

Comment appréhender les compétences évaluatives des enseignants ? Adaptation d'un outil au contexte de l'enseignement secondaire supérieur belge francophone

How to grasp teacher's Assessment Literacy? Adaptation of an instrument to the context of upper secondary school in the French speaking part of Belgium

Jean Baron – jean.baron@uliege.be – <https://orcid.org/0009-0001-5334-7978>

Annick Fagnant – afagnant@uliege.be – <https://orcid.org/0000-0003-1026-3480>

Université de Liège - Belgique

Pour citer cet article : Baron, J. et Fagnant, A. (2025). Comment appréhender les compétences évaluatives des enseignants ? Adaptation d'un outil au contexte de l'enseignement secondaire supérieur belge francophone. *Évaluer. Journal international de recherche en éducation et formation*, 11(1), 29-48. <https://doi.org/10.48782/e-jref-11-1-29>

Résumé

Le temps consacré aux évaluations occupe une place importante dans le temps scolaire. Selon Wiliam (2011, p.13), « l'intégration de l'évaluation dans l'enseignement pourrait avoir un effet important pour améliorer l'engagement des élèves et leurs apprentissages », mais encore faut-il que les enseignants disposent de compétences évaluatives appropriées (ce que d'aucuns qualifient d'*Assessment Literacy*).

Définie au départ comme un ensemble de « connaissances » que devraient posséder les enseignants pour mener à bien des évaluations en classe (Mertler, 2004), la façon dont l'*Assessment Literacy* a été conceptualisée a évolué au cours du temps (Pastore, 2023) pour être progressivement considérée comme une « pratique sociale et dynamique » qui dépend du contexte dans lequel elle est vécue (Willis et al., 2013).

L'article cherche tout d'abord à montrer comment les définitions de l'évaluation et de l'*Assessment Literacy* ont évolué conjointement, en partant d'une vision relativement « techniciste » de l'évaluation pour aller vers une vision plus « opérationnelle » d'une évaluation-soutien d'apprentissage. En parallèle, il décrit l'évolution des outils conçus par les chercheurs pour appréhender l'*Assessment Literacy* et présente un outil spécifique (ACAI, DeLuca et al., 2016a, b) qui répond davantage aux normes contemporaines de l'évaluation. L'article envisage alors la problématique de l'adaptation de cet outil au contexte de l'enseignement secondaire en Belgique francophone en interrogeant sa concordance avec le cadre décretaal et en envisageant les enjeux de traduction/contextualisation d'un questionnaire développé en langue anglaise et principalement utilisé en Amérique du Nord.

Mots-clés

Littératie en évaluation, Pratiques évaluatives des enseignants, Évaluation-soutien d'apprentissage, Belgique francophone, ACAI

Abstract

Time devoted to assessment is an important part of school time. According to Wiliam (2011, p.13), “integrating assessment with instruction may well have unprecedented power to increase student engagement and to improve learning outcomes” but teachers still need to have appropriate assessment skills and knowledge (what some refer to as Assessment Literacy).

Initially defined as a body of “knowledge” that teachers should possess in order to carry out assessments in the classroom (Mertler, 2004), how Assessment Literacy has been conceptualized has evolved over time (Pastore, 2023) to be gradually considered as a “social and dynamic practice” that depends on the context in which it is experienced (Willis et al., 2013).

The paper aims to show how the definitions of assessment and Assessment Literacy have evolved together, moving from a relatively “technical” view of assessment to a more “operational” view of assessment of learning. Simultaneously, the paper describes the evolution of the tools designed by researchers to measure Assessment Literacy and presents a specific tool (ACAI, DeLuca et al., 2016a, b) which better meet contemporary assessment standards. Then, the article considers the possible adaptation of this tool to the context of secondary education in the French-speaking part of Belgium by first questioning its concordance with the educational policy and then considering the issues of translation/contextualization of a questionnaire initially developed in English and mainly used in North America.

Keywords

Assessment Literacy, Teachers' approaches to classroom assessment, Assessment for Learning, French speaking part of Belgium, ACAI

1. Introduction

De nombreuses recherches mettent en évidence les effets positifs de pratiques évaluatives qui soutiennent l'apprentissage des élèves (voir Fagnant, 2023, pour une synthèse). Utilisées adéquatement, de telles pratiques fournissent des informations fiables sur l'apprentissage des élèves, permettent de communiquer plus efficacement les résultats et facilitent l'implication des élèves en tant que partenaires à part entière dans l'évaluation (DeMink-Carthew et al., 2021). Certains travaux montrent que, lorsque l'enseignant intègre pleinement l'évaluation à ses pratiques de classe, on peut constater « des gains significatifs dans la réussite des élèves, les fonctions métacognitives et la motivation à apprendre » (DeLuca et al. 2016a, p.252).

La littérature consultée met en évidence des lacunes au niveau des compétences évaluatives des enseignants débutants (Barnes et al., 2020 ; McGee et Colby, 2014 ; Oo et al., 2022 ; Starck et al., 2018). Certains travaux montrent aussi que l'expérience professionnelle semble avoir un effet positif sur ces pratiques (Unal et Unal., 2019) même si certaines difficultés semblent subsister (DeLuca et al., 2016a). D'autres travaux montrent, quant à eux, des impacts négatifs lors de l'entrée dans le métier, lorsque les jeunes enseignants cherchent à se conformer aux habitudes trop ancrées des écoles (Luisoni et Monnard, 2015).

Pour évoquer les habiletés, connaissances ou compétences requises pour mener des pratiques évaluatives efficaces, la littérature anglo-saxonne parle de *Literatie in Evaluation – Assessment Literacy* (ou AL dans la suite du texte), qui est généralement définie comme « un ensemble de connaissances et compétences utiles à l'enseignant pour mettre efficacement en place l'évaluation en classe » (Pastore, 2023, p.1). Cette définition, pour le moins généraliste et qui a évolué depuis ses prémices, sera précisée ultérieurement. Depuis les premiers travaux de Plake et al. (1993), différents outils ont été conçus pour appréhender l'AL, que ce soit auprès d'enseignants ou de futurs enseignants en formation initiale.

Dans sa revue systématique à propos de l'AL, Pastore (2023) rappelle l'importance d'une bonne compréhension de ce concept ainsi que le besoin d'identifier la façon dont les enseignants intègrent les conceptions contemporaines de l'évaluation au sein de leurs pratiques de classe. Le premier enjeu de cet article s'inscrit dans cette dynamique en proposant, à notre connaissance pour la première fois en langue française, une analyse évolutive des différentes conceptualisations de l'AL et des différents outils conçus par les chercheurs pour l'appréhender. Ce premier enjeu conduit à justifier le choix de l'outil retenu - à savoir l'outil ACAI (*Approaches Classroom Assessment Inventory*, DeLuca et al., 2016b). Aucune recherche appuyée sur cet outil n'ayant été menée dans le contexte éducatif de la Belgique francophone (Fédération Wallonie-Bruxelles ou FWB), le second enjeu de l'article est alors celui de son adaptation à ce contexte. Après avoir vérifié l'adéquation de l'outil avec les normes régissant le système éducatif, la problématique de sa traduction et de son adaptation au contexte spécifique de l'enseignement secondaire en FWB est envisagée. Enfin, la question de son adéquation aux représentations et pratiques évaluatives des enseignants est analysée au départ de focus groups, ayant eux-mêmes conduit à de nouvelles adaptations de l'outil.

2. L'évolution de la conceptualisation de l'*Assessment Literacy* et des outils permettant de l'appréhender

La conceptualisation de l'AL a évolué au cours du temps (Pastore, 2023) passant d'une vision plus « techniciste » de l'évaluation - et, conjointement, d'une définition de l'AL influencée par l'idée qu'il convient d'atteindre un certain nombre de « *standards* » et donc, de disposer pour ce faire de « connaissances » spécifiques – à une vision plus « opérationnelle » d'une évaluation soutien d'apprentissage – liée quant à elle à une vision de l'AL davantage orientée vers les « pratiques évaluatives » des enseignants. Le premier point tente de rendre compte de cette évolution. Parallèlement, les outils conçus par les chercheurs pour appréhender l'AL sont évidemment influencés par leur conceptualisation de celle-ci. Le second point passe en revue différents outils et leur évolution dans le temps, pour aboutir à l'outil ACAI (DeLuca et al., 2016b), outil que nous avons retenu et cherché à adapter au contexte de la FWB.

2.1. *Évolution de la conceptualisation de l'Assessment Literacy*

2.1.1 D'une vision assez « techniciste » de l'évaluation à une vision plus « opérationnelle » d'une évaluation soutien d'apprentissage

Les premiers travaux sur l'AL (Plake et al., 1993 ; Mertler, 2004) se réfèrent largement aux standards de l'*American Federation of Teachers – AFT* (1990, p.2) qui définit l'évaluation comme « un processus qui conduit à obtenir des informations utiles pour prendre des décisions éducatives à propos des progrès, des forces et faiblesses [des élèves] et pour juger de l'efficacité de l'enseignement reçu et de l'atteinte des objectifs visés ». Nous qualifierons cette vision de « techniciste » au sens où l'enjeu va alors être d'identifier les composantes de l'évaluation qui sont essentielles pour permettre aux enseignants de mettre en œuvre une pratique évaluative de qualité, telle qu'attendue par certaines normes évaluatives. Ainsi l'AFT (1990) précise que les enseignants doivent être capables de choisir des méthodes d'évaluation adaptées aux décisions pédagogiques ; d'élaborer des méthodes d'évaluation adaptées à ces décisions ; d'utiliser les résultats des évaluations pour prendre des décisions concernant les élèves et/ou pour planifier l'enseignement ; d'élaborer des procédures de notation (*grading procedure*) valables et de communiquer les résultats des évaluations aux élèves, aux parents et à d'autres publics. Ils doivent aussi être capables d'identifier les méthodes d'évaluation qui seraient contraires à l'éthique ou inappropriées.

Brookhart (2011) regrette la vision trop restreinte de l'évaluation telle que véhiculée par l'AFT (1990), celle-ci ne prenant pas suffisamment en compte les travaux récents en matière d'évaluation-soutien d'apprentissage que l'on peut définir comme une pratique qui « fait partie des pratiques quotidiennes des élèves, des enseignants et des pairs qui recherchent, réfléchissent et réagissent aux informations issues d'échanges, de démonstrations et d'observations afin de développer les apprentissages en cours » (Klenowski, 2009, p.264).

Cette évolution de la conceptualisation de l'évaluation se retrouve dans les différents travaux sur l'AL, qui vont cesser de se référer aux normes de l'AFT (1990) pour s'appuyer sur les normes plus contemporaines du *Joint Committee on Standards for Educational Evaluation* (JCSEE, Klinger et al., 2015 ; voir aussi Zaleski, 2015) qui définit notamment les enjeux suivants : définir les objectifs visés par l'évaluation et identifier comment utiliser les informations en lien pour soutenir les apprentissages ultérieurs ; identifier ce qui est évalué

et à quelles fins ; en fonction de ces éléments, déterminer la méthode la plus adéquate à mettre en œuvre en alignant les visées d'évaluations ; interroger la communication des informations aux personnes concernées et définir la façon d'y impliquer les élèves. Parallèlement à cette vision plus large de l'évaluation qui invite à s'intéresser davantage aux usages de celles-ci qu'aux fonctions traditionnelles qui la caractérisent (Mottier Lopez, 2023), le point suivant mettra en évidence une évolution dans la façon de conceptualiser l'AL.

2.1.2 D'une vérification des connaissances évaluatives des enseignants à une analyse de leurs pratiques évaluatives

C'est à Stiggins (1995) que l'on doit le concept d'AL, que l'on peut situer en lien avec une vision « techniciste » de l'évaluation, appuyée sur les recommandations de l'AFT (1990) précitées :

Il [Stiggins, 1995] a déclaré que les éducateurs compétents en matière d'évaluation (...) entrent dans le domaine en sachant ce qu'ils évaluent, pourquoi ils le font, comment évaluer au mieux les connaissances visées, comment générer de bons exemples permettant d'illustrer les performances des élèves... mais aussi ce qui peut potentiellement mal se passer avec l'évaluation et comment éviter que cela ne se produise. (...) (Mertler, 2004, p. 51).

Plus largement, l'AL est généralement définie comme une compétence comprenant des savoirs et savoir-faire concernant les pratiques évaluatives et les objectifs de l'évaluation (Brookhart, 2011; Levy-Vered et Alhija, 2015 ; Mertler, 2004). Pour Paterno (2001, cité par Mertler, 2004, p. 51), l'AL implique essentiellement de posséder des connaissances en matière évaluative : « des connaissances sur les principes de base d'une bonne pratique d'évaluation, y compris la terminologie, le développement et l'utilisation de méthodologies et de techniques d'évaluation » ; il s'agit de connaître « des normes de qualité en matière d'évaluation » ainsi que des « alternatives aux mesures traditionnelles de l'apprentissage ».

Cette façon de définir l'AL en lien avec les composantes essentielles d'une « bonne évaluation » est très présente dans la littérature (Pastore, 2023). On y distingue d'une part, ce qui a trait aux habiletés liées à la gestion des épreuves, à leur conception, leur administration et leur notation ; viennent ensuite l'interprétation et l'utilisation des résultats de l'évaluation pour soutenir la prise de décision pédagogique et l'apprentissage des élèves ; puis enfin des éléments relatifs à la façon de rendre compte et de communiquer les résultats des évaluations. Les études qui s'appuient sur ces définitions cherchent généralement à voir dans quelle mesure les (futurs) enseignants ont acquis les standards de la profession (cf. AFT, 1990, par exemple).

Pastore (2023) estime que deux phénomènes ont marqué conjointement un changement substantiel dans cette façon de définir l'AL : l'intérêt généralisé pour l'évaluation-soutien d'apprentissage et la reconnaissance du caractère socioculturel de l'évaluation. Ainsi, elle précise que la volonté d'intégrer davantage la vision de l'évaluation-soutien d'apprentissage conduit certains auteurs à définir l'AL comme étant « les connaissances et compétences requises pour évaluer et soutenir les apprentissages des élèves grâce à l'évaluation » (DeLuca et al., 2016b, p. 248, s'inspirant notamment des travaux de Brookhart, 2011). Le caractère socioculturel de l'évaluation se retrouve, quant à lui, assez nettement dans la définition de Willis et al. (2013) pour qui l'AL est une pratique sociale et dynamique qui dépend du contexte dans lequel elle est vécue. Dans ces définitions, l'évaluation joue un

rôle central pour soutenir les élèves dans l'atteinte des objectifs d'apprentissage. C'est donc de « pratiques évaluatives », terme emprunté à DeLuca et al. (2016b), dont il est question ici, et non plus seulement de « connaissances » en matière d'évaluation.

2.2. Évolution des outils permettant d'appréhender l'Assessment Literacy

2.2.1 Une première génération d'instruments basée sur les normes de l'AFT

Au cours des années 1990 et 2000, une série d'instruments ont été élaborés et testés par des chercheurs, essentiellement américains : le *Teacher Assessment Literacy Questionnaire* (TALQ, Plake et al. (1993) ; *l'Assessment Literacy Inventory* (ALI, Campbell et al., 2002, cité par Mertler, 2004) ; le *Classroom Assessment Literacy Inventory* (CALI, Mertler, 2004) ; le *Revised Assessment Literacy Inventory* (ALI, 2^{ème} version ; Mertler et Campbell, 2005). Ces divers instruments semblent tous constituer des modifications progressives de la version d'origine de Plake et al. (1993) ; ils déclarent tous s'appuyer sur les sept domaines de compétences identifiés par l'AFT (1990) ; ils sont tous organisés sous la forme d'analyses de cas ou de scénarios pédagogiques ; ils déclarent tous évaluer des connaissances et se présentent tous sous la forme de questions à choix multiple face auxquelles une seule réponse est attendue. Même si les indicateurs utilisés dans les recherches varient ou ne sont pas toujours précisés, les chercheurs pointent globalement une bonne consistance interne du questionnaire auprès de publics d'étudiants en formation (Mertler, 2004 ; Mertler et Campbell, 2005) mais nettement moins bonne auprès d'enseignants en exercice (Mertler, 2004 ; Plake et al., 1993).

Nous ne pousserons pas plus avant l'investigation de cette première génération d'outils (voir DeLuca et al., 2016a pour une synthèse), mais notons qu'ils soulèvent des interrogations à trois niveaux. Tout d'abord, ils s'appuient sur des normes évaluatives qui peuvent aujourd'hui être considérées comme dépassées, au sens où trop peu de place y est accordée à l'évaluation-soutien d'apprentissage (cf. critique de Brookhart, 2011). La vision assez « techniciste » de l'évaluation qui est à la base de ces outils conduit à évaluer essentiellement des « connaissances » en matière évaluative, pour vérifier si les enseignants ou futurs enseignants rencontrent les standards attendus (Pastore, 2023, évoque une « *professional standards perspective* », p. 18). Cette focale sur les « connaissances » conduit à proposer des scénarios face auxquels une seule réponse est correcte, ce qui place les enseignants dans une situation potentiellement désagréable (se sentir évalué comme des étudiants) et pourrait potentiellement expliquer les différences de consistance interne observées entre les deux types de publics (les futurs enseignants se prêtant peut-être plus volontiers au jeu d'une évaluation assez « scolaire » que les enseignants en exercice).

2.2.2 Une deuxième génération d'instruments basée sur des normes visant davantage à favoriser une évaluation-soutien d'apprentissage

Pour s'assurer de prendre en compte les principales normes et directives contemporaines à propos de l'évaluation telles que définies par le JCSEE (Klinger et al., 2015), DeLuca et al. (2016a) recensent et analysent les documents officiels de six pays et régions anglophones. C'est un total de 35 directives différentes qui sont mises en avant puis regroupées au départ en huit thèmes qu'ils synthétisent finalement en quatre dimensions pour construire leur questionnaire (DeLuca et al., 2016b) : a) Objectifs de l'évaluation ; b) Processus d'évaluation – dont les aspects liés à la communication des résultats – ; c) Équité dans l'évaluation – dont la place de l'éthique – ; et d) Respect de la fiabilité et de la validité au niveau des mesures.

Selon DeLuca et al. (2016a), les évolutions concernant les normes de référence sont importantes : alors que les documents plus anciens visaient « à aider les enseignants à sélectionner et à utiliser des évaluations, principalement sommatives et standardisées afin de prendre et de communiquer des décisions pédagogiques » (p. 262), c'est l'émergence de l'évaluation soutien d'apprentissage qui retient l'attention des auteurs dans plusieurs documents plus récents. A ce jour, les compétences requises concernant l'évaluation formative et diagnostique, l'auto-évaluation et l'évaluation par les pairs ou encore les feedbacks sont considérés comme essentiels. Souhaitant alors prendre en considération ces évolutions, DeLuca et al. (2016b) conçoivent un nouvel instrument de mesure qu'ils intitulent *Approaches to Classroom Assessment Inventory* (ACAI). L'outil mesure les « approches évaluatives des enseignants », autrement dit leurs « positionnements » par rapport à des situations évaluatives possibles, plutôt que « les connaissances » en évaluation. Cela correspond à une des évolutions de la façon de conceptualiser l'AL telle que mise en évidence par Pastore (2023).

Cette 2e génération d'outils, qui correspond aux normes d'évaluation plus contemporaines, présente aussi l'intérêt d'éviter l'évaluation directe de connaissances, ce qui nous paraît plus adapté pour s'adresser à des enseignants en exercice. Par ailleurs, il est présenté par ses auteurs comme un instrument pouvant soutenir le développement de l'AL auprès des enseignants. L'outil ACAI, retenu pour notre étude, est dès lors développé au point suivant.

3. L'outil ACAI et la problématique de son adaptation au contexte de l'enseignement secondaire supérieur en FWB

Les études menées au départ de l'outil ACAI (DeLuca et al., 2016b) se situent principalement en Amérique du Nord et, à notre connaissance, aucune utilisation de l'outil n'a été proposée dans le monde francophone et encore moins en FWB. Après avoir décrit l'outil ACAI et fait état des différentes versions de celui-ci au cours du temps, la question de sa possible adaptation au contexte de la FWB est traitée suivant un double enjeu : (1) vérifier dans quelle mesure l'outil s'accorde aux normes qui régissent le système éducatif en FWB ; (2) traduire et adapter les scénarios au contexte de la FWB, tout en s'assurant qu'ils soient en adéquation avec les représentations et pratiques évaluatives des enseignants dans ce contexte culturel spécifique.

3.1. Description de l'outil ACAI, de son évolution et des questionnements qui demeurent

L'outil ACAI (DeLuca, 2016b) est composé de trois parties : la première concerne les *pratiques évaluatives* ; la deuxième interroge les enseignants sur leur *sentiment d'efficacité personnelle* concernant leurs pratiques évaluatives et la troisième s'intéresse à leurs *souhaits d'apprentissages professionnels* à propos des pratiques évaluatives. L'objectif de cet article étant de présenter et de contextualiser l'outil permettant d'appréhender l'AL, nous nous intéresserons uniquement à la première partie.

En vue de réaliser une première validation de construit, l'instrument a été soumis à une dizaine d'experts nord-américains du domaine de l'évaluation ainsi qu'à une dizaine d'enseignants. Le questionnaire a ensuite été proposé à un échantillon de 404 enseignants canadiens, et des statistiques de mode ont été calculées pour voir comment les réponses se distribuaient parmi les différentes options proposées. Des tests Chi carré de Pearson ont été réalisés pour analyser d'éventuels patterns de réponses selon les caractéristiques des

répondants, mais aucune donnée indiquant la consistance interne du questionnaire n'est proposée pour cette partie du questionnaire. Tout comme dans les travaux précédents, les enseignants sont appelés à se positionner par rapport à des questionnements centrés sur des scénarios décrivant des situations de classe. Toutefois, dans le cas de l'ACAI, toutes les propositions traduisent des visions différentes de l'évaluation mais ne sont pas considérées comme correctes ou incorrectes.

L'outil ACAI a été utilisé dans une panoplie de recherches, en procédant à des utilisations variées et avec une implémentation complète ou partielle. Les recherches ont été principalement menées au Canada et aux Etats-Unis et la plupart du temps, en tout ou en partie, par l'équipe des auteurs. Dans les articles consultés, nous avons pu constater que les recherches ont concerné des enseignants en formation initiale (Coombs et al., 2019 ; DeLuca et al., 2019a ; DeLuca et al., 2019c) ; des enseignants venant de terminer leur formation initiale (DeLuca et al., 2019b) ; des enseignants en exercice (DeLuca et al., 2018 ; Coombs et al., 2021 ; DeLuca et al., 2021), de façon conjointe des enseignants en formation initiale et en exercice (DeLuca et al., 2016b ; Coombs et al., 2017 ; Barnes et al., 2020) ou encore des formateurs d'enseignants (Coombs, 2017).

Les cinq scénarios portent sur les thématiques suivantes : l'évaluation sommative, la notation, la différenciation, la planification de séquences de leçon et les évaluations standardisées. Pour chaque scénario, quatre séries de trois items sont proposées. Chacune interroge une des quatre dimensions identifiées par les auteurs (et rappelées ci-dessous) via trois items décrivant une approche évaluative spécifique (A, B ou C), comme présenté dans le Tableau 1. Pour chacun des 5 scénarios, il y donc *in fine* 12 pratiques évaluatives qui sont décrites.

Tableau 1 : Dimensions de l'AL et approches en lien (DeLuca et al., 2016b, p.252)

| Dimensions de l'AL | Approches de l'AL | | |
|---------------------------|--------------------------------------|--|--|
| | Approche A | Approche B | Approche C |
| Objectifs de l'évaluation | Évaluation <u>de</u> l'apprentissage | Évaluation <u>pour</u> l'apprentissage | Évaluation <u>en tant</u> qu'apprentissage |
| Processus de l'évaluation | Conception | Notation/Critères | Communication |
| Éthique de l'évaluation | Traitement standard | Équité | Différenciation |
| Théorie de la mesure | Fiabilité | Validité | Fiabilité et validité |

Selon DeLuca et al. (2016b), l'approche A est assimilée à un type d'évaluateurs qui se centrent sur l'évaluation sommative ; l'accent est mis sur la conception d'épreuves fiables et administrées de façon standardisée aux différents groupes d'élèves. L'approche B décrit des évaluateurs centrés sur des pratiques d'évaluation formative ; c'est la validité de l'épreuve qui est ici centrale et qui se traduit notamment par l'importance des critères et de l'usage qui sera fait des résultats pour soutenir l'apprentissage des élèves et les traiter de manière équitable. La dernière approche, C, décrit un type d'évaluateurs proposant une évaluation plus individualisée, tout en veillant aux principes de fiabilité et de validité. Dans cette dernière approche, la communication des résultats et le développement d'évaluations différenciées sont importants pour promouvoir une évaluation en tant qu'apprentissage.

Dans la première version du questionnaire (DeLuca et al., 2016b, DeLuca et al., 2018), le répondant sélectionne un item par série. Un profil lui est alors attribué (approche évaluative de type A, B ou C, cf. Tableau 1) selon le type de réponse qu'il produit le plus fréquemment. L'attribution des profils est donc issue d'une valeur modale, calculée pour chacune des dimensions évaluées. Cette façon de procéder peut paraître imprécise. En effet, si un répondant choisit systématiquement la même ligne de conduite (par ex. A pour « les buts de l'évaluation » dans les 5 scénarios), il est catégorisé comme tel. De même, un répondant choisissant à 3 reprises la modalité [A] et à 2 reprises la modalité [B] sera catégorisé « A » également ; tout comme un enseignant qui panacherait en proposant 3 réponses [A], 1 [B] et 1 [C]. Pour nuancer davantage les classements, différentes recherches (Barnes et al., 2020 ; Coombs et al., 2017 ; DeLuca et al. 2019a), vont faire apparaître des catégories hybrides dans l'analyse des résultats, distinguant des profils purs (par ex. exclusivité ou très large majorité de A) et des profils mixtes (par ex. combinaison de A et B). Assez logiquement, cette nouvelle façon de construire les profils donne des résultats très contrastés par rapport à ceux des travaux précédents, rendant difficile toute comparaison entre les études menées.

On constate par la suite une modification importante concernant le mode de réponse : les répondants ne doivent plus choisir une action pédagogique parmi trois propositions mais sont invités à se positionner par rapport à chacun des 60 items (*i.e.* 12 items pour 5 scénarios) via une échelle de Likert à six niveaux. Il n'est donc plus question de prioriser l'action privilégiée mais plutôt d'indiquer dans quelle mesure le répondant est susceptible d'effectuer chacune d'elles. A partir de ce moment-là, les auteurs ne déterminent plus des approches globales (A, B, C) ou mixtes, mais calculent une valeur moyenne, au travers des scénarios toujours, pour chacune des 12 « pratiques évaluatives » répertoriées dans le Tableau 1. Cette nouvelle version a été utilisée dans les études ultérieures (DeLuca et al., 2019b ; DeLuca et al., 2019c ; Coombs et al., 2019 ; Coombs et al., 2021 ; DeLuca et al., 2021) ; elle permet une présentation plus nuancée des résultats puisque les enseignants peuvent se montrer plus ou moins en accord avec chacune des « pratiques évaluatives » plutôt que d'être catégorisé dans une approche.

Parmi les différents articles consultés, la seule recherche pour laquelle les indices de consistance interne des 12 « pratiques évaluatives » (3 approches x 4 dimensions) sont présentés est celle menée par DeLuca et al. (2021) auprès d'enseignants en exercice dans trois pays (Canada, Chine et Etats-Unis). Les indices (alpha de Cronbach) s'étendent de 0.46 à 0.68 et sont donc à prendre avec réserve. Leurs valeurs assez faibles pourraient être liées au nombre réduit d'items par indicateur (5 items, un seul par scénario) et/ou au fait que les enseignants avec de l'expérience professionnelle pourraient faire des distinctions fines selon les scénarios, ce qui rendrait leurs réponses moins cohérentes par indicateur puisque ces derniers sont construits au travers des scénarios. A notre connaissance, il n'y a pas eu de vérification de l'existence de ces dimensions via des analyses factorielles.

3.2. L'outil ACAI est-il en cohérence avec les normes légales en FWB ?

Pour utiliser le questionnaire dans le contexte de la FWB, il convient tout d'abord de vérifier l'adéquation avec les prescrits concernant les pratiques d'évaluation. Nous nous sommes basés sur le Décret Missions (Communauté française, 1997.07.24) définissant les missions prioritaires de l'enseignement, le Décret concernant la formation initiale des enseignants de FWB (Communauté française, 2022a) ainsi que sur deux documents de

référence à propos de l'évaluation rédigés par un réseau libre subventionné (FESeC, 2021) et par un réseau officiel (FWB, 2019a).

Le Décret Missions (Communauté française, 1997.07.24, Article 15) précise que les établissements d'enseignement doivent permettre aux élèves de « progresser à leur rythme en pratiquant l'évaluation formative et la pédagogie différenciée ». De son côté, le décret consacré à la formation initiale des enseignants (Communauté française, 2022a, Article 5 du Chapitre II, § 1^{er}) prévoit que le futur enseignant soit notamment capable d'« agir comme pédagogue au sein de la classe et au sein de l'établissement scolaire dans une perspective collective, notamment au travers de :

- la *construction et l'utilisation* de supports *d'observation et d'évaluation*, cette dernière étant spécifiquement à visée *compréhensive et formative*, favorisant la *responsabilisation* et la *participation de l'élève* dans ses apprentissages ;
- la conception et la mise en œuvre de pratiques de différenciation pédagogique, *d'accompagnement personnalisé* des élèves *tenant compte de leurs acquis antérieurs*, de leur profil d'apprenant et, le cas échéant, de *leurs besoins spécifiques* impliquant la mise en œuvre *d'aménagements raisonnables* et reposant notamment sur le co-enseignement ou la co-intervention pédagogique. »

On identifie très clairement la volonté du législateur de responsabiliser l'élève dans son évaluation, ce qui rejoint, selon nous, les lignes directrices du JCSEE (Klinger et al., 2015) mais aussi l'idée d'évaluation « en tant qu'apprentissage » donnant une place active à l'élève.

Sur une des pages Internet du site officiel de la FWB¹, il est précisé ce qui suit dans un bref passage intitulé « Nouvelle approche de l'évaluation » :

La manière d'envisager les besoins de l'élève s'inscrit dans une approche évolutive des difficultés d'apprentissage dans laquelle les évaluations formatives sont encouragées. L'évaluation combine bienveillance et exigence, elle repose sur une observation attentive au cours de laquelle le ou les enseignants identifient les acquis et les progrès des élèves, comme leurs difficultés. Point essentiel : les élèves en difficulté font l'objet d'un suivi immédiat et le soutien qui leur est apporté est adapté à leurs besoins.

Ces éléments témoignent de la volonté d'intégrer l'évaluation formative dans les situations d'enseignement-apprentissage. Le site ne renvoie pas vers d'autres ressources sur le sujet, mais d'autres éléments mentionnés ailleurs sur ce site officiel² font référence à la volonté de soutenir et accompagner les élèves de manière plus personnalisée.

Ce faisant, le soutien et l'accompagnement s'envisagent de manière plus « personnalisée » pour les élèves, au plus près de leurs besoins, mais aussi de manière plus collective pour l'enseignant, qui partage avec d'autres membres de l'équipe pédagogique et éducative la prise en charge du soutien apporté.

En Ontario, région d'origine des auteurs du questionnaire ACAI, les enseignants disposaient, dès 2010, d'un document de référence leur précisant les lignes de conduite (Ministry of Education Ontario, 2010). En FWB, les réseaux d'enseignement publient

¹ <http://www.enseignement.be/index.php?page=28590>

² <http://www.enseignement.be/index.php?page=28609&navi=4935>

également des recommandations et/ou directives dont certains éléments sont présentés ci-dessous pour un réseau libre subventionné (FESeC, 2021) et un réseau officiel (FWB, 2019a).

Que le processus d'évaluation soit vu comme « la partie la plus délicate de l'acte d'enseigner » pour FWB (2019a, p.1) ou encore comme « un des outils les plus puissants dont disposent les enseignants pour améliorer les apprentissages de leurs élèves » (FESeC, 2021, p.5), chacun s'accorde sur la place à lui réserver au cœur des pratiques d'enseignement-apprentissage. Dans les documents officiels francophones, tout comme pour Brookhart (2011) et le *JCSEE* (2015), on constate que l'évaluation n'est pas une fin en soi mais plutôt une étape de l'activité d'enseignement-apprentissage.

3.3. Traduction et adaptation de l'outil ACAI au contexte de la FWB

L'outil ACAI (DeLuca, 2016b) a été traduit et contextualisé tout en veillant à rester le plus proche possible de l'outil initial afin de permettre les comparaisons statistiques ultérieures avec les autres recherches. Notons toutefois que nous avons d'emblée posé le choix d'une version dans laquelle les enseignants se situeraient par rapport à chacun des items, comme proposé dans les versions les plus récentes.

3.3.1 Procédure de traduction

Pour la traduction proprement dite, alors que certains auteurs (Beaton et al., 2000 ; Epstein et al. 2015) recommandent de mettre en place une étape de rétro-traduction, d'autres (Berh et Shishido, 2009) précisent que, quelle que soit la façon de procéder, il faut garder à l'esprit que la traduction doit être un processus de décision multidimensionnel. Berh et Shishido (2009) rappellent l'importance de la production de traductions « en parallèle », c'est-à-dire produites indépendamment. Une première traduction du questionnaire a ainsi été réalisée en parallèle par deux chercheurs (le premier auteur de l'article et un collègue travaillant dans le champ de l'évaluation dans une autre université), puis les deux traductions ont été confrontées et une première version cible a été retenue.

Par ailleurs, Beaton et al. (2000) et Epstein et al. (2015) recommandent de faire appel à des comités d'experts ou encore à des entretiens de débriefing cognitif. Ces deux étapes ont également été mises en œuvre. Tout d'abord la version cible issue de la traduction croisée a été revue et commentée indépendamment par trois chercheurs (le 2^e auteur de cet article et deux chercheurs, issus de deux autres universités, dont les recherches portent notamment sur les pratiques d'évaluation) puis une 2^e version cible a été retenue par le 1^{er} auteur de l'article, après prise en compte et arbitrage des commentaires proposés par les trois chercheurs. Par la suite, des entretiens de débriefing cognitifs ont été menés sous la forme de focus groups, comme expliqué plus loin.

3.3.2 Adaptations liées au contexte spécifique de la FWB

Pour rendre compte de ces adaptations, nous prendrons un exemple ciblé sur un scénario. Dans la version d'origine, les auteurs évoquent des élèves qui bénéficient de « plans individuels d'apprentissage ». A l'heure actuelle, cette terminologie correspond à un cas de figure qui concerne le début de l'enseignement secondaire³, mais elle ne correspond pas à une situation que peuvent rencontrer les enseignants du secondaire supérieur (grades 10-12,

³ <http://www.enseignement.be/index.php?page=25664>

public cible de notre étude). Un autre dispositif de suivi des élèves en difficulté d'apprentissage est actuellement progressivement mis en place dans les écoles primaires (le dossier d'accompagnement de l'élève ou DAcCE⁴). Ce type de dispositif sera progressivement mis en place dans les classes de grades 10 à 12 en FWB à partir de l'année scolaire 2029-2030 (Communauté française, 2022b), mais il ne correspond donc pas encore à la réalité quotidienne des enseignants interrogés. Pour adapter le scénario au contexte ciblé par notre étude, nous avons donc décidé de faire le lien avec les aménagements dits « raisonnables » qui, eux, sont déjà bien d'actualité. En effet, un élément spécifique du contexte de la FWB concerne la prise en compte des élèves « à besoins spécifiques » et les nécessaires adaptations auxquelles sont tenus les enseignants (Communauté française, 2018 ; Fédération Wallonie-Bruxelles, 2018). Le scénario a dès lors été adapté pour tenir compte du contexte spécifique de l'étude (à titre illustratif, la version traduite et adaptée de cette situation se trouve en annexe 6). Notons que c'est la linéarité de l'écriture qui nous conduit à évoquer les adaptations contextuelles après la problématique de la traduction, mais les deux processus sont évidemment menés conjointement.

3.3.3 Adaptation pour prendre en compte les représentations et pratiques évaluatives des enseignants : organisation de focus groups

Après les étapes de traduction/adaptation des questionnaires au contexte de la FWB, des entretiens de débriefings cognitifs ont été réalisés sous la forme de focus groups, de façon à s'assurer de l'adéquation du questionnaire aux représentations et pratiques évaluatives des enseignants. En accord avec les écrits de Schelings (2021), de trois à cinq groupes semblent constituer un nombre adéquat. Ainsi, 4 focus groups ont été menés par le 1^{er} auteur de l'article avec des enseignantes du secondaire supérieur de FWB.

Le recrutement a été effectué par bouche à oreille au départ de l'entourage du premier auteur. Chaque groupe comptait entre 2 et 4 enseignants qui ont ainsi eu l'occasion de donner leur avis sur le questionnaire. Les enseignants ne sont pas représentatifs des enseignants du contexte de FWB, mais ils ont des profils relativement diversifiés, même si l'on peut regretter l'absence de certaines disciplines telles que les sciences dites « dures » ou un moindre nombre d'enseignants de cours à visée professionnalisante. Concrètement, le groupe est uniquement composé de femmes enseignant dans le secondaire supérieur (grades 10, 11 et 12). Deux tiers enseignent dans l'enseignement de transition et les autres dans l'enseignement de qualification. Elles sont sept participantes avec plus de 15 ans d'ancienneté. Huit d'entre elles font partie du réseau libre (Segec) et quatre du réseau officiel (WBE). Différentes disciplines sont représentées : sciences géographiques (1), histoire (1), langue française (2), langues modernes (2), psychologie (1), religion (2) sciences économiques et sociales (1), sciences sociales (1) et soins animaliers (1).

Ces focus groups, qui ont duré chacun entre deux heures et deux heures et demie, ont conduit à procéder à des adaptations plus ou moins importantes du questionnaire. Par exemple, même s'il était précisé d'emblée que les enseignants devaient répondre aux items en fonction de leur contexte de classe, une modification de forme a été effectuée afin de mettre en évidence cet élément dans les consignes, car au moins une enseignante de chaque groupe a souhaité une clarification à ce sujet après avoir traité la première situation. Certains items ont été simplifiés parce qu'ils proposaient plusieurs informations en même

⁴ <http://www.enseignement.be/index.php?page=28609&navi=4935>

temps ou qu'ils étaient formulés au mode conditionnel, ce qui dans les deux cas rendait complexe le positionnement des répondants.

Des modifications plus importantes ont été envisagées, comme ajouter une modalité de réponse, supprimer un scénario, ne pas placer les items par groupes de trois ou encore ajouter un espace libre pour insérer des commentaires⁵. Premièrement, la modalité « Je ne sais pas » a été ajoutée suite à la suggestion de plusieurs groupes, certaines enseignantes éprouvant parfois des difficultés à se positionner face à un item et trouvant dommage de devoir forcer un choix. Ce n'était pas nécessairement les mêmes items qui questionnaient ainsi les différentes enseignantes, et la solution n'était dès lors pas d'amender l'item lui-même. Par ailleurs, même si elles n'arrivaient pas à se positionner face à certains items, on peut aussi considérer l'intérêt de les faire réfléchir à leurs pratiques évaluatives voire de bousculer leurs représentations. Deuxièmement, la cinquième situation en lien avec les épreuves standardisées a été supprimée parce que plusieurs participantes ont signalé ne pas avoir vécu ce type d'épreuves et éprouver dès lors des difficultés à lui donner du sens dans leurs pratiques professionnelles⁶. Le facteur « temps » était également à prendre en compte car le questionnaire était assez long, ce qui a renforcé le choix de supprimer un scénario. Troisièmement, une enseignante a signalé avoir été interpellée par, selon elle, « une gradation entre les items de chaque série de trois ». Comme précisé précédemment, suite aux premières versions de l'ACAI (DeLuca et al., 2016b) où les répondants devaient choisir une action parmi trois propositions pour les 4 dimensions investiguées, le questionnaire conservait cette organisation des réponses par « paquets » sans que cela ne prenne sens pour les répondants puisqu'ils devaient se positionner sur l'ensemble des 12 propositions pour chaque scénario. Nous les avons donc replacés sans catégorie, les uns à la suite des autres, comme dans l'exemple repris en annexe. Enfin, dans les quatre groupes, les enseignantes ont regretté l'absence d'un espace où partager, si besoin, leur façon « personnelle » de réagir aux situations. Cet espace a été intégré.

Il a également été mis en avant, à différentes reprises, que le contexte de classe (par exemple le nombre d'élèves, le type de discipline enseignée...) avait une influence sur la façon de gérer l'épreuve évaluative et donc de se positionner face aux affirmations proposées. D'une façon générale, les focus groups ont mis en avant d'éventuelles difficultés pratiques au niveau de la gestion des interactions avec « chaque » élève, comme proposé dans le questionnaire d'origine. Par exemple, l'enseignant n'a pas l'opportunité de mener des discussions évaluatives avec « chaque » élève si sa classe compte 28 élèves alors que ce serait faisable avec une plus petite classe. La formulation de plusieurs items avec ce « chaque » semblait ainsi gêner plusieurs enseignantes, qui s'accordaient avec l'idée de s'intéresser à « tous » les élèves de la classe mais sans nécessairement pouvoir individualiser complètement les interventions. Elles éprouvaient alors des difficultés à se positionner sur

⁵ Ces quatre modifications ont été décidées suite aux focus groups, qui étaient eux-mêmes basés sur une traduction/adaptation de la version du questionnaire d'origine (DeLuca et al., 2016b). Il est toutefois important de préciser qu'elles ont également été effectuées dans la version en ligne de l'équipe de DeLuca et al. (s.d.) disponible à l'adresse suivante : <https://interceptum.com/si/en/4800045>. Elles semblent aussi être prises en compte dans l'article de DeLuca et al. (2021) qui fait notamment écho de l'usage de quatre situations, sans toutefois préciser lesquelles, ni les raisons de cette suppression et des autres modifications apportées au questionnaire.

⁶ En FWB et depuis l'année scolaire 2010-2011, les élèves de 6ème année (dernière des années de scolarité obligatoire) sont tenus de présenter des épreuves standardisées qui prennent la forme d'épreuves externes certificatives (CESS). Celles-ci sont communes à tous les établissements mais actuellement, seules les disciplines d'histoire et de français sont présentées.

l'échelle de Likert car elles avaient envie de répondre qu'elles agiraient très probablement... plus ou moins de cette façon. Dès lors, lorsque c'était possible, sans dénaturer complètement le sens de la proposition, nous avons remplacé le terme « chaque » par « tous », en précisant, par exemple, l'idée de favoriser les échanges avec « tous » les élèves de la classe. Dans d'autres cas, même si certaines propositions étaient considérées comme peu faisables voire peu souhaitables (comme par exemple modifier le contenu des épreuves en fonction des compétences des élèves), elles ont été conservées car les enseignantes n'éprouvaient pas de difficulté à se positionner sur l'échelle de Likert en indiquant simplement ici que c'était peu probable qu'elles agissent de la sorte. D'une façon générale, les adaptations proposées visaient ainsi à permettre aux enseignants de se positionner plus aisément face aux propositions, que ce soit pour considérer que cela correspondait à leurs pratiques évaluatives ou au contraire que cela n'en faisait pas partie.

Enfin, concernant leur ressenti général face au questionnaire, les enseignantes des quatre groupes ont qualifié le questionnaire de « normal », de « comme les autres »... ce qui pourrait laisser entendre qu'il pouvait faire écho à leurs représentations en matière d'évaluation, tant au niveau des scénarios retenus que des propositions face auxquels se positionner. Plusieurs ont aussi pointé l'intérêt de la démarche pour questionner leurs pratiques de classe. Dans l'un des groupes, une enseignante a ajouté qu'en « discuter avec les collègues pousse à un éventuel changement des pratiques ».

4. Discussion et perspectives

Le premier enjeu de cet article était de proposer une analyse évolutive des différentes conceptualisations de l'AL et de différents outils conçus par les chercheurs pour l'appréhender. En nous appuyant majoritairement sur des écrits anglo-saxons, nous avons ainsi tenté de rendre compte de la façon dont le concept d'*Assessment Literacy* (AL) a évolué depuis ses prémisses (Pastore, 2023), passant d'une définition plus « techniciste » de l'évaluation visant à vérifier si les enseignants disposent des connaissances correspondant aux standards répertoriés dans les documents officiels (Stiggins, 1995 ; Mertler, 2004) à une vision plus « opérationnelle », dynamique et contextualisée des pratiques évaluatives visant à soutenir l'apprentissage des élèves (Willis et al., 2013 ; Xu et Brown, 2016). Au vu de l'évolution de ce concept, mais également en lien avec l'évolution des standards de la profession (AFT, 1990 ; *JCSEE*, Klinger et al., 2015), les outils de mesure de l'AL ont assez logiquement subi eux aussi une évolution importante, comme nous avons cherché à l'expliquer pour justifier le choix de l'outil retenu.

L'outil ACAI (DeLuca et al., 2016b) a été utilisé dans une variété de recherches récentes, menées dans des contextes différents mais jamais, à notre connaissance, dans un contexte francophone et encore moins en FWB. La problématique de traduction/adaptation à ce contexte spécifique constituait dès lors le deuxième enjeu de cet article.

L'analyse comparative des éléments du *JCSEE* (Klinger et al., 2015) et d'éléments de référentiels francophones a permis de constater la convergence des directives quant à la place de l'évaluation dans les pratiques éducatives, ce qui conforte la possibilité d'utiliser et d'adapter l'instrument au contexte de la FWB. En effet, la plupart des principes directeurs déclinés par Brookhart (2011) tout comme ceux proposés par le *JCSEE* (Klinger et al., 2015) sont présents dans les normes du contexte éducatif de la FWB. Ainsi, les différentes fonctions de l'évaluation sont mises en avant, tout comme la demande de veiller à intégrer

l'évaluation au sein des situations d'enseignement-apprentissage pour prendre les décisions adaptées en termes d'accompagnement pédagogique.

Le questionnaire ACAI a été traduit et adapté afin, d'une part, que les situations évoquées correspondent pleinement au contexte de la FWB et, d'autre part, répondent aux constats mis en avant lors des focus groups avec les enseignants. Concernant les adaptations contextuelles, il nous paraît intéressant de revenir sur l'une des recommandations du *JCSEE* (Klinger et al., 2015) concernant la volonté de différenciation dans les évaluations, de façon à tenir compte des *diversités culturelle et linguistique* ou encore des *besoins spécifiques* d'apprentissages. En FWB, les référentiels et autres documents officiels prévoient la mise en œuvre des pratiques différenciées et inclusives. Cependant, mis à part des conditions de passation relevant d'aménagements « raisonnables » (pour une définition des aménagements raisonnables, voir l'Article 1.1.3.1-1. du Décret du 19.09.2019), il n'est pas demandé de différencier les évaluations. D'ailleurs, le décret précise que « les aménagements d'ordre pédagogique ne remettent pas en cause les objectifs d'apprentissage définis par les référentiels. » « FWB (2019b, Article 1.7.8-1., § 4). Autrement dit, même si des adaptations ont été proposées pour que les scénarios rencontrent des réalités auxquelles les enseignants du secondaire supérieur en FWB pourraient être confrontés, des éléments relevant du contexte institutionnel risquent aussi d'affecter les réponses fournies à certains items. La prise en compte du contexte spécifique de l'étude affecte ainsi non seulement le processus d'adaptation du questionnaire, mais aussi l'analyse et l'interprétation ultérieures des résultats. En plus du contexte institutionnel, les focus groups ont, par ailleurs, mis en évidence que le contexte spécifique de classe pouvait également affecter la façon dont les enseignants géraient une épreuve évaluative. En ce sens, et malgré les adaptations proposées pour aider les enseignants à pouvoir se positionner face aux différents items sur l'échelle de Likert (que ce soit positivement ou négativement), les réponses fournies par les uns et les autres semblent logiquement affectées par le contexte dans lequel ils enseignent. Ainsi, les focus groups ont mis en évidence une possible tension entre les conceptions évaluatives (qui pourraient par exemple soutenir une large implication de l'élève dans l'évaluation) et les pratiques évaluatives (qui pourraient rendre complexe la mise en œuvre de certaines actions avec lesquelles il y aurait pourtant accord). A ce stade, la question de savoir dans quelle mesure ces tensions vont se traduire par des profils de répondants différents (ce qui serait un résultat intéressant) ou risquent d'affecter la validité interne du questionnaire (les enseignants risquant de répondre en suivant des logiques différentes selon les items) reste en suspens et devra faire l'objet des analyses de validité ultérieures.

La contextualisation d'une enquête ne se limite pas aux tâches à réaliser en amont ; elle impacte également les questions de validation psychométrique de l'instrument et d'interprétation des résultats. Rappelons à ce propos que la question de la validité interne du questionnaire de l'ACAI questionne, et ceci même dans la version d'origine. L'outil est conçu de façon à distinguer trois profils d'évaluateurs qui répondraient, en théorie, de façon différenciée aux quatre dimensions mesurées (pour rappel : objectifs de l'évaluation, processus, éthique et théorie de la mesure). Les premières versions du questionnaire (DeLuca et al., 2016b) proposaient ainsi, pour chaque dimension, une série de trois items (correspondant aux trois profils présumés) parmi lesquels les répondants devaient d'abord choisir (DeLuca et al., 2018 ; DeLuca et al. 2019a), puis se positionner sur une échelle de Likert (DeLuca et al, 2019b, c ; DeLuca et al., 2021). Dans le premier cas, un profil (approches A, B, C) était attribué en fonction de valeurs modales ; dans le second cas, c'est la moyenne des items traduisant une même approche qui était calculée pour chacune des 12 approches évaluatives. Ces approches évaluatives sont construites sur une base théorique

solide mais sont peu vérifiés empiriquement. Les indicateurs sont établis au travers des scénarios proposés et, non seulement leur consistance interne n'est pas très élevée (cf. DeLuca et al., 2021), mais aucune analyse factorielle ne semble non plus en avoir attesté l'existence, permettant notamment de démontrer qu'ils transcenderaient les scénarios.

La version adaptée de l'ACAI a été soumise entre les mois de février et de mai 2024 à près de 100 enseignants exerçant dans le secondaire supérieur en FWB. Les premiers résultats montrent que la consistance interne de plusieurs des 12 indicateurs établis par DeLuca et al. (2016b ; 2021) est assez faible. Si l'on ne peut rejeter de possibles problèmes liés aux choix qui ont été posés en termes de traduction/adaptation du questionnaire, on peut aussi questionner la structure théorique du questionnaire d'origine. Lors des focus groups, les enseignantes ont mis en évidence l'impact que leur contexte de classe pouvait avoir sur leur façon de gérer l'évaluation et donc de se positionner face aux différents items du questionnaire. Les situations décrites dans les scénarios sont assez différentes les unes des autres et, même si les quatre scénarios retenus peuvent faire écho à leurs représentations, on peut faire l'hypothèse que ces situations elles-mêmes pourraient influencer les pratiques évaluatives des enseignants. Pour le dire autrement, il se pourrait que, selon la situation décrite dans le scénario, le contexte spécifique de classe influence de façon différentielle les réponses fournies par les enseignants et que, dès lors, les profils d'évaluateurs varient selon les scénarios plutôt que de les « transcender ». Indéniablement, des analyses complémentaires (notamment des analyses factorielles et/ou en composantes principales, Berger, 2021) devront être menées pour mieux comprendre la structure interne du questionnaire, étape préalable pour rendre compte des pratiques évaluatives des enseignants du secondaire supérieur en FWB.

5. Références bibliographiques

- American Federation of Teachers, National Council on Measurement in Education & National Education Association. (1990). Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice*, 9(4), 30-32. <https://eric.ed.gov/?id=ED323186>
- Barnes, N., Gareis, C., DeLuca, C., Coombs, A. et Uchiyama, K. (2020). Exploring the roles of coursework and field experience in teacher candidates' assessment literacy: A focus on approaches to assessment. *Assessment Matters*, 14, 5-41.
- Beaton, D. E., Bombardier, C., Guillemin, F. et Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186-3191. <https://doi.org/10.1097/00007632-200012150-00014>
- Behr, D., et Shishido, K. (2016). The translation of measurement instruments for cross-cultural surveys. *The SAGE handbook of survey methodology*, 55, 269-87.
- Berger, J. L. (2021). Analyse factorielle exploratoire et analyse en composantes principales: guide pratique. <https://hal.archives-ouvertes.fr/hal-03436771v2>
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: issues and practice*, 30(1), 3-12.
- Communauté française (1997, 24 juillet). *Décret définissant les missions prioritaires de l'enseignement fondamental et de l'enseignement secondaire et organisant les structures propres à les atteindre*. https://www.galilex.cfwb.be/document/pdf/21557_021.pdf
- Communauté française (2018). *Décret relatif à l'accueil, à l'accompagnement et au maintien dans l'enseignement ordinaire fondamental et secondaire des élèves présentant des besoins spécifiques*. Docu 44807. https://www.galilex.cfwb.be/document/pdf/44807_000.pdf
- Communauté française (2022a). *Décret modifiant le décret du 7 février 2019 définissant la formation initiale des enseignants*. Docu 50119. https://www.galilex.cfwb.be/document/pdf/50119_000.pdf
- Communauté française (2022b). *Décret portant création du dossier d'accompagnement de l'élève (DACCE)*. Docu 50293. https://www.galilex.cfwb.be/document/pdf/50293_000.pdf
- Coombs, A. J. (2017). *Teacher educators' approaches to assessment* [Thèse de Doctorat, Queen's University].
- Coombs, A., DeLuca, C., et MacGregor, S. (2019). A person-centered analysis of teacher candidates' approaches to assessment. *Teaching and Teacher Education*, 87, 102952.
- Coombs, A., DeLuca, C., LaPointe-McEwan, D., et Chalas, A. (2017). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education*, 71, 134-144. <https://doi.org/10.1016/j.tate.2017.12.010>
- Coombs, A., Rickey, N., DeLuca, C., et Liu, S. (2021). Chinese teachers' approaches to classroom assessment. *Educational Research for Policy and Practice*, 21, 1-18. <https://doi.org/10.1007/s10671-020-09289-z>
- DeLuca, C., Coombs, A. et LaPointe-McEwan, D. (2019a). Assessment mindset: Exploring the relationship between teacher mindset and approaches to classroom assessment. *Studies in Educational Evaluation*, 61, 159-169. <https://doi.org/10.1016/j.stueduc.2019.03.012>
- DeLuca, C., Coombs, A., MacGregor, S. et Rasooli, A. (2019b). Toward a differential and situated view of assessment literacy: Studying teachers' responses to classroom assessment scenarios. *Frontiers in Education*, 4(94). <https://doi.org/10.3389/feduc.2019.00094>

- DeLuca, C., Lapointe-McEwan, D. et Luhanga, U. (2016a). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), 251-272.
- DeLuca, C., LaPointe-McEwan, D. et Luhanga, U. (2016b). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, 21(4), 248-266. <https://doi.org/10.1080/10627197.2016.1236677>
- DeLuca, C., Rickey, N. et Coombs, A. (2021). Exploring assessment across cultures: Teachers' approaches to assessment in the US, China, and Canada. *Cogent Education*, 8(1), 1921903. <https://doi.org/10.1080/2331186X.2021.1921903>
- DeLuca, C., Schneider, C., Coombs, A., Pozas, M. et Rasooli, A. (2019c). A cross-cultural comparison of German and Canadian student teachers' assessment competence. *Assessment in Education: Principles, Policy & Practice*, 27(1), 26-45. <https://doi.org/10.1080/0969594X.2019.1703171>
- DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D. et Luhanga, U. (2018). Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice*, 25(4), 355-375. <https://doi.org/10.1080/0969594X.2016.1244514>
- DeMink-Carthew, J., Smith, C. P., Tinkler, A. S. et Tinkler, B. (2021). The Impact of Experiencing Proficiency-Based Learning (PBL) on Preservice Teacher Perspectives on PBL, Assessment, and Grading. *Teacher Education Quarterly*, 48(4), 52-77. <https://www.jstor.org/stable/27099522>
- Epstein, J., Santo, R. M. et Guillemin, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *Journal of clinical epidemiology*, 68(4), 435-441.
- Fagnant, A. (2023). *Les pratiques d'évaluation en classe : des compétences professionnelles pour soutenir l'apprentissage des élèves*. Cnesco-Cnam.
- Fédération Wallonie-Bruxelles. (2018, 19 septembre). *Circulaire 6831 : Mise en œuvre des aménagements raisonnables permettant l'accueil, l'accompagnement et le maintien dans l'enseignement ordinaire, fondamental et secondaire des élèves présentant des besoins spécifiques*. http://www.enseignement.be/download.php?do_id=16032&do_check=IQJMXTWCSY
- Fédération Wallonie-Bruxelles. (2019a, 28 février). *Circulaire 7012 : L'évaluation des études dans l'enseignement secondaire ordinaire de plein exercice organisé par la Communauté française – Bulletins – Conseils de classe*. https://www.galilex.cfwb.be/document/pdf/46017_000.pdf
- Fédération Wallonie-Bruxelles. (2019b, 1^{er} novembre). *Décret portant les livres 1er et 2 du Code de l'enseignement fondamental et de l'enseignement secondaire, et mettant en place le tronc commun*. https://www.galilex.cfwb.be/document/pdf/47165_000.pdf
- FESeC. (2021). *Évaluation de l'apprentissage et au service des apprentissages*. *Enseignement catholique Secondaire*. <https://extranet.segec.be/gedsearch/document/23113>
- Klenowski, V. (2009). Assessment for Learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice*, 16(3), 263–268. <https://doi.org/10.1080/09695940903319646>
- Klinger, D.A., McDivitt, P.R., Howard, B.B., Munoz, M.A., Rogers, W.T. et Wylie, E.C. (2015). *The Classroom Assessment Standards for PreK-12 Teachers*. Kindle Direct Press. <https://evaluationstandards.org/classroom/>
- Laveault, D. et Allal, L. (dir.). (2016). *Assessment for learning: Meeting the challenge of implementation (Vol. 4)*. Cham, Heidelberg, New York, Dordrecht, London: Springer.
- Levy-Vered, A. et Nasser-Abu Alhija, F. (2015). Modelling beginning teachers' assessment literacy: The contribution of training, self-efficacy, and conceptions of assessment. *Educational Research and Evaluation*, 21(5-6), 378-406

- Luisoni, M. et Monnard, i. (2015). L'évolution des conceptions de l'évaluation scolaire chez les futurs enseignants en formation initiale. Dans P.-F. Coen et M. Bélaïr (dir.), *Évaluation et autoévaluation : quels espaces de formation ?* (p. 175-196). De Boeck Supérieur.
- McGee, J. et Colby, S. (2014). Impact of an assessment course on teacher candidates' assessment literacy. *Action in Teacher Education*, 36(5-6), 522-532.
- Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American secondary education*, 33(1), 49-64. <https://www.jstor.org/stable/41064623>
- Mertler, C. A. et Campbell, C. (2005). *Measuring Teachers' Knowledge & Application of Classroom Assessment Concepts: Development of the "Assessment Literacy Inventory"*. Online Submission. <https://files.eric.ed.gov/fulltext/ED490355.pdf>
- Ministry of Education Ontario. (2010). *Growing Success: Assessment, Evaluation and Reporting in Ontario's Schools: Covering Grades 1 to 12*. Ministry of Education. <https://www.edu.gov.on.ca/fre/policyfunding/growSuccessfr.pdf>
- 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*. Cnesco-Cnam.
- Oo, C. Z., Alonzo, D. et Asih, R. (2022). Acquisition of teacher assessment literacy by pre-service teachers: A review of practices and program designs. *Issues in Educational Research*, 32(1), 352-373.
- Pastore, S. Teacher Assessment Literacy: A Systematic Review. *Frontiers in Education*, 8, 1217167. <https://doi.org/10.3389/feduc.2023.1217167>
- Plake, B. S. et Impara, J. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(1), 21-27. <https://www.mwera.org/MWER/documents/MWER-1993-Winter-6-1.pdf#page=23>
- Schelings, C. (2021). *Cahier méthodologique. Questionnaire, étude de cas, entretien individuel, focus group et observation in situ*. Université de Liège
- Starck, J. R., Richards, K. A. R. et O'Neil, K. (2018). A conceptual framework for assessment literacy: Opportunities for physical education teacher education. *Quest*, 70(4), 519-535.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238. <https://www.jstor.org/stable/20405538>
- Unal, A. et Unal, Z. (2019). An examination of K-12 teachers' assessment beliefs and practices in relation to years of teaching experience. *Georgia Educational Researcher*, 16(1), 4. <https://digitalcommons.georgiasouthern.edu/teach-elementary-facpubs/47>
- Wiliam, D. (2011). What is assessment for learning? *Studies in educational evaluation*, 37(1), 3-14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Willis, J., Adie, L. et Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, 40, 241-256.
- Xu, Y. et Brown, G. T. (2016). Teacher assessment literacy in practice: *A reconceptualization*. *Teaching and Teacher Education*, 58, 149-162. <https://doi.org/10.1016/j.tate.2016.05.010>
- Zaleski, D. (2015). *Guiding principles for classroom assessment*. <https://www.isbe.net/Documents/guiding-principles.pdf>

6. Annexe - Exemple de scénario traduit et adapté de l'outil ACAI (DeLuca et al., 2016b)

Scénario 3 : Sur les 28 élèves de votre classe, quatre sont officiellement identifiés comme ayant des besoins spécifiques et devant bénéficier d'aménagements raisonnables. Plusieurs autres élèves semblent également avoir besoin d'aménagements mais ne sont pas reconnus officiellement. Vous devez décider de la façon d'évaluer avec précision les apprentissages de tous les élèves de votre classe.

En tant qu'enseignant dans cette situation, selon votre contexte, indiquez la probabilité d'effectuer les actions suivantes.

1 : *Tout à fait improbable*, 2 : *Très peu probable*, 3 : *Peu probable*, 4 : *Un peu probable*, 5 : *Très probable*, 6 : *Absolument probable*

Vous mettez en œuvre, pour toutes les épreuves sommatives, les aménagements requis pour les quatre élèves identifiés. 1 2 3 4 5 6
Je ne sais pas.

Vous mettez en place, avec tous vos élèves, des évaluations formatives adaptées en fonction de leurs besoins d'apprentissage, et ceci jusqu'à l'évaluation de fin de chapitre tout en conservant le même niveau d'exigence pour tous. 1 2 3 4 5 6
Je ne sais pas.

Vous permettez aux élèves de développer leur propre plan individualisé d'apprentissage basé sur leurs points forts, leurs besoins, les objectifs d'apprentissage tout en conservant le même niveau d'exigence pour tous. 1 2 3 4 5 6
Je ne sais pas.

Vous concevez et proposez une variété de tâches d'évaluation qui permettent aux élèves de démontrer la façon dont ils atteignent les acquis d'apprentissage tels que définis dans les programmes. 1 2 3 4 5 6
Je ne sais pas.

Vous adaptez les grilles de correction et les échelles de cotation pour prendre en compte les adaptations et les modifications effectuées dans l'évaluation par rapport au programme initial. 1 2 3 4 5 6
Je ne sais pas.

Vous expliquez aux élèves et à leurs parents le but des adaptations proposées et la manière dont elles seront mises en œuvre et communiquées dans les bulletins des élèves. 1 2 3 4 5 6
Je ne sais pas.

Vous notez les travaux de tous les élèves sur la base des mêmes tâches. 1 2 3 4 5 6
Je ne sais pas.

Vous vous assurez que les tâches évaluatives permettent de mesurer adéquatement les acquis de tous les élèves, y compris ceux à besoins spécifiques. 1 2 3 4 5 6
Je ne sais pas.

Vous discutez et vous négociez les épreuves évaluatives différenciées avec les élèves en fonction de leurs besoins d'apprentissage. 1 2 3 4 5 6
Je ne sais pas.

Vous utilisez la même grille de correction pour tous les élèves. 1 2 3 4 5 6
Je ne sais pas.

Vous créez des grilles de correction différentes pour les quatre élèves à besoins spécifiques. 1 2 3 4 5 6
Je ne sais pas.

Vous utilisez la même grille de correction pour tous les élèves de la classe tout en faisant appel à votre jugement professionnel pour appliquer les critères différemment en fonction des capacités spécifiques de chacun d'eux. 1 2 3 4 5 6
Je ne sais pas.

Souhaitez-vous partager l'une de vos pratiques pour ce genre de situation ? Oui Non

Vous pouvez partager ici l'une de vos pratiques pour ce genre de situation.
.....