées. Le Tableau 1 exemplifie ce codage.

---

<sup>4</sup> Par deux des auteurs de l'article.

**Tableau 1 :** Exemple de codage des définitions produites par les étudiants à propos de l'évaluation

<b>Evaluer c'est ...</b>	<b>Mesure</b>	<b>Note</b>	<b>Critères</b>	<b>Binaire</b>	<b>Régulation</b>	<b>Test</b>
Juger des connaissances ou des compétences à des fins d'apprentissage, de diagnostic et parfois de certification. C'est donc un outil pédagogique.	0	0	0	0	1	0
Donner des critères pour savoir si les éléments sont acquis ou non.	0	0	1	1	0	0
Déterminer le niveau d'adéquation d'une performance/d'un travail par rapport à un objectif choisi.	1	0	1	0	0	0
Vérifier l'acquisition et la compréhension de connaissances par un test.	0	0	0	0	0	1

Note. Pour ce codage, 0 = non et 1 = oui

Le codage des définitions de l'évaluation formative s'est centré sur la présence ou non des éléments suivants : la référence à un « test à blanc » (non noté) ; une projection de régulation de l'enseignement (adaptation, différenciation, etc.) ; le rôle informatif pour l'enseignant (diagnostic, constat, etc.) et pour l'élève (feedbacks, critères, objectifs, etc.). Le Tableau 2 illustre ces modalités.

Le codage des définitions de l'erreur rend compte de la présence ou non d'une vision de l'erreur négative (échec, faute, écart, etc.), neutre (une étape) ou positive (constitutive de l'apprentissage, nécessaire, essentielle, etc.).

L'adhésion des étudiants aux différentes visées de l'évaluation est exprimée par un score moyen compris entre 0 et 3. Ce score est calculé en faisant la moyenne des scores obtenus pour les différentes affirmations associées à chaque visée de l'évaluation (échelle de Likert). L'adhésion de l'étudiant est considérée comme faible si le score est inférieur ou égal à 1, l'adhésion est dite moyenne s'il est supérieur à 1 et inférieur ou égal à 2 et elle est forte si le score est supérieur à 2.

En outre, le choix des métaphores évaluatives a fait l'objet d'une analyse permettant de mettre en évidence la diversité de postures rapportées (opérationnalisée par le nombre de postures sélectionnées), ainsi que leur place dans la séquence d'enseignement fictive, c'est-à-dire les moments de la séquence d'enseignement auxquels les postures sont rapportées par l'étudiant.

Les traitements statistiques ont été réalisés avec le logiciel SPSS version 25.0 (IBM SPSS Statistics, IBM Corporation). Pour évaluer la significativité des résultats observés, le test d'indépendance du Khi-deux avec un seuil de 0,05 a été utilisé. De plus, la force d'association a été mesurée avec le V de Cramer (variables à plus de deux modalités) et le coefficient Phi (variables à deux modalités).

## 5.4. Résultats de l'étude 1

### 5.4.1 Conceptions intuitives de l'évaluation et de l'erreur

Parmi les 87 définitions produites par les étudiants, 42,5 % mobilisent les principes de la mesure. La référence à une épreuve formelle apparaît dans 5,7 % des définitions, 17,2 % associent l'évaluation à la note et 23,0 % déclarent que l'évaluation se base sur des critères ou objectifs. Enfin, 11,0 % y associent une régulation à posteriori et 2,3 % des réponses évoquent une visée de classement pour l'évaluation. En outre, la variable de mobilisation du concept de note et celle de mesure ont une association moyenne et significative ( $\chi^2 (1, N = 87) = 4,32, p = 0,038, \Phi = 0,223$ ).

En ce qui concerne les 88 définitions à propos de l'erreur, 86,4 % des réponses associent l'erreur à l'échec, lui donnant ainsi un statut exclusivement négatif. 11,4 % des productions définissent l'erreur comme un élément constitutif de l'apprentissage (vision positive). Les réponses subsistantes expriment une vision neutre de l'erreur (ni un échec, ni nécessaire à l'apprentissage).

Sur les 85 définitions de l'évaluation formative, 24,7 % l'associent à un test à blanc, 42,4 % expriment que l'évaluation formative informe l'élève, 61,2 % qu'elle a une fonction de diagnostic pour l'enseignant et 40,0 % évoquent explicitement une régulation de l'enseignement.

### 5.4.2 Degré d'adhésion aux fonctions normative, formative et certificative

Le Tableau 3 présente les scores d'adhésion (de 0 à 3) calculés à la suite du positionnement des étudiants par rapport aux principes des évaluations formative, certificative critériée et normative.

**Tableau 3 :** Moyennes des scores d'adhésion en fonction du type d'évaluation

	N = 88	
Évaluation formative	M = 2,34	SD = 0,52
Évaluation normative	M = 0,82	SD = 0,38
Évaluation certificative critériée	M = 2,18	SD = 0,58

Les étudiants expriment une adhésion relativement forte (score > 2 sur 3) pour l'évaluation certificative critériée et formative et plutôt faible pour l'évaluation normative. Pour estimer la significativité statistique de ces résultats, une comparaison des moyennes deux à deux est nécessaire. Ainsi, un test ANOVA à mesure répétée auquel est appliquée une correction de Bonferroni (nombre de comparaisons faible) permet d'affirmer que les écarts de scores sont significatifs ( $F(2, 174) = 313,416, p < 0,001$ ).

En outre l'adhésion forte à l'évaluation formative est fortement associée à une adhésion forte pour l'évaluation certificative critériée ( $\chi^2 (2, N = 88) = 24,04, p < 0,001$  et V de Cramer = 0,523). Les analyses montrent que 27,0 % de la variance du score d'adhésion à l'une de ces deux formes d'évaluation est expliquée par l'autre ( $R^2$  ajusté = 0,2704).

Le score d'adhésion est plus bas pour l'évaluation normative. En effet, 68,2 % des répondants expriment une faible adhésion aux principes d'évaluation normative (score  $\leq 1$  sur 3).

#### 5.4.3 Diversité des postures évaluatives au fil d'une séquence fictive

Il ressort que 70,4 % des étudiants mobilisent une unique posture évaluative dans la séquence d'enseignement fictive (N = 89). Parmi les réponses exprimant une seule posture, celle de contrôleur est largement minoritaire avec moins de 2,0 % des réponses. Au sein des 28,6 % de réponses mobilisant plusieurs postures, 58,3 % sollicitent celle du contrôleur. Elle est majoritairement choisie dans la phase de production finale et les métaphores de conseiller, consultant et pisteur sont davantage mobilisées dans les autres phases de la séquence d'enseignement. Le Tableau 4 détaille ces résultats.

**Tableau 4 :** Proportions de mobilisations déclarées des postures d'évaluateur en fonction du moment de la séquence d'enseignement

	<b>Contrôleur (N=16)</b>	<b>Pisteur (N=41)</b>	<b>Conseiller (N=46)</b>	<b>Consultant (N=24)</b>
Mise en situation	0 %	<b>70,8 %</b>	32,6 %	41,6 %
Production initiale	43,8 %	<b>70,8 %</b>	<b>67,4 %</b>	50,1 %
Modules	18,8 %	73,2 %	<b>69,6 %</b>	<b>62,6 %</b>
Production finale	<b>81,3 %</b>	7,3 %	30,4 %	12,6 %

En considérant un réseau sémantique « évaluation-sanction » il s'agit de questionner l'association entre la posture de contrôleur, la conception négative de l'erreur, la conception de l'évaluation comme mesure et l'adhésion à l'évaluation normative. Les mesures d'association réalisées deux à deux entre ces variables par le biais de coefficients de Cramer ne montrent pas de relation statistiquement significative pour chaque paire de ces quatre variables. Il ressort tout de même que 9,1 % des participants expriment une adhésion moyenne à forte à l'évaluation normative tout en mobilisant la posture évaluative de contrôleur. Cette relation est faible (V de Cramer = 0,184) mais a une tendance à la significativité ( $\chi^2$  (1, N = 88) = 2,98,  $p = 0,084$ ).

Par ailleurs dans un réseau sémantique « évaluation-soutien », des analyses de force d'association sont réalisées entre la conception positive ou neutre de l'erreur, la régulation des apprentissages et les critères comme éléments constitutifs de l'évaluation, l'adhésion à l'évaluation formative et à l'évaluation critériée et la mobilisation des postures de pisteur, conseillé ou consultant. Une faible association, à tendance significative, est observée entre la posture du pisteur et l'adhésion à l'évaluation formative ( $\chi^2$  (1, N = 88) = 3,08,  $p = 0,079$  et V de Cramer = 0,187). Cette double mobilisation (pisteur et évaluation formative) concerne 29,5 % des participants. Il est aussi intéressant de noter que 17,2 % des participants expriment à la fois la notion de critères dans leur définition de l'évaluation et adhèrent fortement à l'évaluation certificative critériée alors que 75,8 % des personnes adhèrent à l'évaluation certificative critériée sans exprimer la notion de critères dans leur définition de l'évaluation. Ce constat est tout de même à nuancer au regard du manque de significativité statistique ( $\chi^2$  (2, N = 87) = 4,22,  $p = 0,121$  et V de Cramer = 0,220). Les autres paires de variables testées ne montrent qu'une très faible association (V de Cramer < 0,2 et  $p > 0,1$ ).

### 5.5. Discussion de l'étude 1

L'hypothèse selon laquelle les futurs enseignants expriment une conception de l'évaluation proche d'une fonction certificative et une vision négative de l'erreur est vérifiée par cette première étude. Les résultats montrent une adhésion forte pour l'évaluation formative lorsque les réponses sont proposées dans les items fermés du questionnaire, ce qui est cohérent avec des études antérieures (e.g. Fagnant et Goffin, 2017 ; Issaieva *et al.*, 2015). Or, lorsque les étudiants définissent le terme « évaluation » (sans que sa fonction soit spécifiée dans la consigne), ils l'associent principalement à l'évaluation certificative (mesure, note), négligeant des éléments tels que la régulation, qui renverraient à sa fonction formative. Le concept « évaluation » semble rattaché à la conception intuitive « évaluation certificative » pour les étudiants interrogés, ce qui rejoint le constat de Mottier Lopez (2023), et conduit notamment à ce que la fonction de l'évaluation privilégiée par les futurs enseignants soit certificative. Cette conception intuitive pourrait provenir de leur propre expérience scolaire (Mansfield et Volet, 2010). En outre, le caractère inconscient des conceptions intuitives pourrait constituer une explication au fait que les étudiants interrogés déclarent adhérer fortement aux principes d'évaluation formative tout en exprimant la prévalence de la visée certificative dans leur définition de l'évaluation.

Dans les formulations, l'erreur est majoritairement définie en termes d'échec. Cela rejoint la perspective certificative de l'évaluation, avec une dichotomie « réussite-échec » qui peut être liée à la posture évaluative du contrôleur (Jorro, 2000). Dans les réponses, la métaphore du contrôleur est associée à la phase de production finale, laissant penser que, pour définir l'erreur, les futurs enseignants auraient principalement pensé à une évaluation certificative en fin de séquence d'enseignement. Ces résultats donnent donc à voir la prévalence de la conception certificative de l'évaluation comme ancrage pour penser l'évaluation, mais également pour penser la notion d'erreur.

Lorsque les étudiants proposent une définition de l'évaluation formative, leurs réponses montrent que celle-ci joue davantage un rôle informatif pour l'enseignant que pour l'élève. Or, la valeur informative de l'évaluation devrait être élevée à la fois pour l'enseignant et les élèves (Hadji, 2017 ; Morales Villabona *et al.*, 2023). Par ailleurs, la perspective de régulation est présente dans moins de la moitié des définitions, pourtant elle est censée être constitutive de l'évaluation formative.

La faible adhésion des répondants à l'évaluation normative entre en concordance avec le peu de mentions d'éléments définitoires de l'évaluation évoquant le classement. L'évaluation critériée semble donc bien établie dans les conceptions des futurs enseignants du secondaire, se détachant de la conception historique de l'évaluation ayant comme fonction le classement des élèves (Crahay *et al.*, 2019 ; Perrenoud, 1998). Cela rejoint les constats de Crahay *et al.* (2013) à propos des étudiants en formation d'enseignant primaire à Genève. Ces réponses peuvent aussi être influencées par la désirabilité sociale de correspondre à la culture du contexte de formation. Il s'agirait également d'observer concrètement si ces principes restent pérennes après la formation et permettent à ces enseignants novices d'échapper à la « constante macabre » identifiée par Antibii (2003), qui dénonce la pression implicite de maintenir artificiellement une certaine proportion d'échecs lors de la notation et créer des hiérarchies scolaires inconscientes.

La majorité des futurs enseignants du secondaire interrogés déclarent une unique posture d'évaluation dans la séquence d'enseignement alors que les étudiants en enseignement

primaire de Genève n'étaient qu'un quart à ne pas diversifier les postures évaluatives (Gvozdic *et al.*, sous presse). Cet écart interroge la différence de culture de l'évaluation entre le primaire et le secondaire. Il pourrait être expliqué par un manque de connaissance des postures évaluatives possibles de la part des futurs enseignants du secondaire ou par le fait de ne pas être assez informés de l'importance, pour l'élève, de varier les postures d'enseignement (Bucheton et Soulé, 2009).

Enfin, le peu de lien observable entre les variables d'un même réseau sémantique au sein des réponses formulées, que ce soit pour l'évaluation-sanction ou l'évaluation-soutien à l'apprentissage, pourrait être expliqué par le fait que les étudiants n'aient pas fait de lien entre les divers items du questionnaire en ligne. Plus largement il s'agirait de questionner explicitement les relations qu'ils pourraient établir entre ces différents concepts enseignés en formation. En effet, Dupuch et Sander (2007) expliquent que les relations entre les concepts d'un même réseau sémantique peuvent être ignorées des apprenants car le mode d'apprentissage n'a pas soutenu leur construction. La significative statistique limitée de ces résultats pourrait aussi être due au nombre trop restreint de réponses recueillies.

Dans une perspective de formation, il s'agit de soutenir les étudiants dans leur changement conceptuel à propos de l'évaluation, notamment au travers du statut de l'erreur, de la place de l'élève et des visées de l'évaluation. Dans le cursus de formation à l'enseignement secondaire de Genève cela passe par la valorisation du réseau sémantique de l'évaluation-soutien, prônant une diversification des postures évaluatives, mettant en avant l'intérêt pédagogique de l'erreur et la promotion du rôle de l'évaluation comme soutien à l'apprentissage (régulation, critères) par rapport à une évaluation normative, sanction (classement). Il s'agit en outre de rendre explicite les relations entre les différents concepts du champ de l'évaluation. C'est dans ce double objectif que la seconde étude a été construite.

## **6. Étude 2 (posttest) : évolution des conceptions avec une tâche de comparaison analogique**

Cette étude, expérimentale, vise à comprendre dans quelle mesure la pratique guidée de comparaison analogique de textes métaphoriques à propos de l'évaluation stimule le changement conceptuel des participants à propos de l'erreur et de l'évaluation par rapport à une tâche contrôle.

### **6.1. Participants, protocole et méthode d'analyse**

Les participants sont les mêmes que pour la première étude. Ils ont été aléatoirement répartis en deux groupes, expérimental et contrôle, à effectifs équivalents en suivant leur regroupement habituel de séminaire. Le questionnaire de l'étude 1, prétest de cette seconde étude, a été réalisé par les participants quelques jours avant de bénéficier d'un module d'enseignement. Ces trois cours, dispensés à tous les étudiants indépendamment de leur assignation, traitaient de l'apport de la sociologie à propos du classement à l'école, des pratiques de différenciation pédagogique et des troubles des apprentissages.

Un ultime séminaire (1h30) concernait la prise en compte des besoins éducatifs particuliers dans un contexte d'évaluation. Les étudiants du groupe contrôle et du groupe expérimental étaient dans des salles séparées avec des chargés d'enseignement différents. Afin de s'assurer qu'au-delà de la manipulation expérimentale étudiée, les autres aspects soient les plus équivalents possible dans les deux groupes comparés, les chargés d'enseignement ont été

préparés lors d'une réunion, avec un des chercheurs menant l'étude, à suivre le protocole précis de séance (contenus, modalités, temps, consignes, supports). La seconde moitié de la séance était dédiée à une analyse de textes (condition expérimentale ou contrôle) et la passation d'un questionnaire posttest (identique au prétest). Dans les deux groupes, les étudiants devaient effectuer une tâche de comparaison analogique de textes traitant de trois types d'évaluation : normative, formative et certificative critériée. Chaque groupe devait lire puis analyser ses textes, guidé par un tableau de comparaison (Figure 2).

	Evaluation type 1	Evaluation type 2	Evaluation type 3
Rôle(s) de l'enseignant·e			
Rôle(s) de l'élève			
Statut de l'erreur			
Place accordée au suivi des apprentissages, des progrès			
Objectif(s) de l'évaluation			

**Figure 2 :** Tableau de synthèse guidant la comparaison analogique, commun aux deux conditions

Le groupe contrôle a effectué cette analyse en s'appuyant sur un texte académique, extrait du chapitre « L'évaluation des élèves : Docteur Jekyll and Mister Hyde de l'enseignement » (Crahay *et al.*, 2019). Le contenu sélectionné abordait la visée de l'évaluation en opposant perspective d'excellence (évaluation normative) et de maîtrise (évaluation formative et certificative critériée), la régulation de l'enseignement-apprentissage et la coresponsabilité enseignant-élève dans l'évaluation et la régulation. La figure ci-après est la diapositive projetée pendant ce temps de travail spécifique au groupe contrôle.

**Extraits du texte : L'évaluation Des Élèves : Docteur Jekyll And Mister Hyde de L'enseignement (Crahay, Mottier Lopez & Marcoux, 2019).**

Aide à la lecture : tableau de synthèse

- 10 min de lecture individuelle guidée par les items du tableau.
- 10 min par groupe de 3 : tableau de synthèse.
- 15 min en collectif : l'essentiel selon vous pour passer de l'évaluation sanction à l'évaluation soutien ?
- 15 min en individuel : prise de recul (questionnaire)

	Eval normative	Eval certif critériée	Eval formative
Rôle(s) de l'enseignant·e			
Rôle(s) de l'élève			
Statut de l'erreur			
Place accordée au suivi des apprentissages, des progrès			
Objectif(s) de l'évaluation			

**Figure 3 :** Guide pour la tâche de comparaison analogique du groupe contrôle

Les étudiants du groupe expérimental devaient aussi effectuer une tâche de comparaison analogique, mais basée cette fois sur trois textes métaphoriques (Annexe B) illustrant chacune des trois formes d'évaluation sans qu'elles ne soient explicitement nommées (normative, certificative critériée et formative). La création de ces textes a été assistée par l'intelligence artificielle conversationnelle, basée sur la technologie *generative pre-trained transformer* (GPT), afin qu'ils aient la même structure, le même niveau de langue et la même longueur. Tous trois filent la métaphore automobile, mais se distinguent par le rôle de l'enseignant, celui de l'élève, le statut de l'erreur, le suivi des apprentissages et les visées de l'évaluation. Cela diffère en fonction du type d'évaluation (formative, certificative critériée ou normative). L'évaluation formative est explicitée au travers de la métaphore d'un trajet guidé par GPS, tandis que l'évaluation certificative critériée est associée à l'examen du permis de conduire. Le troisième texte présente l'évaluation normative à travers la métaphore d'une course automobile.

Les étudiants des deux groupes disposaient d'une dizaine de minutes pour lire les textes, seuls. Puis, en trio, ils réalisaient la comparaison analogique, guidés par le tableau (Figure 2) pour questionner le rôle de l'enseignant, de l'élève, du statut de l'erreur, du suivi des progrès et des objectifs pour chacun des trois types d'évaluation (pendant 10 minutes). Pour soutenir la lecture du texte académique, plus long (11870 caractères contre 5530), les types d'évaluation étaient précisés (formative, certificative critériée et normative). Ce n'était pas le cas des textes métaphoriques. Par ailleurs, pour des étudiants en dernière année de formation universitaire les textes académiques sont davantage familiers que des textes à dominante métaphorique. Les participants assignés au groupe expérimental avaient donc un texte plus court à lire mais devaient eux-mêmes inférer la source et la cible de l'analogie portée par les textes métaphoriques. La durée accordée a convenu aux deux groupes.

À la suite, un échange collectif d'une dizaine de minutes, dans chacune des deux conditions, visait à répondre à la problématique de départ : comment passer de l'évaluation-sanction à l'évaluation-soutien à l'apprentissage ? Enfin chaque étudiant répondait au questionnaire en ligne, identique au prétest (voir l'étude 1 et l'Annexe A). Le posttest a donc eu lieu une semaine après le prétest.

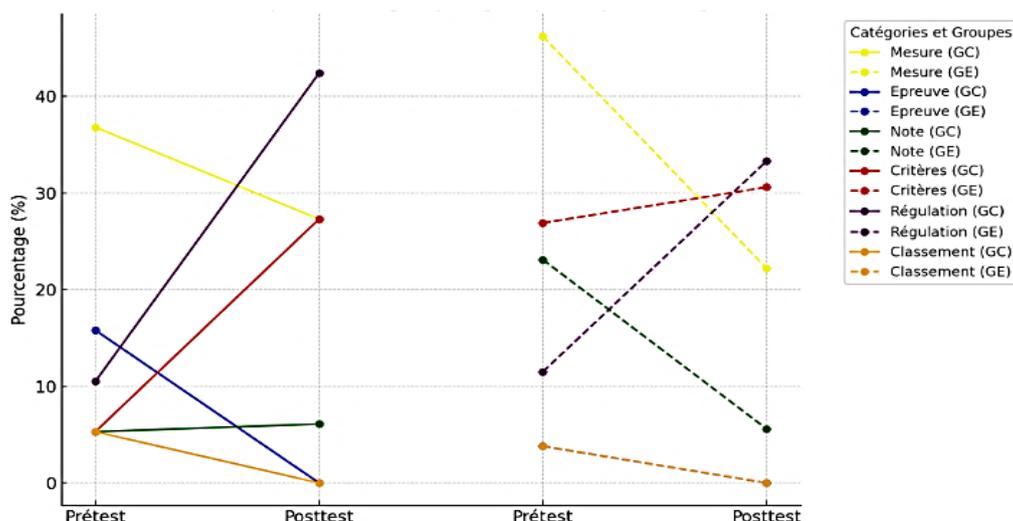
Les méthodes de codage et d'analyse des données qualitatives et quantitatives posttest ont suivi les mêmes procédés que pour la première étude. Il s'agit aussi de comparer les réponses produites en pré et posttest en fonction du groupe de participants, expérimental ou contrôle.

## **6.2. Résultats de l'étude 2**

### **6.2.1 Conceptions intuitives de l'évaluation et de l'erreur**

En posttest, 36 définitions de l'évaluation ont été formulées par le groupe expérimental (GE) et 33 pour le groupe contrôle (GC). Le graphique suivant (Figure 4) compare le pourcentage de mobilisation des concepts en fonction des groupes (contrôle et expérimental) et du moment du test (pré ou post).

En posttest, 36 définitions de l'évaluation ont été formulées par le groupe expérimental (GE) et 33 pour le groupe contrôle (GC). Le graphique suivant (Figure 4) compare le pourcentage de mobilisation des concepts en fonction des groupes (contrôle et expérimental) et du moment du test (pré ou post).



**Figure 4 :** Graphique illustrant l'évolution des réponses en fonction du test et du groupe

Les différences, parfois conséquentes, de résultats entre le groupe contrôle et le groupe expérimental au prétest, amènent à considérer davantage les écarts de progression pré/posttest entre les deux groupes plutôt que de s'intéresser uniquement aux comparaisons de résultats posttest. Le Tableau 5 rend compte de ces résultats.

**Tableau 5 :** Comparaison des définitions de l'évaluation produites par les étudiants

	Pré GC	Post GC	Ecart pré/post GC	Pré GE	Post GE	Ecart pré/post GE
Mesure	36,8 %	27,3 %	-11,5	46,2 %	22,2 % *	-24,0
Epreuve	15,8 %	0 % *	-15,8	3,8 %	0 %	-3,8
Note	5,3 %	6,1 %	0,8	23,1 %	5,6 % *	-17,5
Critères	5,3 %	27,3 % *	22,0	26,9 %	30,6 %	3,7
Régulation	10,5 %	42,4 % *	31,9	11,5 %	33,3 % *	21,8
Classement	5,3 %	0 %	-5,3	3,8 %	0 %	-3,8

Note. \*  $p \leq 0,05$  pour les différences pré et posttest d'un même groupe.

En posttest, les écarts de résultats entre les deux groupes ne sont pas statistiquement significatifs, quelle que soit la variable observée. Certains écarts descriptifs sont à considérer. En effet, en ce qui concerne la mobilisation de la dimension de régulation, l'écart en posttest entre les deux groupes n'est pas statistiquement significatif ( $\chi^2 (1, N = 69) = 0,60, p = 0,436$ ), mais en rapportant les cotes des deux groupes en posttest, il apparaît que les étudiants du groupe contrôle expriment 1,27 fois plus le concept de régulation dans leurs définitions (+31,9) que le groupe expérimental (+21,8). Une comparaison intragroupe permet d'affirmer que la mobilisation de la notion de régulation dans les définitions de l'évaluation a évolué de façon positive et significative pour les deux groupes entre le pré et posttest.

En prétest, l'écart entre les réponses des deux groupes quant à la mobilisation du concept de note n'était pas statistiquement significatif ( $\chi^2 (1, N = 45) = 2,65, p = 0,103$ ) mais tout de même de l'ordre de 17,8 points de pourcentage. Entre le pré et le posttest, les définitions du

groupe expérimental mobilisent significativement moins la note ( $\chi^2 (1, N = 62) = 4,12, p = 0,042$ ), se rapprochant ainsi des proportions du groupe contrôle.

Le phénomène inverse s'est produit pour la notion de critère d'évaluation. En effet, en posttest, les réponses du groupe contrôle ont atteint le niveau de celles du groupe expérimental. Leur écart en prétest avait une tendance à la significativité ( $\chi^2 (1, N = 45) = 3,52, p = 0,065$ ) avec 21,6 points d'écart (5,3 % et 26,9 %). Entre le pré et le posttest, l'augmentation de la mobilisation du concept de critère dans les définitions de l'évaluation est significative pour le groupe contrôle avec 22 points de pourcentage en plus ( $\chi^2 (1, N = 52) = 3,76, p = 0,052$ ) mais pas pour le groupe expérimental qui exprimait déjà un fort taux en prétest (3,7 points en plus au posttest avec  $p = 0,492$ ).

La diminution de l'expression d'épreuve formelle est plus importante pour le groupe contrôle ( $\chi^2 (1, N = 52) = 5,53, p = 0,019$ ) que pour le groupe expérimental ( $p = 0,419$ ) qui avait déjà un taux très bas en prétest.

Le concept de mesure est significativement moins mobilisé entre le pré et posttest pour le groupe expérimental avec 24 points de pourcentage de moins ( $\chi^2 (1, N = 62) = 3,96, p = 0,047$ ). Pour le groupe contrôle, cette évolution est de -11 points de pourcentage mais n'est pas statistiquement significative ( $p = 0,472$ ). En outre, le concept de classement, déjà peu présent en prétest, n'apparaît plus dans les définitions du posttest, quel que soit le groupe d'étudiants.

Le Tableau 6, ci-dessous, expose l'analyse des définitions de l'évaluation formative au travers des concepts mobilisés en fonction du moment et du groupe d'étudiants.

**Tableau 6 :** Comparaison des définitions de l'évaluation formative produites par les étudiants

	Pré GC	Post GC	Ecart pré/post GC		Pré GE	Post GE	Ecart pré/post GE
Test à blanc	19,0 %	0,0 %*	- 19		38,5 %	12,1 %*	- 26,4
Diagnostique	76,2 %	78,6 %	2,4		50,0 %	60,6 %	10,6
Informe l'élève	47,6 %	53,6 %	6		42,3 %	39,4 %	- 2,9
Régulation	42,9 %	17,9 %*	- 25		38,5 %	27,3 %	- 11,2

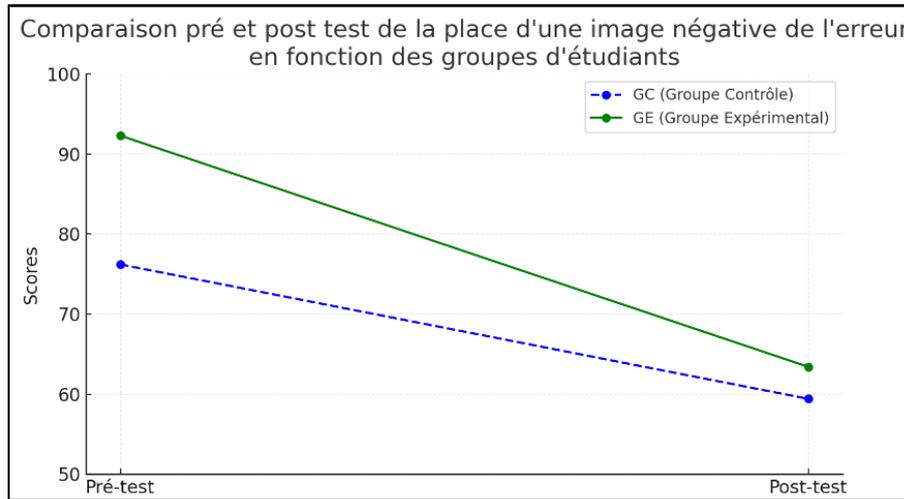
Note. \*  $p \leq 0,05$  pour les différences pré et posttest d'un même groupe.

L'évaluation formative est toujours décrite comme plus informative pour l'enseignant que pour l'élève, quel que soit le groupe de participants. Par rapport au prétest, elle est moins définie comme un « test à blanc ». En effet, pour les étudiants du groupe contrôle aucune définition ne mobilise cette idée et cet écart pré-posttest de 19,0 % est significatif ( $\chi^2 (1, N = 49) = 5,80, p = 0,016$ ). Pour le groupe expérimental, l'évolution est plus forte (- 26,4 %) et significative ( $\chi^2 (1, N = 59) = 5,57, p = 0,018$ ). L'écart entre les réponses des deux groupes en posttest est significatif ( $\chi^2 (1, N = 61) = 3,63, p = 0,057$ ) alors qu'il ne l'était pas en prétest ( $\chi^2 (1, N = 47) = 2,09, p = 0,148$ ).

Dans les définitions d'évaluation formative, la perspective de régulation est significativement moins présente dans les réponses du groupe contrôle en posttest par rapport au prétest ( $\chi^2 (1, N = 49) = 3,67, p = 0,055$ ) alors qu'elle n'a pas significativement évolué pour le groupe

expérimental ( $\chi^2 (1, N = 59) = 0,83, p = 0,361$ ). En posttest, l'écart entre les deux groupes n'est pas significatif ( $\chi^2 (1, N = 61) = 0,76, p = 0,384$ ).

L'évolution de l'image de l'erreur est illustrée par le graphique suivant (Figure 5).



**Figure 5 :** Évolution des pourcentages de réponses attribuant une image négative de l'erreur en fonction du groupe d'étudiants

Alors que 86,4 % des réponses associaient l'erreur à l'échec en prétest, cette proportion d'image négative est réduite de 16,8 % entre le pré et le posttest dans les réponses du groupe contrôlé et de 28,9 % pour le groupe expérimental ( $\chi^2 (1, N = 67) = 7,01, p = 0,008$ ).

### 6.2.2 Degré d'adhésion aux fonctions normative, formative et certificative

L'objectif des analyses suivantes est de vérifier si les étudiants ayant participé à la tâche expérimentale présentent une adhésion différente de celle des étudiants du groupe contrôlé à propos de l'évaluation normative, formative et certificative critériée. Concernant l'évaluation normative, les résultats posttest du groupe contrôlé et du groupe expérimental ont des valeurs proches sur le plan descriptif et l'écart est non significatif ( $\chi^2 (2, N = 75) = 0,65, p = 0,722$ ). Le Tableau 7 rend compte des résultats des deux groupes entre le pré et posttest.

**Tableau 7 :** Niveaux d'adhésion à l'évaluation normative

	Pré GC	Post GC	Pré GE	Post GE
Adhésion forte	0,0 %	3,0 %	7,7 %	2,4 %
Adhésion moyenne	35,0 %	39,4 %	30,8 %	31,0 %
Adhésion faible	65,0 %	57,6 %	61,5 %	66,7 %

Note. \*  $p \leq 0,05$  pour les différences pré et posttest d'un même groupe.

Le Tableau 8 résume, quant à lui, l'évolution des fortes adhésions pour l'évaluation formative et certificative critériée en fonction des groupes d'étudiants. Des écarts descriptifs apparaissent sans qu'ils soient significatifs.

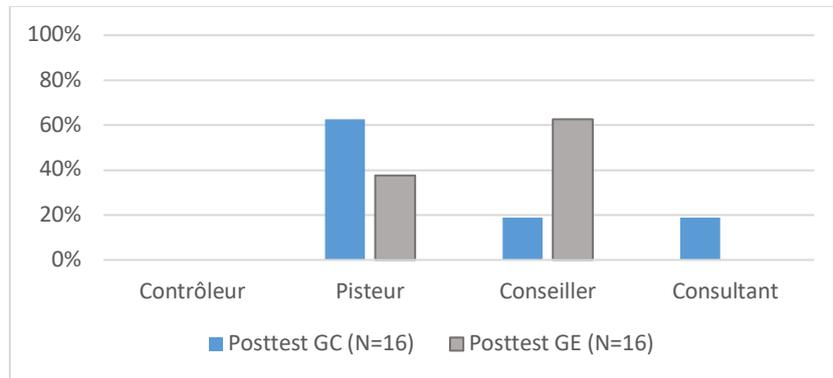
**Tableau 8** : Taux d'adhésion forte aux évaluations formative et certificative critériée en fonction des groupes d'étudiants

	Pré GC	Post GC	<i>p</i> value (pré-post)		Pré GE	Post GE	<i>p</i> value (pré-post)
Évaluation formative	45,0 %	57,6 %	0,374		50,0 %	57,1 %	0,565
Évaluation certificative critériée	60,0 %	72,7 %	0,382		50,0 %	57,1 %	0,405

### 6.2.3 Diversité des postures d'enseignant évaluateur au fil d'une séquence

Les analyses ont pour but de vérifier si la diversité des postures évaluatives choisies a évolué entre le pré et le posttest, et dans quelle mesure. Pour le groupe expérimental, le nombre d'étudiants déclarant une seule posture évaluative est significativement réduit entre le pré et le posttest, passant de 75,0 % à 41,0 % ( $\chi^2(1, N = 63) = 6,90, p = 0,009$ ). Cette évolution n'est pas significative pour le groupe contrôle (de 66,7 % à 57,1 % avec  $p = 0,518$ ). En posttest, la différence de réduction entre les deux groupes est significative : - 9,6 points pour le GC et - 34,0 points pour le GE ( $\chi^2(3, N = 107) = 7,91, p = 0,048$ ).

Le graphique ci-après (Figure 6) illustre la proportion de chaque métaphore dans les réponses ne mobilisant qu'une seule posture évaluative en posttest.



**Figure 6** : Répartition des choix de métaphores dans les réponses mobilisant une seule posture

Dans les réponses exprimant une posture unique, la métaphore de conseiller est majoritaire pour le groupe expérimental avec 62,5 % (18,8 % pour le groupe contrôle avec une différence significative  $\chi^2(9, N = 62) = 19,97, p = 0,018$ ). Parmi les réponses du groupe contrôle ne mobilisant qu'une seule posture, la métaphore majoritaire est celle du pisteur avec 62,5 % (37,5 % pour le groupe expérimental).

En prenant en compte toutes les réponses (une ou plusieurs postures exprimées), le Tableau 9 fait état des répartitions de choix de métaphore selon le groupe d'étudiants interrogés.

**Tableau 9** : Proportions de mobilisation des postures évaluatives en fonction des groupes

	Pré GC	Post GC	Pré GE	Post GE
Contrôleur	4,8	33,3**	23,1	38,1
Pisteur	47,6	33,3	30,8	40,5
Conseiller	47,6	33,3	73,1	42,9**
Consultant	33,3	18,2	15,4	31,0

Note. \* $p \leq .05$  \*\* $p \leq .01$  \*\*\* $p < .001$  pour les différences pré et posttest d'un même groupe

La répartition des choix de postures évaluatives semble plus équilibrée en posttest pour les deux groupes d'étudiants. Comme en prétest, la posture de contrôleur apparaît dans les réponses mobilisant plusieurs postures, et elle est majoritaire pour la phase de production finale alors qu'elle est absente des réponses mobilisant une posture unique.

Les mesures d'association réalisées deux à deux entre les variables du réseau conceptuel « évaluation-sanction », par le biais de coefficients de Cramer, montrent une relation entre les variables mesure et erreur négative. Sans que ces résultats ne soient statistiquement significatifs, la force d'association est modérée pour le groupe expérimental et concerne 11,1 % des répondants de ce groupe ( $V$  de Cramer = 0,337,  $p = 0,129$ ), et faible pour le groupe contrôle et concerne 21,9 % des étudiants de ce groupe ( $V$  de Cramer = 0,243,  $p = 0,390$ ). En outre, la relation entre la vision négative de l'erreur et une adhésion à l'évaluation normative apparaît comme forte et significative pour le groupe expérimental et concerne 26,8 % des participants ( $\chi^2$  (2,  $N = 41$ ) = 6,87,  $p = 0,032$ ,  $V$  de Cramer = 0,410). Pour le groupe contrôle, cette association est modérée et non significative mais concerne tout de même 28 % des participants ( $p = 0,366$ ,  $V$  de Cramer = 0,259). Les forces d'association des autres paires de variables sont faibles à très faibles et sans significativité statistique.

En appliquant le même traitement statistique aux variables du réseau « évaluation-soutien », il ne ressort pas de lien statistiquement significatif entre elles dans les réponses du groupe contrôle. Toutefois, l'association entre une vision positive de l'erreur et l'adhésion à l'évaluation formative est modérée et concerne 21,9 % des étudiants de ce groupe ( $V$  de Cramer = 0,248,  $p = 0,373$ ). Celle-ci n'apparaît pas dans les réponses du groupe expérimental ( $V$  de Cramer = 0,006,  $p = 0,970$ ). Au sein des définitions du groupe expérimental, il y a une force d'association moyenne à tendance significative entre la mobilisation des critères d'évaluation et la visée de régulation, sans que ces variables ne soient prédictives l'une de l'autre ( $\chi^2$  (1,  $N = 36$ ) = 3,20,  $p = 0,073$ ,  $V$  de Cramer = 0,298 et  $R^2$  ajusté = 0,062). Il existe aussi une association modérée entre le fait de mobiliser la posture de conseiller et la notion de critères dans la définition de l'évaluation ( $\chi^2$  (1,  $N = 36$ ) = 4,42  $p = 0,035$ ,  $V$  de Cramer = 0,351 et  $R^2$  ajusté = 0,097). Toujours pour le groupe expérimental, une association forte réside entre l'adhésion à l'évaluation formative et celle à l'évaluation certificative critériée, et 36 % de la variance du score d'adhésion à l'une de ces deux formes d'évaluation est expliquée par l'autre ( $\chi^2$  (1,  $N = 42$ ) = 15,68,  $p < 0,001$ ,  $V$  de Cramer = 0,611 et  $R^2$  ajusté = 0,358). Ainsi, les étudiants ayant réalisé la tâche expérimentale et adhérant à l'évaluation formative ont davantage tendance à être aussi en accord avec les principes de l'évaluation certificative critériée. Cette relation est plus forte qu'en prétest (27 %) pour le groupe expérimental mais reste faible et sans significativité statistique pour le groupe contrôle ( $V$  de Cramer = 0,209,  $p = 0,486$ ).

### 6.3. Discussion de l'étude 2

L'hypothèse d'un effet spécifique de la comparaison analogique fondée sur des textes métaphoriques par rapport à celle fondée sur un texte académique n'a pas pu être vérifiée. En dépit de différences ponctuelles qui ont pu être observées selon la condition, les résultats montrent une évolution des conceptions des étudiants à propos de l'évaluation et de l'erreur qui ne diffère pas significativement entre la lecture d'un texte académique (groupe contrôle) et celle de textes métaphoriques (groupe expérimental).

Concernant les définitions de l'évaluation et de l'évaluation formative, les différences entre les deux groupes au prétest ne sont pas négligeables. Bien que les participants soient répartis aléatoirement, des écarts peuvent être observés lorsque l'échantillon est de petite taille. Ils sont liés notamment à des différences de connaissances, de compétences et de motivation. Il ressort tout de même que, dans les définitions de l'évaluation, les concepts mobilisés en prétest étaient majoritairement en lien avec l'évaluation certificative et que cela semble évoluer dans le sens des contenus de formation : réduction significative des éléments relatifs à la mesure et la note dans le cas du groupe expérimental, et augmentation significative des éléments relatifs à la régulation dans les deux groupes. Cette augmentation est plus prononcée dans les réponses posttest du groupe contrôle. Cette différence pourrait être expliquée par le fait que le terme régulation soit présent 13 fois dans le texte de la tâche contrôle alors qu'il ne l'est qu'une seule fois dans ceux de la tâche expérimentale. Du point de vue du modèle de prévalence, ces évolutions pourraient être exprimées ainsi : les définitions de l'évaluation formulées par les deux groupes donnent à voir une inhibition des concepts de classement et d'épreuve et une prévalence de celui de régulation.

Les résultats ne montrent pas d'évolution quant à l'adhésion à l'évaluation normative et ce pour les deux groupes. Cela contraste avec l'absence d'éléments liés à la référence normative (classement) dans les définitions des étudiants. Cela pourrait être expliqué par l'inhibition du terme de classement dans l'expression libre, liée à la désirabilité sociale. Selon la littérature, des croyances ancrées, provenant d'expériences préalables, et centrales dans le système de croyances de l'individu seraient plus difficiles à faire évoluer (Crahay *et al.*, 2010).

La tâche expérimentale basée sur des textes métaphoriques ne semble pas avoir eu un effet différent sur les définitions de l'évaluation formative des étudiants par rapport à la comparaison de textes académiques. En effet, l'idée de « test à blanc » est significativement moins mobilisée dans les réponses des deux groupes en posttest. Le fait que l'évaluation formative puisse être considérée comme un « test à blanc » était un élément identifié pour expliquer le recours moins prégnant à cette évaluation (Goffin *et al.*, 2014). Cela pourrait donc être une première étape vers un changement des pratiques évaluatives des étudiants participant à cette recherche. L'échec a, quant à lui, moins de prévalence pour définir l'erreur dans les deux groupes. Cette évolution est plus forte pour la condition expérimentale. La prévalence de l'erreur comme outil pédagogique est essentielle, car la conception à propos de l'erreur influence les pratiques pédagogiques et celles-ci ont un impact sur la gestion émotionnelle et cognitive de l'erreur chez l'élève (Gagné *et al.*, 2023).

A la suite de la formation, le recours à une seule posture évaluative a significativement diminué pour le groupe expérimental. Cela pourrait être dû à une meilleure compréhension des rôles de l'enseignant et de l'élève au sein de chacune des postures et donc un développement conceptuel à propos de l'agir évaluatif. Le fait que les textes métaphoriques analysés mettent concrètement en exergue la pratique enseignante en exposant trois postures

différentes même si elles n'étaient pas nommées a pu avoir une incidence. L'abstraction des propriétés des postures évaluatives serait alors soutenue par le rattachement à l'expérience concrète (Sander, 2018). De surcroît, l'observation de certaines réponses d'étudiants lors de la tâche expérimentale permet de constater que plusieurs utilisaient les métaphores des postures évaluatives pour comparer le rôle de l'enseignant dans les trois textes. Il serait opportun de réaliser une analyse systématique de ces réponses. En outre, les étudiants du groupe expérimental semblent donner davantage de place à l'élève en contexte d'évaluation, puisqu'ils privilégient la posture de conseiller par rapport à celle de pisteur, majoritairement choisie par le groupe contrôle.

Bien que les étudiants du groupe expérimental semblent faire davantage de liens entre les concepts du champ de l'évaluation à la suite de la formation, cela reste très sporadique. Ce résultat rejoint les constats de Dupuch *et al.* (2015) : les relations d'inclusion de classe entre les concepts d'un même champ sont majoritairement inconscientes pour les apprenants et c'est encore plus prégnant quand un même terme désigne la classe et la sous-classe (Dupuch et Sander, 2007), comme ici avec le terme « évaluation » qui désigne le champ conceptuel et l'un des concepts. Par ailleurs, la robustesse de certaines conceptions, et donc la plus grande résistance au changement conceptuel, serait causée notamment par l'appartenance de ces concepts à même réseau conceptuel (Chi, 2008). Ils s'influenceraient mutuellement. Ce pourrait être le cas pour les concepts du champ de l'évaluation, mobilisés dans cette recherche.

## 7. Conclusion

L'évaluation est un aspect fondamental du métier d'enseignant, avec une influence réciproque entre pratique et conception. Cette recherche explore les effets d'un dispositif de formation basé sur la comparaison analogique, conçu pour aider les futurs enseignants du secondaire à comprendre et s'appropriier les contenus de formation à propos de l'évaluation et de l'erreur.

La première étude identifie les conceptions des étudiants. Bien qu'exprimant une forte adhésion à l'évaluation formative, lorsque les étudiants définissent eux-mêmes l'évaluation, celle-ci est principalement associée à l'évaluation certificative (note, mesure) en négligeant des éléments comme la régulation, essentielle à l'évaluation formative. Les expériences éducatives passées ayant un fort impact émotionnel – comme c'est le cas avec l'évaluation certificative – engendrent des croyances qui évoluent plus difficilement que d'autres (Mansfield et Volet, 2010). La prévalence de la définition de l'évaluation comme une évaluation certificative chez les étudiants interrogés pourrait refléter un tel phénomène.

Par ailleurs, l'adhésion à l'évaluation normative, synonyme de classement entre élèves, est faible, ce qui coïncide avec le peu de mention du concept de classement dans les définitions produites, tandis que l'évaluation critériée est bien intégrée avec l'emploi prégnant du concept de critère. Ces résultats, encourageants car s'éloignant du paradigme de l'évaluation normative, ne peuvent être exempts de marque de désirabilité sociale liée au contexte de formation dans lequel évoluent les étudiants. Il est également opportun de remarquer que les réponses montrent très peu de liens entre des concepts appartenant pourtant au même réseau sémantique (évaluation, erreur, critères, régulation, postures, etc.), ce qui permet d'inférer que ces relations sont encore implicites. Quant à l'erreur, sa définition mobilise largement l'échec, ce qui argumente en faveur d'une vision certificative de l'évaluation, centrée sur la dichotomie « réussite-échec ». En outre, les résultats font état de peu de diversité dans le

choix des postures évaluatives proposées, ce qui contraste avec les données recueillies auprès de futurs enseignants de primaire de Genève (Gvozdic *et al.*, sous presse).

La seconde étude vise à étudier dans quelle mesure la pratique de comparaison analogique de textes métaphoriques (condition expérimentale) stimule le changement conceptuel des participants à propos de l'erreur et de l'évaluation par rapport à une tâche similaire avec des textes académiques (condition contrôle). Pour les deux conditions, une augmentation de la mobilisation du concept de régulation dans les définitions de l'évaluation est constatée ainsi qu'une régression de l'image négative de l'erreur. Ces constats rejoignent les contenus de formation visant à concevoir l'évaluation comme soutien à l'apprentissage. Les résultats montrent que la réduction de l'association erreur-échec est plus forte chez les étudiants ayant analysé les textes métaphoriques. Il en est de même pour le recul des concepts de note et de mesure dans les définitions formulées et pour l'augmentation de la diversité des postures évaluatives choisies. Si ces évolutions montrent que la formation, et en particulier l'analyse de textes métaphoriques illustrant les pratiques d'évaluation formative, normative et certificative critériée, permet de soutenir le changement conceptuel des étudiants à propos de l'erreur et de l'évaluation, il s'agit de rester prudents dans les interprétations : un biais de désirabilité sociale (répondre au questionnaire dans le sens des contenus dispensés en formation) n'est pas à exclure. Ensuite, les évolutions de croyances à la suite d'une intervention de formation ne sont pas nécessairement pérennes, et certaines recherches longitudinales constatent un « effet à rebours » avec l'entrée dans le métier (Boraita et Crahay, 2013). Proposer une nouvelle fois le questionnaire, quelques temps après la prise de fonction des participants, permettrait d'observer si les changements constatés perdurent. Du reste, il serait pertinent d'étudier si le fait de dénommer les types d'évaluations illustrées par les textes métaphoriques influence les inférences des étudiants et leur changement conceptuel (Yamauchi et Markman, 2000).

À la lumière des résultats de l'étude 2 montrant une évolution des conceptions assez similaire dans les deux conditions (contrôle et expérimentale), une interrogation méthodologique émerge : peut-on réellement qualifier le groupe ayant travaillé sur le texte académique de « groupe contrôle », ou ne constituerait-il pas un groupe expérimental alternatif ? En effet, bien que les tâches de comparaison analogiques proposées aux deux groupes reposent sur des textes différents (métaphoriques versus académiques), elles partagent toutes deux une forme de réflexion structurée. Une direction méthodologique intéressante pourrait consister à reproduire cette étude en intégrant une configuration à trois groupes : deux groupes expérimentaux (chacun bénéficiant d'une tâche de comparaison analogique basée sur des textes différents, comme dans la présente étude) et un groupe contrôle qui ne bénéficierait d'aucune tâche de comparaison analogique.

Dans une perspective de formation, ces résultats invitent, au vu de la complémentarité potentielle des apports, à associer une analyse de textes métaphoriques avec comparaison analogique à l'étude de textes académiques. De plus, il serait intéressant d'étendre les recherches sur le changement conceptuel à propos de l'évaluation en questionnant les liens entre les concepts de l'évaluation et leur influence chez les futurs enseignants. En effet, les concepts d'un même champ n'étant pas indépendants, les relations qu'ils entretiennent font partie des connaissances à acquérir (Dupuch et Sander, 2007). Ainsi, soutenir les étudiants dans la construction d'un réseau sémantique à propos de l'évaluation pourrait être un levier à leur changement conceptuel.

## 8. Références bibliographiques

- Alfieri, L., Nokes-Malach, T. J. et Schunn, C. D. (2013). Learning Through Case Comparisons: A Meta-Analytic Review. *Educational Psychologist*, 48(2), 87-113. <https://doi.org/10.1080/00461520.2013.775712>
- Allal, L. (2008). Évaluation des apprentissages. In A. van Zanten (Ed.), *Dictionnaire de l'éducation* (p. 311-314). Presses Universitaires de France.
- Allal, L. et Mottier Lopez, L. (2005). L'évaluation formative de l'apprentissage : revue de publications en langue française. In OCDE, *L'évaluation formative : pour un meilleur apprentissage dans les classes secondaires* (pp. 265-290). Éditions OCDE/CERI.
- Antibi, A. (2003). *La constante macabre ou comment a-t-on découragé des générations d'élèves*. Math'Adore.
- Astolfi, J.-P. (2015). *L'erreur, un outil pour enseigner (12e ed.)*. ESF éditeur.
- Bachelard, G. (1993). *La formation de l'esprit scientifique*. Vrin.
- Bloom, B. S., Hastings, T. et Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. McGraw-Hill.
- Boraita, F. et Crahay, M. (2013). Les croyances des futurs enseignants ? est-il possible de les faire évoluer en cours de formation initiales et, si oui, comment ? *Revue française de pédagogie. Recherches en éducation*, 183, 99-158. <https://doi.org/10.4000/rfp.4186>
- Bucheton, D., Soulé, Y. (2009). Les gestes professionnels et le jeu des postures de l'enseignant dans la classe : un multi-agenda de préoccupations enchâssées. *Education et didactique*, 3(3), 29-48. <https://doi.org/10.4000/educationdidactique.543>
- Bullough, R. V. Jr (1991). Exploring personal teaching metaphors in preservice teacher education. *Journal of teacher education*, 42(1), 43-51. <https://doi.org/10.1177/002248719104200107>
- Bullough Jr R.-V. & Stokes D.-K. (1994). Analyzing personal teaching metaphors in preservice teacher education as a means for encouraging professional development. *American Educational Research Journal*, 31(1), 197-224.
- Cardinet, J. (1985). Des instruments d'évaluation pour chaque fonction. *Mesure et évaluation en éducation*, 8(1-2), 45-118.
- Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. Dans R. Giere (dir.), *Cognitive models of science: Minnesota Studies in the Philosophy of Science* (p. 129-186). University of Minnesota Press.
- Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. Dans S. Vosniadou (dir.), *Handbook of research on conceptual change* (p. 61-82). Hillsdale, NJ: Erlbaum.
- Cnesco (2023). *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts*. Cnesco-Cnam.
- Collins, A. M. et Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, 8(2), 240-247.
- Crahay, M., Issaieva, E. et Marbaise, C. (2013). *Conceptions de l'évaluation et principes de justice chez des enseignants primaires en Fédération Wallonie-Bruxelles*. Actes du 25e colloque de l'ADMEE-Europe. Évaluation et autoévaluation, quels espaces de formation, Fribourg, Suisse. [http://www.admee2013.ch/ADMEE-2013/7\\_files/Crahay-Issaieva-Marbaise-ADMEE-2013.pdf](http://www.admee2013.ch/ADMEE-2013/7_files/Crahay-Issaieva-Marbaise-ADMEE-2013.pdf)

- Crahay, M., Mottier Lopez, L. et Marcoux, G. (2019). L'évaluation des élèves : Docteur Jekyll and Mister Hyde de l'enseignement. Dans M. Crahay (dir.), *Peut-on lutter contre l'échec scolaire ?* (p. 358-425). De Boeck.
- Crahay, M., Wanlin, P., Issaieva, E. et Laduron, I. (2010). Fonctions, structuration et évolution des croyances (et connaissances) des enseignants. *Revue française de pédagogie*, 172, 85-129. <https://doi.org/10.4000/rfp.2296>
- De Ketele, J. M. (2001). Les concepts d'ami critique et de maillage au cœur de l'accompagnement, du suivi et de l'évaluation. *Bulletin Psychologique*, 3, 72-82.
- Dupuch, L., Bosc-Miné, C. et Sander, E. (2015). Évaluer les connaissances à l'école élémentaire à partir d'un réseau d'inclusion de classes. *Psychologie française*, 60(1), 17-34.
- Dupuch, L. et Sander, E. (2007). Apport pour les apprentissages de l'explicitation des relations d'inclusion de classes. *L'Année psychologique*, 107(4), 565-596.
- 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. <https://doi.org/10.7202/1041002ar>
- Gagné, P.-P., Leblanc, N., Rousseau, A., Lachaux, J.-P. et Mazeau, M. (2023). *Moi je fais de l'oupsologie ! Pour une pédagogie de l'erreur*. Cheneliere Mcgraw-Hill.
- Gentner D., Holyoak K.-J. et Kokinov B. N. (Eds.) (2001). *The analogical mind: Perspectives from cognitive science*. MIT press.
- Girardet, C. (2017). *Explaining the evolution of vocational teachers' classroom management beliefs and practices during teacher education* [Phd Thesis]. Université de Genève.
- Girardet C. (2020). Origines et évolutions des croyances et pratiques des enseignant(e)s de la formation professionnelle en Suisse. *Revue internationale d'éducation de Sèvres*, 84, 135-143. <https://doi.org/10.4000/ries.9633>
- Goffin, C., Renson, J. M. et Fagnant, A. (2014). Comment les futurs enseignants du secondaire conçoivent-ils l'évaluation formative et ses difficultés de mise en œuvre ? In *Colloque international de l'Admée-Europe. Cultures et politiques de l'évaluation en éducation et en formation*.
- Gray, M. E. et Holyoak, K. J. (2021). Teaching by analogy: From theory to practice. *Mind, Brain, and Education*, 15(3), 250-263.
- Gvozdic, K., gir, S. et Tessaro, W. (sous presse). Chapitre 4. Les métaphores de l'agir évaluatif : diversité et évolution des postures. Dans E. Sander (dir.), *Les Métaphores pour l'Education* (p. 153-179). ISTE.
- Hadji, C. (2017). Savoir mettre en œuvre une évaluation constructive. *L'essentiel de la pédagogie*, 229-250.
- Hofstadter, D. et Sander, E. (2013). *L'Analogie, cœur de la pensée*. Odile Jacob.
- Issaieva, E., Yerly, G., Petkova, I., Marbaise, C. et Crahay, M. (2015). Conceptions et prises de position des enseignants face à l'évaluation scolaire dans quatre systèmes éducatifs : quel est le reflet des cultures et politiques évaluatives ? Dans L. Belair et P.-F. Coen (dir.), *Évaluation et auto-évaluation : quels espaces de formation ?* (p. 73-98). De Boeck Université.
- Jorro, A. (1996). Pour une culture plurielle de l'évaluation, entre usages et archétypes. *Mesure et évaluation en éducation*, 19(2), 5-21.
- Jorro, A. (2000). *L'enseignant et l'évaluation des gestes évaluatifs en question*. De Boeck Université.
- Jorro, A. (2006). Devenir un ami critique. Avec quelles compétences et quels gestes professionnels. *Mesure et évaluation en éducation*, 29(1), 31-44. <https://doi.org/10.7202/1086966ar>

- Jorro, A. (2013). L'accompagnement des enseignants dans l'activité évaluative face à des situations de production écrite. *Revue française de linguistique appliquée*, 13(1), 107-116. <https://doi.org/10.3917/rfla.181.0107>
- Kagan, D. M. (1992). Implication of research on teacher belief. *Educational psychologist*, 27(1), p. 65-90. [https://doi.org/10.1207/s15326985ep2701\\_6](https://doi.org/10.1207/s15326985ep2701_6)
- Kurtz, K. J., Miao, C. H. et Gentner, D. (2001). Learning by analogical bootstrapping. *The Journal of the Learning Sciences*, 10(4), 417-446.
- Lakoff, G. et Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Lautrey, J., Rémi-Giraud, S., Sander, E. et Tiberghien, A. (2008). *Les connaissances naïves*. Armand Colin.
- Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and instruction*, 11(4-5), 357-380.
- Loewenstein J., Thompson L. et Gentner D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review*, 6(4), 586-597. <https://doi.org/10.3758/bf03212967>
- Mansfield, C. et Volet, S. (2010). Developing beliefs about classroom motivation: Journeys of preservice teachers. *Teaching and Teacher Education*, 26(7), 1404-1415. <https://doi.org/10.1016/j.tate.2010.04.005>
- Morales Villabona, F., Pasquini, R. et Strehmel, B. (2023). Mettre en œuvre des pratiques d'évaluation sommative notée au service des apprentissages de tous les élèves. Une étude de cas à l'école secondaire. *Journal international de recherche en éducation et formation*, 9(2), 1-26. <https://doi.org/10.48782/e-jiref-9-2-1>
- Masson, S. (2012). Neuroéducation : mieux comprendre le cerveau pour mieux enseigner [éditorial]. *Neuroéducation*, 1(1), 3-4.
- Mottier Lopez, L. (2015). *Évaluations formative et certificative : enjeux pour l'enseignement*. De Boeck.
- Mottier Lopez, L. (2023). Quelles sont les nouvelles conceptions de l'évaluation en classe ? Vers des évaluations pour apprendre durablement. Dans *Conférence de consensus du Cnesco l'évaluation en classe, au service de l'apprentissage des élèves : Notes des experts* (p. 13-22). Cnesco-Cnam.
- Nersessian N. J. (1999). Model-based reasoning in conceptual change. Dans L. Magnani, N.-J. Nersessian et P. Thagard (dir.), *Model-based reasoning in scientific discovery* (p. 5-22). Springer.
- Patrick H. et Pintrich P.-R. (2001). Conceptual change in teachers' intuitive conceptions of learning, motivation, and instruction: The role of motivational and epistemological beliefs. Dans B. Torff et R.J. Sternberg (dir.), *Understanding and teaching the intuitive mind: Student and teacher learning* (p. 117-143). Lawrence Erlbaum.
- Perrenoud, P. (1998). L'évaluation des élèves. *De la fabrication de l'excellence à la régulation des apprentissages*. De Boeck.
- Potvin, P. (2013). Proposition for improving the classical models of conceptual change based on neuroeducational evidence: Conceptual prevalence. *Neuroeducation*, 1(2), 16-43. <https://doi.org/10.24046/neuroed.20130201.16>
- Potvin, P., Sauriol, É. et Riopel, M. (2015). Experimental evidence of the superiority of the prevalence model of conceptual change over the classical models and repetition. *Journal of Research in Science Teaching*, 52(8), 1082-1108. <https://doi.org/10.1002/tea.21235>
- Raths J. (2001). « Teachers' beliefs and teaching beliefs ». *Early Childhood Research and Practice*, 3(1), 385-391
- Roegiers, X. (2010). *L'école et l'évaluation*. De Boeck. <https://doi.org/10.3917/dbu.roegis.2010.01>
- Sander, E. (2000). *L'analogie, du naïf au créatif : analogie et catégorisation*. Editions l'Harmattan.

- Sander E. (2017). Les connaissances issues de la vie quotidienne et les apprentissages scolaires. Dans *Psychologie du Développement*, (p. 217-225). Elsevier Masson.
- Sander E. (2018). Transformer l'inconnu par le connu : Constructions et interventions analogiques pour les apprentissages scolaires. Dans J.-M. Barbier and M. Durand (dir.) *Représenter/ transformer. Débats en analyse des activités*. (p. 259-279). Éditions L'Harmattan.
- Sander, E., Gros, H., Gvozdic, K. et Scheibling-Sève, C. (2018). *Les neurosciences en éducation*. Retz.
- Sanders, D. P. et McCutcheon, G. (1986). The development of practical theories of teaching. *Journal of Curriculum and Supervision*, 2(1), 50-67.
- Soussi, A., Ducrey, F., Ferrez, E., Nidegger, C., Viry, G. (2006). *EVALEPCOPO – Pratiques d'évaluation : ce qu'en disent les enseignants*. Service de la recherche en éducation, Genève. [https://www.syneval.ch/database/pdf/Soussi\\_et\\_al\\_2006\\_Evalepcopo.pdf](https://www.syneval.ch/database/pdf/Soussi_et_al_2006_Evalepcopo.pdf)
- Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: the componential analysis of human abilities*. Lawrence Erlbaum.
- Thibodeau, P. H., Crow, L. et Flusberg, S. J. (2017). The metaphor police: A case study of the role of metaphor in explanation. *Psychonomic bulletin & review*, 24(5), 1375-1386. <https://doi.org/10.3758/s13423-016-1192-5>
- Thibodeau, P. H. et Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. *PloS one*, 6(2), e16782. <https://doi.org/10.1371/journal.pone.0016782>
- Vosniadou S. (2003). Exploring the relationships between conceptual change and intentional learning. Dans G.-M. Sinatra et P.-R. Pintrich (dir.), *Intentional Conceptual Change* (p. 407-428). Lawrence Erlbaum Associates, Inc.
- Willame, B. et Snauwaert, P. (2018). Entraînement au contrôle inhibiteur et apprentissage en chimie dans le secondaire supérieur : favoriser un changement de prévalence conceptuelle. *Neuroéducation*, 5(2), 73-92. <https://doi.org/10.24046/neuroed.20180502.73>
- Yamauchi, T. et Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 776.

## 9. Annexes

### 9.1. *Annexe A : Questionnaire complet utilisé en pré et posttest*

---

Start of Block: Evaluer déf

Formuler une **définition** pour expliquer ce qu'"évaluer" veut dire. *"Evaluer c'est ..."*

---

---

---

---

---

End of Block: Evaluer déf

---

Start of Block: analogie évaluer

Formuler une **analogie** pour illustrer ce qu'"évaluer" veut dire. Par exemple, si une personne voulait expliquer comment fonctionne le cerveau, elle pourrait proposer l'analogie suivante "le cerveau c'est comme un ordinateur car il stocke les informations et on peut ensuite aller les chercher lorsqu'on en besoin". *"Evaluer c'est comme ..."*

---

---

---

End of Block: analogie évaluer

---

Start of Block: Crahay 2013



Exprimez votre degré d'accord avec les affirmations suivantes.	Pas du tout d'accord	Pas vraiment d'accord	Plutôt d'accord	Tout à fait d'accord
Les évaluations permettent de repérer les élèves qui ont besoin d'une explication supplémentaire	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lors des évaluations, les réponses des élèves permettent de cerner les progrès qu'ils ont réalisés depuis le début de l'année.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
L'analyse des résultats des élèves aux évaluations permet d'ajuster l'enseignement.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Les évaluations permettent de repérer ce qu'il faudrait retravailler avec les élèves.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Les évaluations permettent de savoir si je peux avancer dans la discipline ou si je dois reprendre certaines notions ou exercices.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Les résultats des élèves aux évaluations permettent d'identifier ceux qui n'ont pas les aptitudes naturelles.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Les évaluations permettent de sanctionner les élèves qui n'ont pas travaillé.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Les évaluations sont l'occasion pour les élèves de montrer ce qu'ils valent.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Les évaluations permettent de distinguer les bons et les mauvais élèves.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Les évaluations permettent de voir si les objectifs du programme sont atteints.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
En fin d'enseignement, il faut évaluer dans quelle mesure les élèves maîtrisent les compétences minimales.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
En fin d'enseignement, il faut évaluer les acquis des élèves, en évitant tout classement.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Les épreuves bilan permettent d'évaluer les effets de l'enseignement dispensé.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Les évaluations permettent de voir si les élèves maîtrisent le contenu de l'enseignement.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
L'évaluation est plus utile à l'enseignant qu'à l'élève.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

End of Block: crahay 2013

**Start of Block: Erreur analogieVSdef**

Formuler une **définition** pour expliquer ce qu'est une "**erreur**". *"Une erreur c'est ..."*

---

---

---

**End of Block: Erreur analogieVSdef**

---

**Start of Block: Analogie erreur**

Formuler une **analogie** pour illustrer ce qu'est une "**erreur**". Par exemple, si une personne voulait expliquer comment fonctionne le cerveau, elle pourrait proposer l'analogie suivante "le cerveau c'est comme un ordinateur car il stocke les informations et on peut ensuite aller les chercher lorsqu'on en a besoin". *"Une erreur c'est comme ..."*

---

---

---

**End of Block: Analogie erreur**

---

### Start of Block: postures évaluatives

**Dans votre classe...** Vous êtes titulaire d'une classe et avez prochainement planifié une séquence d'environ quatre semaines. En premier lieu, vous proposez aux élèves une *mise en situation* (S1), puis une *production initiale* (S2). Celle-ci sera suivie de différents *modules* (S3) qui permettront d'aborder les sous objectifs lors de tâches spécifiques. En fin de séquence, les élèves réaliseront une *tâche finale* (S4) que vous évaluerez par une note.

---



Voici les descriptions de différentes postures d'évaluateur-trice. Choisissez la-les posture(s) que vous adopteriez en cliquant sur sa description.

- En tant que **contrôleur-se**, j'effectue un jugement sur les performances des élèves en identifiant l'atteinte des objectifs. J'annote les productions en complément des appréciations.
- En tant que **pisteur-se**, j'anticipe les cheminements et suis la progression de l'élève. Je jalonne le parcours d'objectifs intermédiaires. J'encourage l'effort et j'instille le goût du savoir.
- En tant que **conseiller-ère**, j'analyse les tâches, je m'intéresse aux démarches des élèves et j'incite à l'auto-évaluation. J'adapte mes interventions aux besoins des élèves.
- En tant que **consultant-e**, je reste en retrait et j'instaure une communauté de pairs. Je favorise la parole des élèves et m'intéresse à leur rapport au savoir.
- Je n'adopterais aucune de ses postures.

End of Block: postures évaluatives

---

Start of Block: Postures vs situations

Display This Question:

*If Voici les descriptions de différentes postures d'évaluateur-trice. Choisissez la-les posture(s) q... , Je n'adopterais aucune de ses postures. Is Displayed*

*Carry Forward Selected Choices from "Voici les descriptions de différentes postures d'évaluateur-trice. Choisissez la-les posture(s) que vous adopteriez en cliquant sur sa description."*



**Vignette** Vous êtes titulaire d'une classe de 6P et avez prochainement planifié une séquence de production écrite en Français. Vous y traiterez le genre narratif pendant environ quatre semaines, en proposant aux élèves en premier lieu une *mise en situation* (S1), puis une *production initiale* (S2). Celle-ci sera suivie de différents *modules* (S3) qui permettront d'aborder les caractéristiques du genre dans des tâches spécifiques. En fin de séquence, les élèves produiront un *texte final* (S4) que vous évalueriez par une note. Vous avez choisi une ou plusieurs postures d'évaluateur-trice. Indiquez à quel(s) moment(s) de la séquence d'enseignement vous pensez les mettre en œuvre. (Une même posture peut être utilisée à plusieurs reprises.)

	S1	S2	S3	S4
En tant que <b>contrôleur-se</b> , j'effectue un jugement sur les performances des élèves en identifiant l'atteinte des objectifs. J'annote les productions en complément des appréciations.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
En tant que <b>pisteur-se</b> , j'anticipe les cheminements et suis la progression de l'élève. Je jalonne le parcours d'objectifs intermédiaires. J'encourage l'effort et j'instille le goût du savoir.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
En tant que <b>conseiller-ère</b> , j'analyse les tâches, je m'intéresse aux démarches des élèves et j'incite à l'auto-évaluation. J'adapte mes interventions aux besoins des élèves.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
En tant que <b>consultant-e</b> , je reste en retrait et j'instaure une communauté de pairs. Je favorise la parole des élèves et m'intéresse à leur rapport au savoir.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Je n'adopterais aucune de ses postures.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

End of Block: Postures vs situations

---

Start of Block: Evaluation formative analogie

Veillez formuler une **analogie** pour illustrer la notion d'**évaluation formative**.

---

---

---

End of Block: Evaluation formative analogie

---

Start of Block: Eval form def

Veillez formuler une **définition** pour expliquer ce qu'est l'**évaluation formative**.

---

---

---

End of Block: Eval form def

---

## 9.2. *Annexe B : Textes métaphoriques, supports de la tâche expérimentale*

### **Texte 1**

La conception de Camille à propos de l'évaluation peut être comparée à la navigation assistée par GPS lors d'un voyage en voiture. Voici comment cette analogie fonctionne.

Tout d'abord, vous déterminez votre destination finale, tout comme Camille fixe des objectifs d'apprentissage pour ses élèves.

Le GPS ne se contente pas de donner la distance à parcourir, mais il guide à chaque étape du voyage, en indiquant les virages, les rues à emprunter et les changements de direction. De la même façon, Camille fournit des feedbacks tout au long de l'apprentissage pour aider les élèves à atteindre leurs objectifs.

Si vous prenez un mauvais virage ou faites une erreur, le GPS recalcule automatiquement l'itinéraire pour vous ramener sur la bonne voie. C'est similaire à la régulation de l'enseignement en fonction des besoins des élèves, en fournissant des conseils et des feedbacks pour les aider à corriger leurs erreurs.

Le GPS fournit des informations en temps réel sur la progression, y compris le temps estimé pour atteindre la destination. De même, l'évaluation construite par Camille lui permet de suivre en temps réel les progrès des élèves et de leur donner une idée de l'endroit où ils/elles en sont par rapport à leurs objectifs d'apprentissage.

Le/la conducteur-riche a la responsabilité de rentrer dans le GPS les caractéristiques du trajet (éviter les autoroutes, faire au plus court...) et de suivre les indications du GPS tout en faisant les ajustements nécessaires en cours de route pour atteindre la destination. De la même manière, les élèves de Camille participent à la construction de la grille de critères d'évaluation et sont responsables de leur propre évaluation en comparant leur travail à ces critères tout en prenant en compte les feedbacks de Camille pour améliorer leur production.

En fin de compte, l'évaluation conçue par Camille, comme le GPS, vise à guider les apprenant-es vers leur destination d'apprentissage tout en fournissant des ajustements en cours de route pour maximiser leur réussite. Camille les aide à prendre des décisions éclairées, à corriger leurs erreurs et à rester sur la voie de l'apprentissage.

## Texte 2

L'évaluation telle que la conçoit Andrea peut être illustrée en utilisant une analogie basée sur l'obtention d'un permis de conduire. Voici comment cela fonctionne. Imaginez une personne passant un examen pour obtenir un permis de conduire.

L'objectif principal est d'obtenir le permis de conduire, ce qui signifie que la personne doit passer un examen et démontrer qu'elle possède les compétences et les connaissances nécessaires pour conduire en toute sécurité. De la même manière, l'évaluation construite par Andrea a pour objectif de vérifier si ses élèves ont acquis un ensemble spécifique de compétences ou de connaissances.

L'examen pour obtenir le permis de conduire est standardisé et suit un ensemble de critères prédéfinis. C'est similaire à l'évaluation mise en œuvre par Andrea, qui est basée sur des critères établis.

Lors de l'examen pour le permis de conduire, le candidat doit démontrer sa capacité à conduire en respectant les règles de la route, à effectuer des manœuvres, etc. Dans l'évaluation d'Andrea, les élèves sont évalués sur leur capacité à mobiliser des compétences ou des connaissances spécifiques conformément aux critères établis par Andrea.

Le résultat de l'examen du permis de conduire est binaire : les candidat-es réussissent et obtiennent le permis, ou échouent et doivent repasser l'examen. De manière similaire, l'évaluation faite par Andrea peut aboutir à la validation d'une compétence ou au fait que des élèves n'aient pas atteint le niveau requis.

Une fois le permis de conduire obtenu, il sert de preuve que la personne a acquis les compétences nécessaires pour conduire légalement. De la même manière, une validation obtenue à la suite de l'évaluation construite par Andrea prouve que les élèves ont atteint un certain niveau de compétence ou de connaissance dans un domaine donné.

Cette analogie avec l'obtention d'un permis de conduire met en évidence le caractère décisif et la nature binaire de l'évaluation telle que la conçoit Andrea, où l'objectif principal est de déterminer si l'élève a atteint un certain niveau de compétence ou de connaissance.

## Texte 3

L'évaluation telle que la conçoit Alex peut être illustrée en faisant l'analogie avec une course automobile. Voici comment cela fonctionne : imaginez un circuit sur lequel des pilotes de différents niveaux de compétence s'affrontent.

Dans une course automobile, l'objectif est de déterminer qui est le/la plus rapide parmi les participant-es. C'est similaire à l'évaluation d'Alex, où l'objectif est de classer les élèves en fonction de leur performance par rapport au groupe de pairs.

À la fin de la course, il y a un classement en fonction du temps. Les personnes les plus rapides sont en tête du classement, tandis que les plus lentes sont classées plus bas. Dans l'évaluation d'Alex, les élèves sont classés en fonction de leurs résultats par rapport à la moyenne de leur groupe.

Les résultats des participant-es à la course forment une distribution, où la majorité se situe autour de la moyenne, une petite partie est au-dessus de la moyenne, et d'autres sont en dessous. De même, lors des évaluations d'Alex, la plupart des élèves obtiennent des résultats proches de la moyenne, tandis que certain-es se démarquent positivement et d'autres obtiennent des résultats inférieurs à la moyenne.

Les pilotes comparent leur classement avec d'autres adversaires pour voir où ils/elles se situent par rapport à la concurrence. Les élèves font de même en comparant leur classement ou leur note avec leurs camarades de classe.

L'évaluation telle que la conçoit Alex ne se concentre pas sur la performance individuelle absolue, mais sur la performance relative par rapport à un groupe donné. Cela signifie que, même si l'élève a acquis de solides compétences, il/elle peut être classé-e plus bas que d'autres s'il y a des élèves aux résultats exceptionnels dans le groupe classe.

Cette analogie avec une course automobile met en évidence le principe de l'évaluation selon Alex, où la performance d'un individu est évaluée en fonction de sa position relative par rapport à d'autres membres d'un groupe donné, plutôt que sur une échelle absolue de compétences.