

L'évaluation certificative en éducation physique et à la santé : pratiques et défis déclarés des enseignants

David Bezeau

Université de Sherbrooke

Maïa Savard

Université de Sherbrooke

William Lemieux

Université de Sherbrooke

Résumé

L'évaluation certificative en éducation physique représente un enjeu important. Cette étude tente de décrire le processus d'évaluation certificative en éducation physique et à la santé (ÉPS) et d'en identifier les principaux défis. Les données permettant de répondre à ces objectifs ont été collectées auprès d'enseignants d'ÉPS par l'entremise d'entretiens individuels ($n = 11$) et d'entretiens d'instruction au sosie ($n = 10$). La description du processus d'évaluation des participants fait ressortir une diversité de pratiques, de stratégies et d'outils mobilisés dans les phases de planification, de collecte des traces des prestations des élèves, d'interprétation et de jugement, mais également plusieurs défis importants vécus par les enseignants d'ÉPS.

Mots-clés : éducation physique et à la santé, pratiques évaluatives, évaluation des apprentissages, processus d'évaluation certificative, défis

Abstract

Summative assessment in physical education is an important issue. This study attempts to better understand the summative assessment process in physical and health education (PHE) and to identify its main challenges. Data to meet these objectives were collected from PE teachers through individual interviews ($n = 11$) and look-alike interviews ($n = 10$). The description of the participants' assessment process highlights a diversity of practices, strategies and tools mobilized in the planning, information gathering, interpretation and grading phases, but also several important challenges experienced by PHE teachers.

Keywords: physical and health education, assessment practices, assessment of learning, summative assessment, challenges

Introduction

L'adoption d'un mode de vie sain et actif chez les enfants d'âge scolaire est associée à une meilleure santé physique et mentale ainsi qu'à de meilleures performances scolaires (Dubuc et al., 2022). Les connaissances et savoir-faire nécessaires au développement d'un mode de vie sain et actif sont principalement enseignés en cours d'éducation physique et à la santé (ÉPS; Alvarez-Bueno et al., 2017), qui joue un rôle clé dans les écoles en outillant les élèves pour faire des choix éclairés concernant leur santé et leur bien-être (Dubuc et al., 2022). Un enseignement de qualité en ÉPS favorise l'adoption, par les élèves, d'un mode de vie sain et actif en dehors des cours (Chen et al., 2014). Selon Penney et al. (2009), une ÉPS de qualité repose sur trois piliers : le curriculum, la pédagogie et l'évaluation. Cette dernière, lorsqu'elle est bien menée, aide les élèves à constater leurs progrès (Tolgfors, 2018), ce qui peut avoir des effets positifs sur leur apprentissage (Lyngstad et al., 2022; Mougnot, 2016; Tolgfors, 2018) et leur engagement (Duda et al., 2018).

L'évaluation remplit trois fonctions : 1) orienter l'action (diagnostique); 2) réguler l'apprentissage (formative); et 3) certifier les acquis (certificative) (De Ketele, 2010). S'intéresser à la démarche d'évaluation permet de mieux comprendre comment les enseignants articulent ces fonctions dans leur pratique, au-delà du simple choix des objets évalués, ce qui en fait un levier essentiel pour améliorer la qualité de l'enseignement et favoriser la réussite éducative (Ní Chróinín et Cosgrave, 2013).

La présente étude porte sur l'évaluation certificative, identifiée par plusieurs chercheurs comme un enjeu central en ÉPS (Dinan-Thompson et Penney, 2015 ; Leirhaug et Annerstedt, 2016). López-Pastor et al. (2013) l'identifient même comme « *one of the most fraught and troublesome issues physical educators have had to deal with over the past 40 years or so* » (p. 57). Pour que les enseignants d'ÉPS mettent en œuvre une évaluation de qualité, elle doit être spécifique aux compétences que les élèves ont développées dans la situation d'apprentissage et d'évaluation (SAÉ) (Penney et al., 2009). Les enseignants doivent également utiliser des outils de collecte d'information cohérents et inclusifs (Borghouts et al., 2016). Ils doivent placer leurs élèves dans des situations d'évaluation authentiques, ce qui signifie que l'évaluation doit permettre aux élèves d'accomplir des tâches représentatives de contextes réels d'activités sportives pour démontrer leur capacité à mobiliser leurs compétences (López-Pastor et al., 2013 ; Tolgfors et al., 2022). De plus, ils doivent donner des directives claires sur ce qu'ils attendent lors de l'évaluation certificative afin que les élèves comprennent sur quoi ils sont évalués (Borghouts et al., 2016 ; Tolgfors et al., 2022).

Cependant, la majorité des études effectuées en lien avec l'évaluation certificative en ÉPS porte sur les objets d'évaluation, soulignant la prédominance des savoir-faire (Borghouts et al., 2016 ; Dinan-Thompson et Penney, 2015 ; Williams et al., 2020). D'autres études ont permis de mieux connaître trois défis persistants de l'évaluation en ÉPS, soit : 1) le manque de cohérence entre l'enseignement et l'évaluation ; 2) les connaissances limitées des enseignants ; et 3) le caractère intuitif de l'évaluation (Borghouts et al., 2016 ; Dinan-Thompson et Penney, 2015 ; Georgakis et Wilson, 2012 ; Lund et Veal, 2008 ; Marmeleira et al., 2020 ; Williams et al., 2020). En s'intéressant aux pratiques évaluatives sous l'angle de la démarche complète d'évaluation, cette étude cherche à mieux comprendre les choix concrets que posent les enseignants tout au long du processus évaluatif, ce qui pourrait faire émerger des éléments nouveaux ou peu explorés jusqu'ici, et ainsi contribuer à l'amélioration des pratiques en milieu scolaire et, par le fait même, des apprentissages réalisés par les élèves (Mougenot, 2016).

À notre connaissance, la documentation scientifique actuelle ne permet pas de décrire les pratiques évaluatives des enseignants d'ÉPS au-delà des objets d'évaluation. Les questions de recherche auxquelles cet article tentera de répondre sont les suivantes : 1) comment se structure le processus complet d'évaluation certificative mis en œuvre par des enseignants d'ÉPS québécois ? ; et 2) quels défis ces enseignants rencontrent-ils dans ce processus ?

Cadre conceptuel

Cette étude mobilise les quatre premières phases de la démarche d'évaluation des apprentissages de Fontaine et Loye (2017) développée au Québec : 1) planifier les tâches d'évaluation ; 2) collecter les traces de l'apprentissage ; 3) interpréter les résultats ; et 4) poser un jugement.

Dans la première phase, l'enseignant planifie simultanément l'enseignement et l'évaluation. Il détermine alors les savoirs, savoir-faire et savoir-être à enseigner et ceux ciblés comme étant à maîtriser au sein de la compétence lors de la situation d'évaluation (Fontaine et Loye, 2017). Au Québec, les programmes officiels exigent l'enseignement et l'évaluation de trois compétences disciplinaires, soit les compétences « Agir dans divers contextes d'activité physique », « Interagir dans divers contextes d'activité physique » et « Adopter un mode de vie sain et actif » (ministère de l'Éducation, du Loisir et du Sport [MELS], 2006). Chacune de ces compétences est développée à travers différents niveaux taxonomiques, soit les savoirs (connaissances ; ex. : muscles du corps humain), les savoir-faire (habiletés et stratégies ; ex. : lancer par-dessus l'épaule ou rôle dans l'équipe) et les savoir-être (ex. : sécurité et éthique). Deux documents officiels, influencés par l'approche socioconstructiviste, guident cette planification : la *Progression des apprentissages* (document non contraignant ; MELS, 2009) et le *Cadre d'évaluation des apprentissages* (document contraignant ; MELS, 2011). La *Progression des apprentissages* identifie les contenus (savoirs, savoir-faire, savoir-être) à enseigner chaque année selon le niveau de l'élève, soit de la 1^{re} année du primaire (6 ans) à la fin de son secondaire (17 ans). Pour sa part, le *Cadre d'évaluation des apprentissages* explicite ce qui doit être évalué par l'enseignant d'ÉPS. Les trois critères d'évaluation identifiés sont les mêmes pour les trois compétences disciplinaires, tout au long du parcours scolaire de l'élève : 1) la cohérence de la planification ; 2) l'efficacité de l'exécution ; et 3) la pertinence du retour réflexif (MELS, 2011).

La deuxième phase concerne la collecte des traces d'apprentissage pour consigner les apprentissages des élèves (Fontaine et Loye, 2017). L'enseignant d'ÉPS doit alors développer les outils et les stratégies de prise d'informations lui permettant de recueillir toutes les données nécessaires au regard des critères établis dans la première phase (planification). En parallèle, l'enseignant doit interpréter ses observations pour leur attribuer une valeur (Fontaine et Loye, 2017). Dans la troisième phase, cette

interprétation s'effectue souvent à l'aide d'une grille d'évaluation, dans laquelle l'enseignant d'ÉPS utilise divers symboles, tels que ++, +, – et -- (où ++ indique que l'élève a parfaitement réussi, + signifie que l'élève a très bien réussi malgré quelques erreurs, – indique que l'élève a eu quelques succès, et -- montre que l'élève a échoué à la tâche). D'autres peuvent utiliser des chiffres ou des lettres au lieu de symboles et certains préfèrent des indications qualitatives, telles que «réussi parfaitement», «réussi avec de l'aide», «n'a pas réussi» et «n'a pas réussi même avec de l'aide».

Enfin, la quatrième phase — le jugement — constitue le cœur du processus d'évaluation (Fontaine et Loye, 2017). Elle consiste à porter un regard global sur les apprentissages à partir des traces et de leur interprétation. Cette phase est souvent confondue avec l'interprétation, puisque l'enseignant porte en effet un jugement lorsqu'il interprète la prestation d'un élève et qu'il détermine les traces à inscrire sur sa grille d'évaluation. Cependant, le jugement renvoie surtout à la notation. Par exemple, deux enseignants ayant les mêmes observations peuvent accorder des pondérations différentes aux critères, menant à des notes distinctes. L'interprétation permet d'attribuer une valeur à une action, alors que la notation synthétise l'ensemble des interprétations en une note (Durand et Chouinard, 2012).

Les objectifs du présent article sont : 1) de décrire le processus complet d'évaluation certificative d'enseignants d'ÉPS québécois à partir de ces quatre phases ; et 2) d'identifier les défis rencontrés par les enseignants dans ce processus.

Méthodologie de la recherche

Le comité d'éthique de l'Université de Sherbrooke a préalablement approuvé cette étude qualitative. Onze enseignants d'ÉPS y ont participé volontairement après avoir répondu à une invitation envoyée à la Fédération des éducateurs et éducatrices physiques enseignants du Québec. Parmi eux, dix enseignent au primaire (6 hommes, 4 femmes) et un au secondaire (1 homme). L'expérience moyenne en enseignement est de 14,18 ans ($\overline{ET} = 7,11$).

La collecte de données s'est déroulée en deux temps. D'abord, chaque participant a pris part à un entretien individuel semi-dirigé d'environ 45 minutes, réalisé et enregistré sur Teams en octobre 2020. Le guide d'entretien comportait 14 questions ouvertes réparties selon les quatre phases du processus d'évaluation (Fontaine et Loye, 2017) :

1) la planification (six questions; ex. : «Qu'évaluez-vous principalement en ÉPS?»); 2) la collecte des traces (trois questions; ex. : «Lors de votre situation d'évaluation, quelles stratégies utilisez-vous pour vous assurer de recueillir les informations nécessaires auprès de tous vos élèves?»); 3) l'interprétation (trois questions; ex. : «Comment déterminez-vous quelle trace attribuer à ce que vous observez chez un élève?»); et 4) le jugement (deux questions; ex. : «En fin d'étape, comment établissez-vous la note inscrite au bulletin à partir de l'ensemble des évaluations réalisées?»). Ensuite, 10 de ces 11 participants ont pris part à un entretien d'instruction au sosie en dyade¹. Ce type d'entretien a permis d'enrichir les réponses obtenues dans les entretiens individuels et de mieux décrire l'activité réelle du geste évaluatif des participants. L'entretien d'instruction au sosie repose sur la coanalyse d'une situation de travail entre deux personnes, l'instructeur et le sosie :

L'objectif est de conduire l'instructeur qui s'adresse à son remplaçant potentiel, le sosie, à être confronté à son expérience en dialoguant avec un autre. Il s'agit d'inciter l'instructeur à regarder son expérience avec les yeux du sosie dans la mesure où le sosie ne peut pas la voir de la même façon que l'instructeur. Répondre aux questions du sosie oblige ainsi l'instructeur à rendre compte de son travail et à l'amener suite aux questions du sosie, à exprimer des éléments qu'il n'aurait pas pensé verbaliser. (Brau-Antony et Hubert, 2014, p. 5)

La consigne donnée par le chercheur à un des participants était la suivante : «imagine que ce soit ton sosie et qu'il doit te remplacer dans ta situation d'évaluation sans que personne ne s'en rende compte. Quelles sont les informations que tu devrais lui transmettre?». Pendant les 30 premières minutes, l'instructeur contextualisait le cours et détaillait son processus d'évaluation en fournissant le plus de détails possible (ex. : «Tu devras t'assurer d'avoir un atelier pour laisser le temps aux élèves de s'exercer avant de venir te voir pour se faire évaluer»). Le sosie pouvait poser des questions à tout moment pour s'assurer de bien comprendre (ex. : «Je dois planifier un atelier pour que les élèves s'exercent, mais à quel moment dans le cours je peux commencer à évaluer

1 Un des participants ayant participé aux entretiens individuels ne souhaitait pas participer aux entretiens d'instruction au sosie et son choix a été respecté.

les premiers élèves ? »). Les 30 dernières minutes étaient divisées en deux étapes : 1) le sosie « racontait son histoire », c'est-à-dire qu'il résumait le processus d'évaluation de l'instructeur et ce dernier devait seulement intervenir si certains éléments de l'histoire ne correspondaient pas à sa façon de faire ; et 2) le sosie comparait ses pratiques à celles de l'instructeur afin de mieux comprendre les motifs derrière leurs choix respectifs. Cet entretien de 60 minutes, réalisé sur Teams en novembre 2020, était enregistré et jumelait les participants en fonction de leurs disponibilités. L'assignation en tant que sosie ou instructeur s'effectuait au hasard avant le début de la rencontre.

Une analyse de contenu inductive a été menée en respectant les étapes proposées par Creswell et Creswell (2022) : transcription intégrale des entretiens, lecture exploratoire avec prise de notes, identification d'énoncés significatifs, attribution initiale d'unités de classification et catégorisation inductive. Comme ils portaient sur les mêmes objets, à savoir le processus d'évaluation et les défis rencontrés, l'analyse des deux types d'entretien a eu lieu conjointement. Toutefois, l'équipe a porté une attention particulière pour éviter le double comptage : si une information apparaissait deux fois pour un même participant, on la comptabilisait seulement une fois. Cette approche visait à enrichir la compréhension sans biaiser les fréquences. Un étudiant de 2^e cycle (maîtrise) et une enseignante en ÉPS ont ensuite effectué des ententes interjuges. Ils ont classé 40 % des énoncés dans les catégories définies par le chercheur principal (Patton, 2015), avec des taux d'entente respectifs de 92,1 % et 97,6 %. Le chercheur principal a aussi procédé à une entente intrajuge, obtenant un taux de 95,5 %.

Résultats et discussion

Les deux objectifs de recherche — 1) décrire le processus complet d'évaluation certificative d'enseignants d'ÉPS québécois ; et 2) identifier les défis rencontrés par les enseignants dans ce processus — guident la présentation des résultats. La discussion des résultats à la lumière de la documentation scientifique suit immédiatement afin de faire ressortir les principaux éléments de réflexion.

Processus complet d'évaluation certificative d'enseignants d'ÉPS québécois

Cette première section présente les résultats pour chacune des quatre phases du processus d'évaluation certificative, accompagnés d'éléments de discussion en lien avec les écrits scientifiques existants.

Planification de l'évaluation

Le Tableau 1 montre que les principaux objets d'évaluation ciblés par les participants sont : 1) les compétences disciplinaires ($n = 6$); 2) les savoir-faire ($n = 6$); 3) les trois critères d'évaluation des documents officiels ($n = 5$); et 4) les savoir-être ($n = 5$).

Tableau 1

Principaux objets d'évaluation ciblés par les participants

Objets d'évaluation	<i>n</i>
Compétences disciplinaires (<i>Agir, Interagir, Adopter</i>)	6
Savoir-faire (ex. : lancer à une main, démarquage, circulation de l'objet)	6
Critères d'évaluation officiels (planification, exécution, réflexion)	5
Savoir-être (ex. : respect des adversaires, respect du matériel)	5
Participation	3
Port de vêtements adéquats pour pratiquer de l'activité physique	2
Comportements	2
Savoirs (ex. : connaissance des règles au badminton ou des aliments sains)	2

Note. n représente le nombre de participants ayant ciblé l'objet d'évaluation ($N = 11$).

Ces quatre principaux objets d'évaluation montrent que les participants tiennent compte des documents officiels lors de la planification de leur évaluation. En effet, les trois compétences disciplinaires et leurs critères d'évaluation, éléments centraux du programme, figurent parmi les objets d'évaluation les plus fréquemment ciblés. Jules² mentionne notamment : « Mon collègue et moi travaillons ensemble sur nos évaluations et ce qu'on évalue principalement, ce sont les trois compétences du programme ». Les savoir-faire ($n = 6$) et les savoir-être ($n = 5$), également reconnus dans les documents

2 Tous les noms utilisés dans l'article sont fictifs pour respecter la confidentialité des participants.

officiels, sont bien représentés, contrairement aux savoirs qui semblent être moins évalués par les participants ($n = 2$). Trois objets d'évaluation ciblés par certains participants ne figurent pas dans les documents officiels : la participation ($n = 3$), le port de vêtements adéquats ($n = 2$) et le comportement ($n = 2$).

Les résultats révèlent une priorité accordée aux savoir-faire, ce qui correspond aux constats observés lors d'études antérieures (Dinan-Thompson et Penney, 2015 ; Williams et al., 2020). Toutefois, nos résultats montrent une valorisation marquée des savoir-être (ex. : respect des adversaires, respect du matériel), généralement absents des études précédentes. Cette différence pourrait s'expliquer par leur importance dans les programmes québécois, mais elle soulève la question de leur place dans les cadres d'évaluation proposés dans d'autres contextes géographiques et disciplinaires, notamment ceux où l'accent est presque exclusivement mis sur les savoir-faire.

Collecte des traces de l'apprentissage et les principales stratégies

Le Tableau 2 présente quatre principales stratégies utilisées pour recueillir les informations nécessaires auprès de tous les élèves.

Tableau 2

Principales stratégies de prise d'informations

Stratégies de prise d'informations	<i>n</i>
Occuper les élèves avec un jeu autonome	8
Prendre des traces de façon continue	6
Utiliser la captation vidéo	4
Donner des rôles aux élèves	3

Note. *n* représente le nombre de participants ayant adopté cette stratégie de prise d'informations ($N = 11$).

La stratégie la plus fréquente consiste à occuper les élèves avec un jeu autonome afin d'évaluer sans interruption. Mario, par exemple, utilise un jeu de ballon lorsqu'il évalue la compétence « Agir dans divers contextes d'activité physique » : « habituellement, j'ai une section évaluation qui prend le tiers du gymnase, les deux autres tiers, c'est un jeu autonome où les élèves sont capables de se gérer seuls ».

Six enseignants prennent des traces de manière continue tout au long des cours, ce qui allège leur travail lors des périodes d'évaluation :

Je suis pas mal tout le temps en train de prendre des traces. Ça dépend de ce que j'évalue, mais j'aime ça le faire en continu pour me libérer davantage quand j'évalue des éléments plus précis à la fin. Le danger, c'est que ça peut devenir lourd, il faut trouver le juste milieu. (Suzie)

Quatre enseignants utilisent la captation vidéo pour s'assurer de recueillir des informations sur tous les élèves en différé. Cette méthode permet de réviser les prestations de certains élèves ou certaines équipes, au besoin, dans des situations d'apprentissage plus complexes :

J'utilise la vidéo pour les situations d'apprentissage plus complexes où il y a beaucoup d'éléments techniques à observer. Par exemple, en flag-football, c'est difficile de voir les principes de synchronisation en pleine action. Donc, je vais utiliser la vidéo pour avoir l'évaluation la plus exacte possible de chaque enfant. (Robert)

Enfin, trois enseignants attribuent des rôles aux élèves afin de se dégager du temps pour observer.

Des questions visaient également à connaître les outils de prise d'informations qu'utilisent les participants. Parmi ces outils figure l'application iDoceo sur iPad (la plus populaire ; $n = 7$) permettant, entre autres, de créer des grilles d'évaluation et d'y consigner des observations. Sylvie précise que : « c'est assez simple de prendre des traces en les associant avec les icônes disponibles dans l'application. Je peux prévoir mes critères d'avance et juste choisir la bonne icône qui représente mes observations ». Viennent ensuite les grilles d'évaluation en format papier ($n = 5$) et le cahier de l'élève³ ($n = 4$).

Malgré cette variété apparente d'outils et de stratégies, l'observation demeure la méthode d'évaluation quasi exclusive en ÉPS. Ces résultats rejoignent ceux d'études précédentes (Borghouts et al., 2016 ; Dinan-Thompson et Penney, 2015 ; Sofó et al., 2013) et sont cohérents avec la principale façon de démontrer ses apprentissages en ÉPS : le mouvement. Dans leur étude, Sofó et al. (2013) mentionnent qu'il est pourtant crucial

3 Un cahier de l'élève est habituellement une feuille ou un document qui permet à l'élève de répondre aux critères de « cohérence de la planification » et de « pertinence du retour réflexif » en y inscrivant son enchaînement individuel ou sa stratégie d'équipe, et en justifiant ses choix, puis d'effectuer un retour réflexif dans le format souhaité par l'enseignant à la fin de la situation d'évaluation (questions fermées/ouvertes, autoévaluation, etc.).

d'utiliser une variété de méthodes d'évaluation adaptées aux besoins divers des élèves pour assurer leur développement et leurs apprentissages. Les 39 enseignants d'ÉPS interrogés dans l'étude de van Rossum et Morley (2018) ont également exprimé le besoin d'avoir des outils pratiques et simples pour être en mesure de diversifier et d'étendre la portée de leurs méthodes d'évaluation au-delà de l'observation directe. Ce décalage révèle des obstacles à l'intégration d'outils variés, notamment en lien avec la formation continue et le soutien institutionnel nécessaires pour appuyer une telle diversification. Marmeleira et al. (2020) et Williams et al. (2020) ajoutent que les enseignants d'ÉPS ont de la difficulté à colliger des traces de façon efficace et rigoureuse sur tous leurs élèves. Le principal enjeu scientifique au regard de la collecte des traces, c'est qu'au-delà de ces quelques études, les pratiques et les stratégies efficaces des enseignants d'ÉPS dans cette phase restent encore largement sous-explorées. Pourtant, être rigoureux dans sa prise d'informations est essentiel pour assurer une évaluation fiable et valide (Borghouts et al., 2016; Hirsch, 2017). Ainsi, des recherches futures pourraient non seulement décrire de manière plus approfondie différents outils et différentes méthodes ou stratégies de prise d'informations utilisés par des enseignants d'ÉPS, et leur impact sur la cohérence de l'évaluation, mais aussi explorer les facteurs qui influencent leur adoption et leur mise en œuvre.

Interprétation

Trois principaux éléments guident l'interprétation : 1) l'expérience ($n = 5$); 2) une grille d'interprétation ($n = 4$); et 3) le document de *Progression des apprentissages* (MELS, 2009; $n = 3$). Cinq enseignants s'appuient sur leur intuition et leur expérience pour associer une trace évaluative (ex. : ++ / + / - / --) à ce qu'ils observent chez un élève : «Je te dirais que c'est vraiment l'expérience qui m'aide à évaluer plus rapidement. Je n'ai plus besoin de me préparer des grilles avec des critères précis. J'ai des attentes en tête qui guident les traces que je prends» (Sara).

Quatre enseignants préfèrent spécifier la valeur de chaque trace à partir d'une grille d'interprétation. Une telle grille définit la valeur des traces selon l'élément observé :

J'ai une grille d'interprétation qui me permet d'avoir des éléments observables pour chaque critère d'évaluation. J'ai habituellement un élément pour la cohérence de la planification, trois pour l'efficacité de l'exécution et un pour le retour réflexif. Pour chaque élément observable, je vais spécifier

ce que je vais regarder durant l'évaluation. Par exemple, si j'évalue la position d'un élève par rapport à la cible dans une activité collective, je vais regarder s'il se dirige vers la cible et vers les espaces libres. La grille d'interprétation me permet de déterminer les niveaux attendus pour chaque élément observable et guider ma prise de traces. Par exemple, si l'élève se déplace toujours vers la cible ou dans un espace libre, il aura la trace «++». S'il le fait souvent, il aura la trace «+», et ainsi de suite. (Paul)

Enfin, trois enseignants mentionnent utiliser le document ministériel de la *Progression des apprentissages* (MELS, 2009) pour savoir si les élèves dépassent les attentes ou non selon leur niveau d'enseignement.

Peu d'études explorent les facteurs influençant l'interprétation en ce qui appelle à une réflexion plus approfondie sur la manière dont les facteurs contextuels, sociaux et personnels influencent cette interprétation, au-delà de la simple expérience et intuition des enseignants (Borghouts et al., 2016). D'autres disciplines confirment ce rôle central de l'expérience dans le processus d'interprétation (Lunenbergh et al., 2007). Cependant, une question reste en suspens : dans quelle mesure ces facteurs personnels influencent-ils la cohérence des évaluations, et dans quelles conditions ces biais peuvent-ils être atténués par des outils ou des pratiques plus standardisés ? Néanmoins, la question de la subjectivité mérite d'être mise en débat, notamment en comparant les pratiques d'évaluation dans différentes disciplines. Est-il possible que la subjectivité, bien que perçue comme un biais, puisse aussi être une force pour mieux comprendre les besoins individuels des élèves, selon les contextes d'enseignement (Lee, 2014) ? Les éléments de discussion en lien avec les défis de l'évaluation en ÉPS (objectif #2) permettront d'aller plus loin à ce sujet en abordant l'enjeu de subjectivité, issu entre autres de ces pratiques d'interprétation intuitive.

Jugement

Deux principales méthodes sont utilisées pour porter un jugement à la fin d'une SAÉ. La première est de pondérer chacun des éléments évalués, puis de les additionner ($n = 4$) :

Ça prend moins de temps de juste faire des additions. Donc, ce que je fais, c'est que je détermine la pondération de chaque élément observable évalué et j'additionne tout simplement le total obtenu pour chacun de ceux-ci en

fonction des traces que j'ai prises. Ça me donne ensuite une note sur 100. L'avantage, c'est que tout se fait automatiquement avec ma tablette, ça me sauve un temps fou. (Sara)

La deuxième est d'utiliser un outil d'aide au jugement qui permet de placer les élèves dans des catégories associées à des notes préétablies ($n = 4$) :

J'utilise un outil d'aide au jugement à la fin d'une SAÉ pour donner une note à mes élèves. Cet outil permet d'identifier, avec des bonds de 8 %, la note que devrait obtenir un élève. Par exemple, si un étudiant obtient 100 %, ça signifie qu'il a toujours surpassé les attentes et n'a jamais eu besoin d'aide. Un élève qui a 92 % a souvent surpassé les attentes et n'a pas eu besoin d'aide... et ça continue comme ça jusqu'à 44 % je crois. (Robert)

Pour mettre la note au bulletin en fin d'étape à partir de l'ensemble des SAÉ, trois principales méthodes sont utilisées. Celle qui est la plus utilisée est la même qu'à la fin d'une SAÉ, soit de pondérer chacun des éléments évalués, puis de les additionner ($n = 6$). Paul précise qu'«une pondération est accordée à chaque SAÉ faite avec les élèves durant l'étape pour en arriver à une note sur 100». Les deux autres méthodes utilisées sont de : pondérer chacun des éléments évalués et de les additionner, mais ensuite de porter un jugement sur la note ainsi obtenue et l'ajuster au besoin ($n = 3$); et de porter un jugement global à partir de toutes les traces collectées et de l'image qu'ils ont des élèves ($n = 2$).

Au-delà des méthodes utilisées, huit enseignants considèrent les traces prises tout au long de la SAÉ dans la note finale, bien que l'importance accordée varie : 1) trois leur accordent la même importance que les traces prises pendant la situation d'évaluation ; 2) trois mentionnent qu'elles influencent la note, mais dans une moindre mesure en comparaison à celles prises lors de la situation d'évaluation ; 3) un soutient qu'elles influencent la note, mais seulement positivement ; et 4) un dernier déclare qu'elles permettent uniquement de redonner une chance à certains élèves.

Trois éléments influencent le jugement en fin d'étape. D'abord, huit participants considèrent l'aide apportée à l'élève pendant la SAÉ. Un élève qui réussit bien les différentes tâches proposées et de manière autonome obtiendra une meilleure note qu'un élève qui réussit bien, mais qui a besoin d'aide de la part de l'enseignant. Pour les participants, cette aide était souvent liée au respect du plan (ex. : se rappeler son

enchaînement) ou à la pertinence du retour réflexif (ex. : guider l'élève dans ses réponses). Ensuite, cinq enseignants considèrent la progression des élèves dans leur note à la fin de l'étape. Cet élément divisait davantage les participants. Même parmi ceux qui la considèrent, certains croient que la progression ne doit pas permettre à un élève moins fort d'avoir une meilleure note qu'un élève considéré fort alors que d'autres croient que la progression est plus importante que le niveau de performance : «Ceux qui s'améliorent davantage ont une meilleure note. L'élève qui est faible au début et qui termine moyen va avoir une meilleure note que celui qui est moyen du début à la fin» (Laurie).

Je vais généralement bonifier de 5 à 10 % pour la progression, mais un élève de niveau moyen qui progresse ne pourra jamais avoir la même note que l'élève de niveau fort si les deux travaillent bien dans les cours, peu importe sa progression. (Philippe)

Enfin, cinq enseignants jugent que les trois compétences disciplinaires doivent avoir une importance équivalente dans la notation, alors que quatre accordent une moins grande valeur à la compétence «Adopter un mode de vie sain et actif».

Les méthodes variées et les critères personnels montrent que deux enseignants, à partir des mêmes traces, peuvent aboutir à des jugements très différents. La subjectivité, présente dans l'interprétation comme dans le jugement, sera abordée plus en profondeur dans la discussion sur les défis évaluatifs.

Défis rencontrés par les enseignants d'ÉPS dans le processus d'évaluation

Dans le cadre du deuxième objectif de recherche, les entretiens individuels ont permis de dégager plusieurs défis en lien avec l'évaluation en ÉPS (voir Tableau 3).

Tableau 3*Défis de l'évaluation en ÉPS*

Défis	<i>n</i>
Associer une valeur à des observations (interprétation)	7
Mauvaise compréhension de l'évaluation de la part des enseignants	5
Aucune uniformité dans les différentes pratiques évaluatives	5
Évaluer la compétence « Interagir dans divers contextes d'activité physique »	5
Gérer les comportements des élèves pendant l'évaluation	3
Évaluer la compétence « Adopter un mode de vie sain et actif »	3

Note. *n* représente le nombre de participants ayant mentionné ce défi ($N = 11$).

L'un des principaux défis soulevés par sept participants est la difficulté d'associer une valeur à des observations, en raison de la forte subjectivité de l'interprétation en ÉPS. Comme le mentionne Patrick : « Comment je fais pour décider quelles traces vaut ce que j'observe ? C'est ça l'autre grande difficulté. Être capable de dire que ce que je vois ça dépasse mes attentes ? Ça y répond à la moitié ? Ça n'y répond pas du tout ? ».

Cinq enseignants mentionnent une compréhension limitée de l'évaluation. En effet, comme le rapporte Paul, chacun fait du mieux qu'il peut selon ce qu'il comprend des documents officiels :

L'évaluation en ÉPS, c'est un tabou parce que je pense qu'il n'y a personne qui est à l'aise avec la façon d'évaluer et de mettre une note ; chacun fait du mieux qu'il peut de son bord en pensant bien faire et selon sa compréhension des documents.

Cinq enseignants parlent du manque d'uniformité dans les pratiques évaluatives, que celles-ci varient grandement d'une école à l'autre, et même entre collègues. Ces enseignants soulèvent une absence d'orientations claires dans les documents officiels au regard de l'évaluation en général, mais particulièrement en ce qui a trait à l'interprétation et à la notation : « La principale difficulté, c'est qu'il manque des critères clairs [...] il manque d'uniformité entre les profs d'éduc par rapport à c'est quoi les attentes, parce que ces attentes sont imprécises » (Patrick).

Cinq enseignants identifient aussi la compétence « Interagir dans divers contextes d'activité physique » comme particulièrement difficile à évaluer. En effet, comme le

mentionne Mario, la valeur de l'évaluation peut être remise en question lorsqu'on a si peu de temps pour porter un jugement sur une compétence complexe à évaluer :

L'interagir, c'est hyper complexe [...] avec le nombre d'élèves, le nombre de périodes accordées pour l'évaluation et le temps réel disponible pendant les cours, tu vas avoir deux minutes à peine pour regarder un élève. Qu'est-ce que tu vois en deux minutes ? C'est ça la complexité je trouve [...]. Parce que la norme administrative dit qu'au 16 octobre, il faut que j'aie sorti une note, bien tu mets vite une note, tu l'as vu deux minutes et c'est ça que ça donne [...]. Mais la valeur de ton évaluation dans tout ça ?

Enfin, la gestion des comportements des élèves en attente d'être évalués ou qui l'ont déjà été, ainsi que l'évaluation de la compétence «Adopter un mode de vie sain et actif» ont chacune été mentionnées par trois participants.

Les entretiens d'instruction au sosie ont fait émerger trois défis qui n'avaient pas été nommés lors des entretiens individuels. D'abord, tous les participants disent évaluer surtout pour répondre à des obligations administratives, seulement trois expriment également évaluer afin de soutenir l'apprentissage des élèves.

Le deuxième défi est illustré par les propos de Mario :

...c'est fou le temps que tu perds à juste faire ça (évaluer). Pis ce temps-là, je ne l'ai pas mis ailleurs [...] en quoi ça nous avance d'y envoyer le message qu'il a 55 % [...] ? Est-ce que je suis en train de répondre à ma mission de le rendre actif si tous les 4 mois je lui dis qu'il échoue ?

Au cours des entretiens d'instruction au sosie, 7 des 13 participants ont remis en question la pertinence d'évaluer dans leur discipline, y voyant une perte de temps qui nuit à l'enseignement.

Le dernier défi concerne la subjectivité de l'évaluation en ÉPS. Bien que les entretiens individuels aient permis d'identifier cet élément, les données issues des entretiens d'instruction au sosie permettent de préciser que les participants ne se sentent pas confortables lorsqu'ils mettent une note, parce qu'ils ne sont pas en mesure de la justifier. Comme le mentionne Patrick : «on ne sait pas ce qui est important à évaluer. Pour le savoir-faire "Passer à un joueur mieux placé par rapport à la cible", toi et moi on peut avoir des attentes différentes et ça devient subjectif». Les participants font ressortir

trois éléments qui illustrent bien cette notion de subjectivité dans l'évaluation causée par des attentes jugées imprécises dans les documents ministériels : 1) quand on met une note, doit-on pondérer et calculer, ou plutôt porter un jugement global sur l'élève ? ; 2) est-ce que les trois compétences et chacune de leurs composantes doivent avoir une pondération équivalente dans le calcul de la note à la fin d'une étape ? ; et 3) doit-on évaluer la progression de l'élève ou non ? Ces questionnements ont suscité de longues discussions entre les participants lors du second entretien. Faute de repères, chacun développe des pratiques qu'il juge « les moins pires et les moins lourdes » (François).

La mauvaise compréhension de l'évaluation de la part de certains enseignants d'ÉPS dans notre étude nous amène à remettre en question la clarté et la pertinence des documents proposés aux enseignants d'ÉPS pour planifier leur évaluation. Nos résultats indiquent que les enseignants d'ÉPS participants se réfèrent aux documents officiels pour planifier leur évaluation. Toutefois, bien que ces derniers affirment que leur évaluation respecte les orientations émises dans ces documents, certains identifient une incohérence dans le processus d'évaluation en raison d'attentes imprécises. Comme l'ont indiqué des études précédentes (Borghouts et al., 2016 ; Van der Mars et al., 2018 ; Williams et al., 2020), cela fait alors en sorte que chacun interprète les documents officiels à sa façon et évalue d'autres éléments, comme la présence au cours, la participation, le comportement, l'effort et le port de vêtements adéquats. Leirhaug et Annerstedt (2016), dans leur étude effectuée auprès de 1486 élèves et 23 enseignants dans 6 écoles secondaires norvégiennes, arrivent à des résultats similaires et identifient également une dissonance entre les objets d'évaluation utilisés par les enseignants d'ÉPS et ce qui est proposé dans les documents officiels. Perez-Roux (2011) explique cette situation par les relations dynamiques constituant l'identité professionnelle des enseignants d'ÉPS. En effet, selon cette auteure, les actions et les décisions de ces derniers sont influencées par une composante de l'identité professionnelle en particulier, soit les représentations de la discipline, qui les pousseraient à accorder une importance plus grande à certains contenus dans leurs pratiques d'enseignement. Ce sentiment d'incertitude dans l'évaluation par compétences face à la justification des notes illustre plus largement les défis inhérents à l'observation et au jugement professionnel dans plusieurs disciplines scolaires (Blömeke, 2017 ; Govaerts, 2016). Ainsi, bien que ces enjeux soient discutés ici dans le contexte particulier de l'ÉPS, ils rejoignent des préoccupations plus larges en éducation quant à la subjectivité du

jugement enseignant et à l'interprétation des attentes dans un cadre curriculaire fondé sur le développement de compétences (Blömeke, 2017; Morrissette et Legendre, 2012).

Au Québec, sachant que toute la documentation offerte aux enseignants d'ÉPS concerne uniquement la phase de planification (contenus à enseigner et critères à évaluer), faut-il s'étonner que chaque enseignant teinte les zones grises des phases d'interprétation et de jugement en fonction de son identité professionnelle? Ce deuxième élément de discussion, issu des difficultés de l'évaluation en ÉPS identifiées par les participants, nous ramène à la subjectivité soulevée précédemment dans les phases d'interprétation et de jugement. Les réponses de nos participants dans l'étude montrent que : 1) l'interprétation repose principalement sur l'expérience et l'intuition; 2) le processus de jugement varie entre chaque enseignant, que ce soit au regard de l'attribution de la note en soi (calculs ou jugement? pondérations accordées?) ou bien des éléments qui influencent celle-ci (ex. : aide apportée, progression de l'élève); 3) la subjectivité occupe une grande place dans l'interprétation et dans le jugement; et 4) les attentes ministérielles pour ces deux phases sont imprécises ou inexistantes, ce qui entraîne une disparité dans les méthodes utilisées par chacun. Nos résultats corroborent ainsi ceux d'autres études (Dinan-Thompson et Penney, 2015; Svennberg et al., 2018; Tolgfors, 2018) qui relèvent l'existence d'un défi majeur dans l'interprétation en ÉPS, particulièrement au regard de la cohérence des pratiques et des orientations ministérielles. Deux autres études identifient également la nature subjective de l'interprétation en ÉPS (Borghouts et al., 2016; Georgakis et Wilson, 2012). Marmeleira et al. (2020) abordent également cette difficulté pour la phase de jugement. Ces auteurs mentionnent qu'il n'y a pas d'uniformité dans l'attribution d'une note au bulletin, chacun fait à sa façon. Dans leur étude, Williams et al. (2020) identifient une disparité dans les éléments d'évaluation qui détermineront la note finale. Leurs résultats montrent que la pondération accordée à chacun des quatre grands domaines de leur programme (activités physiques, condition physique, connaissances en ÉPS, et participation et comportement) pour accorder une note sur le bulletin est différente dans chacune des écoles et entre chaque enseignant d'une même école. Ces disparités entraînent un questionnement : comment s'assurer d'une meilleure uniformité du processus évaluatif en ÉPS?

Cette question est complexe à répondre, pas seulement d'un point de vue opérationnel, mais aussi idéologique. En effet, la subjectivité, lorsqu'elle est circonscrite et appropriée, peut améliorer le processus d'évaluation en incorporant des valeurs et des

attitudes professionnelles (ministère de l'Éducation, 2025). Cela est particulièrement intéressant dans une évaluation par compétences (comme c'est le cas au Québec), puisque la subjectivité est essentielle pour évaluer certaines composantes de la compétence (ex. : pertinence du retour réflexif). Dans le domaine de l'ÉPS, Graham et al. (2020) soutiennent que les performances des élèves peuvent être multifacettes (ex. : technique, créativité, tactique, savoir-être) et que, dans ces cas, une certaine subjectivité est inévitable et même nécessaire pour évaluer de manière holistique les élèves. Kirk (2010) ajoute que les situations d'évaluation authentiques que l'on souhaite de plus en plus mettre en place en ÉPS sont multifacettes ; elles exigent donc une certaine subjectivité dans l'interprétation et le jugement. Pour combiner subjectivité avec fiabilité, validité et équité de l'évaluation, Fletcher et Mandigo (2008) précisent que l'enseignant doit être expert des contenus évalués. Le défi est alors double pour les formations professionnelles : développer les compétences en évaluation des enseignants d'ÉPS, mais également améliorer leurs connaissances didactiques.

Ainsi, les études recensées montrent que viser l'uniformité des pratiques évaluatives serait contre-productif, mais également qu'une subjectivité mal balisée aurait un impact sur la fiabilité, la validité et l'équité de l'évaluation. Des recherches futures seraient nécessaires pour mieux comprendre comment équilibrer la subjectivité nécessaire à une évaluation holistique en ÉPS avec le besoin d'avoir une évaluation rigoureuse et équitable pour tous les élèves.

Conclusion

Cette étude avait pour objectif de décrire le processus complet d'évaluation certificative mis en place par des enseignants d'ÉPS et d'identifier les défis inhérents à ce processus. Les résultats mettent en lumière les tensions qui traversent le processus évaluatif en ÉPS, en particulier dans les phases d'interprétation et de jugement. Si l'expérience et l'intuition des enseignants apparaissent comme des repères majeurs dans l'interprétation des traces évaluatives, ces pratiques s'inscrivent dans un contexte où les outils de référence manquent de clarté et de précision, laissant une large place à la subjectivité et à un certain inconfort chez les enseignants face à la justification des notes. Ce flou interprétatif engendre une grande variabilité des pratiques, non seulement entre les établissements, mais aussi au sein d'une même école, compromettant ainsi la cohérence et l'équité de

l'évaluation. La compétence « Interagir dans divers contextes d'activité physique » et celle visant l'adoption d'un mode de vie sain et actif posent, à cet égard, des défis particuliers.

Face à ces constats, plusieurs recommandations se dégagent. D'abord, il est essentiel de bonifier les outils de référence fournis aux enseignants afin de préciser les critères de réussite, les niveaux attendus et les éléments observables. Ensuite, la mise en place d'espaces collaboratifs entre enseignants — comme des communautés de pratique — pourrait favoriser l'élaboration de repères communs pour guider l'interprétation et le jugement. De plus, la formation initiale et continue devrait accorder une plus grande place à l'analyse de cas concrets d'évaluation, au développement de l'objectivité professionnelle et à la clarification des visées éducatives de l'ÉPS. En somme, pour que l'évaluation en ÉPS soit perçue non plus comme un mal nécessaire, mais comme un levier d'apprentissage, il importe d'agir à la fois sur les outils, les pratiques professionnelles et les conditions collectives d'exercice du jugement enseignant.

Bien que la présente étude offre un éclairage riche et approfondi du processus d'évaluation certificatif en ÉPS, certaines limites doivent être considérées. D'abord, l'échantillon composé de 11 enseignants provenant majoritairement de l'ordre d'enseignement primaire permet de refléter une diversité d'expériences professionnelles, mais limite la généralisation des résultats, particulièrement pour l'ordre d'enseignement secondaire. Ensuite, le recrutement volontaire peut avoir attiré des participants particulièrement engagés ou sensibilisés aux enjeux de l'évaluation, ce qui peut influencer la portée des constats émis précédemment. La collecte des données, réalisée à distance dans le contexte pandémique, a permis d'assurer la sécurité des participants, mais pourrait avoir teinté certains propos en lien avec les adaptations spécifiques à ce contexte. Enfin, bien que l'approche méthodologique rigoureuse ait permis de trianguler les méthodes utilisées, les résultats reposent sur des pratiques déclarées, sans observation directe de leurs pratiques évaluatives. Nous suggérons donc que d'autres chercheurs s'intéressent au processus d'évaluation dans sa globalité en adoptant une méthodologie davantage centrée sur les pratiques observées lors des situations d'évaluation en ÉPS. Une meilleure connaissance de ce processus permettrait à la communauté scientifique d'établir des collaborations avec des enseignants d'ÉPS afin de développer et d'expérimenter des stratégies simples et efficaces pour répondre à certaines difficultés issues de nos résultats.

Références

- Alvarez-Bueno, C., Pesce, C., Cavero-Redondo, I., Sánchez-López, M., Martínez-Hortelano, J. A. et Martínez-Vizcaíno, V. (2017). The effect of physical activity interventions on children's cognition and metacognition: A systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, 56(9), 729–738. <https://doi.org/10.1016/j.jaac.2017.06.012>
- Blömeke, S. (2017). Assuring quality in competence assessments: The value added of applying different assessment approaches to professional education. Dans M. Mulder (dir.), *Competence-based vocational and professional education* (p. 631–648). Springer. https://doi.org/10.1007/978-3-319-41713-4_29
- Borghouts, L. B., Slingerland, M. et Haerens, L. (2016). Assessment quality and practices in secondary PE in the Netherlands. *Physical Education and Sport Pedagogy*, 22(5), 473–489. <https://doi.org/10.1080/17408989.2016.1241226>
- Brau-Antony, S. et Hubert, F. (2014). Curriculum en éducation physique et sportive et évaluation certificative au baccalauréat. *Questions Vives*, (22), 1–16. <https://doi.org/10.4000/questionsvives.1596>
- Chen, S., Kim, Y. et Gao, Z. (2014). The contributing role of physical education in youth's daily physical activity and sedentary behavior. *BMC Public Health*, 14, article 110. <https://doi.org/10.1186/1471-2458-14-110>
- Creswell, J. W. et Creswell, J. D. (2022). *Research design: Qualitative, quantitative, and mixed methods approaches* (6^e éd.). SAGE.
- De Ketele, J.-M. (2010). Ne pas se tromper d'évaluation. *Revue française de linguistique appliquée*, 15(1), 25–37. <https://doi.org/10.3917/rfla.151.0025>
- Dinan-Thompson, M. et Penney, D. (2015). Assessment literacy in primary physical education. *European Physical Education Review*, 21(4), 485–503. <https://doi.org/10.1177/1356336X15584087>
- Dubuc, M.-M., Aubertin-Leheudre, M. et Karelis, A. D. (2022). Predictors of academic performance in high school students: The longitudinal ASAP study. *International Journal of Exercise Science*, 15(4), 615–630. <https://doi.org/10.70252/zazq9758>

- Duda, J. L., Appleton, J. J., Stebbings, J. et Balaguer, I. (2018). Towards more empowering and less disempowering environments in youth sport. Dans C. J. Knight, C. G. Harwood et D. Gould (dir.), *Sport psychology for young athletes* (p. 81–93). Routledge.
- Durand, M.-J. et Chouinard, R. (2012). *L'évaluation des apprentissages : de la planification de la démarche à la communication des résultats*. Marcel Didier.
- Fletcher, T. et Mandigo, J. (2008). *Physical education and health: Global perspectives and best practice*. Human Kinetics.
- Fontaine, S. et Loye, N. (2017). L'évaluation des apprentissages : une démarche rigoureuse. *Pédagogie Médicale*, 18(4), 189–198. <https://doi.org/10.1051/pmed/2018013>
- Georgakis, S. et Wilson, R. (2012). Australian physical education and school sport: An exploration into contemporary assessment. *Asian Journal of Exercise & Sports Science*, 9(1), 37–52.
- Govaerts, M. J. B. (2016). Competence in assessment: Beyond cognition. *Medical Education*, 50(5), 502–504. <https://doi.org/10.1111/medu.13000>
- Graham, G., Holt/Hale, S. A., Parker, M., Hall, T. et Patton, K. (2020). *Children moving: A reflective approach to teaching physical education* (10^e éd.). McGraw-Hill.
- Hirsch, S. (2017). *Measurement in education: Test score and beyond* [Thèse de doctorat, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20170111sh>
- Kirk, D. (2010). *Physical education futures*. Routledge.
- Lee, C.-j. (2014). In search of subjectivity: A reflection of a teacher educator in a cross-cultural context. *Educational Philosophy and Theory*, 46(13), 1427–1434. <https://doi.org/10.1080/00131857.2014.964165>
- Leirhaug, P. E. et Annerstedt, C. (2016). Assessing with new eyes? Assessment for learning in Norwegian physical education. *Physical Education and Sport Pedagogy*, 21(6), 616–631. <https://doi.org/10.1080/17408989.2015.1095871>

- López-Pastor, V. M., Kirk, D., Lorente-Catalán, E., MacPhail, A. et Macdonald, D. (2013). Alternative assessment in physical education: A review of international literature. *Sport Education and Society*, 18(1), 57–76. <https://doi.org/10.1080/13573322.2012.713860>
- Lund, J. L. et Veal, M. L. (2008). Measuring pupil learning – How do student teachers assess within instructional models? *Journal of Teaching in Physical Education*, 27(4), 487–511. <https://doi.org/10.1123/jtpe.27.4.487>
- Lunenberg, M., Korthagen, F. et Swennen, A. (2007). The teacher educator as a role model. *Teaching and Teacher Education*, 23(5), 586–601. <https://doi.org/10.1016/j.tate.2006.11.001>
- Lyngstad, I., Bjerke, Ø., Bang, K. M. et Ligestad, P. (2022). Norwegian upper secondary students' experiences of their teachers' assessment of and for learning in physical education: Examining how assessment is interpreted by students of different physical abilities. *Sport, Education and Society*, 27(3), 320–331. <https://doi.org/10.1080/13573322.2020.1842728>
- Marmeleira, J., Folgado, H., Guardado, I. M. et Batalha, N. (2020). Grading in Portuguese secondary school physical education: Assessment parameters, gender differences and associations with academic achievement. *Physical Education & Sport Pedagogy*, 25(2), 119–136. <https://doi.org/10.1080/17408989.2019.1692807>
- Ministère de l'Éducation du Québec [MEQ]. (2025). *Programme éducatif CAPS-II – Compétences axées sur la participation sociale. Guide de soutien en évaluation des apprentissages*. <https://cdn-contenu.quebec.ca/cdn-contenu/education/soutien-eleves/Guide-Programme-CAPS-II.pdf>
- Ministère de l'Éducation, du Loisir et du Sport [MELS]. (2006). *Programme de formation de l'école québécoise. Éducation préscolaire Enseignement primaire*. <https://numerique.banq.qc.ca/patrimoine/details/52327/55029>
- Ministère de l'Éducation, du Loisir et du Sport [MELS]. (2009). *Progression des apprentissages au primaire. Éducation physique et à la santé*. <https://cdn-contenu.quebec.ca/cdn-contenu/education/pfeq/primaire/progressions-apprentissages/PFEQ-progression-apprentissages-education-physique-primaire.pdf>

- Ministère de l'Éducation, du Loisir et du Sport [MELS]. (2011). *Cadre d'évaluation des apprentissages en éducation physique et à la santé – Enseignement primaire 1^{er}, 2^e et 3^e cycle*. https://www.education.gouv.qc.ca/fileadmin/site_web/documents/dpse/evaluation/education-physique-et-a-la-sante-pri.pdf
- Morrisette, J. et Legendre, M.-F. (2012). L'évaluation des compétences en contexte scolaire : des pratiques négociées. *Education Sciences and Society*, 2(2), 120–132. <http://hdl.handle.net/1866/21528>
- Mougenot, L. (2016). Régulation différée et apprentissage en sport collectif. *Staps*, 111(1), 29–41. <https://doi.org/10.3917/sta.111.0029>
- Ní Chróinín, D. et Cosgrave, C. (2013). Implementing formative assessment in primary physical education: Teacher perspectives and experiences. *Physical Education and Sport Pedagogy*, 18(2), 219–233. <https://doi.org/10.1080/17408989.2012.666787>
- Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice*. SAGE.
- Penney, D., Brooker, R., Hay, P. J. et Gillespie, L. (2009). Curriculum, pedagogy and assessment: Three message systems of schooling and dimensions of quality physical education. *Sport, Education and Society*, 14(4), 421–442. <https://doi.org/10.1080/13573320903217125>
- Perez-Roux, T. (2011). *Identité(s) professionnelle(s) des enseignants : les professeurs d'ÉPS entre appartenance et singularité*. Éditions EP&S.
- Sofo, S., Ocansey, R. T., Nabie, M. J. et Asola, E. F. (2013). Assessment practices among secondary physical education teachers in Ghana. *IOJES*, 5(2), 274–281. https://iojes.net/?mod=tammetin&makaleadi=&makaleurl=IOJES_799.pdf
- Svennberg, L., Meckbach, J. et Redelius, K. (2018). Swedish PE teachers struggle with assessment in a criterion-referenced grading system. *Sport, Education and Society*, 23(4), 381–393. <https://doi.org/10.1080/13573322.2016.1200025>
- Tolgfors, B. (2018). Different versions of assessment for learning in the subject of physical education. *Physical Education and Sport Pedagogy*, 23(3), 311–327. <https://doi.org/10.1080/17408989.2018.1429589>

- Tolgfors, B., Quennerstedt, M., Backman, E. et Nyberg, G. (2022). Enacting assessment for learning in the induction phase of physical education teaching. *European Physical Education Review*, 28(2), 534–551. <https://doi.org/10.1177/1356336X211056208>
- Van der Mars, H., McNamee, J. et Timken, G. (2018). Physical education meets teacher evaluation: Supporting physical educators in formal assessment of student learning outcomes. *The Physical Educator*, 75(4), 581–615. <https://doi.org/10.18666/TPE-2018-V75-I4-8471>
- van Rossum, T. et Morley, D. (2018). The role of digital technology in the assessment of children’s movement competence during primary school physical education lessons. Dans J. Koekoek et I. van Hilvoorde (dir.), *Digital technology in physical education: Global perspectives* (p. 39–51). Routledge. <https://doi.org/10.4324/9780203704011>
- Williams, J., Pill, S. et Hewitt, M. (2020). ‘I think everyone is on board with changing how we do things, but we are yet to find a best fit model’: A figural study of assessing games and sport in physical education. *Sport, Education and Society*, 26(3), 253–266. <https://doi.org/10.1080/13573322.2020.1716212>