

Annabelle GRANDCHAMP
Jacques MÉARD
Grégory QUIN

ANALYSE DE L'ÉVOLUTION DES PRATIQUES ÉVALUATIVES EN ÉDUCATION PHYSIQUE

REVUE DE LITTÉRATURE

Résumé : Les pratiques d'évaluation en éducation physique et sportive (EPS) ont été largement analysées et discutées dans les problématiques éducatives des pays industrialisés. Loin d'obtenir un consensus conceptuel, les contradictions et les tensions persistent notamment entre les concepts théoriques et les pratiques des enseignants. Cette revue de littérature tente dans un premier temps de clarifier les ambiguïtés intrinsèquement liées à l'évaluation en EPS en proposant un rappel des définitions et des différents objets de l'évaluation. Dans un deuxième temps, cet article présente les évaluations alternatives apparues dans certains pays en réponse à une insatisfaction en termes d'apprentissage des évaluations traditionnelles. Nous verrons toutefois que ces orientations plus formatrices peinent à se généraliser. Enfin, nous tenterons de dégager quelques hypothèses permettant d'expliquer la relative continuité des pratiques évaluatives en EPS.

Mots-clés : éducation physique, évaluation sommative, évaluation formative, réforme, contrôle.

INTRODUCTION

Les pratiques d'évaluation continuent d'alimenter de nombreux débats dans les systèmes éducatifs des pays industrialisés (López-Pastor, Kirk, Lorente-Catalan, MacPhail & Macdonald, 2013). Bien que cet objet ait été largement analysé et discuté, la littérature ne témoigne pas d'un consensus conceptuel (De Ketele, 2012), mais au contraire de nombreuses contradictions entre les concepts théoriques et les pratiques des enseignants (Lau, 2016). Une discipline illustre ce paradoxe mieux qu'aucune autre, l'éducation physique et sportive (EPS) où de nombreuses réformes et prescriptions à propos des examens (Avery, 2012 ; Holt/Hale & Persse, 2015 ; Larsson & Nyberg, 2017 ; Leirhaug & Annerstedt, 2015 ; Leirhaug, MacPhail & Annerstedt, 2016 ; MacPhail & Murphy, 2017) montrent cette absence de consensus et les confusions qui persistent dans les conceptions des enseignants et des politiques (López-Pastor *et al.*, 2013). Certaines spécificités de l'EPS en font sans doute un lieu où les tensions et les contradictions liées à l'évaluation sont exacerbées : sans parler du statut secondaire de cette discipline, le fait par exemple que l'évaluation s'y réalise « à chaud » (c'est-à-dire sans traces de productions d'élèves que l'enseignant peut analyser à froid) (Mougenot, 2013) ou le fait que les élèves performants le soient la plupart du temps parce qu'ils ont construits des compétences en dehors de l'école (ce qui est beaucoup moins le cas en mathématiques, sciences, histoire, etc.) amplifient les effets potentiellement pervers de la certification et attisent les discussions sur la

Spiral-E - Revue de Recherches en Éducation – 2018
Supplément électronique au N° 61 (3-17)

place à accorder à une évaluation « pour apprendre » (Méard & Grandchamp, 2018). Comme nous l'évoquerons plus loin, la discipline EPS partage sur ces deux points des préoccupations communes avec d'autres enseignements telles que celui de la technologie, de la musique, des compétences de communication en langues étrangères (Aubin, 2010) ou encore de l'oralité en langue première (Simard, Dufays, Dolz-Mestre & Garcia-Debanc, 2010 : 283 *seq.*).

La revue de littérature que nous présentons ici tente donc d'abord de clarifier certaines ambiguïtés intrinsèquement liées à l'évaluation en EPS au travers des travaux scientifiques les plus récents. Mais une seconde question se pose : celle de la pérennité des pratiques évaluatives elles-mêmes, notamment par la prédominance de formes traditionnelles d'évaluation, reposant sur les performances ou les habiletés motrices. Certains auteurs relèvent, au travers des recherches contemporaines, une homogénéité de ces pratiques quels que soient les contextes (Hay & Penney, 2009), ce qui rend compte de la difficulté de l'EPS dans le monde à changer, malgré les nombreuses critiques à l'égard des pratiques évaluatives existantes (López-Pastor, Kirk, Lorente-Catalan, MacPhail & Macdonald, 2013). Le second objectif de la synthèse présentée ici consiste donc à pointer dans la littérature récente les travaux portant sur les évaluations alternatives en EPS et sur l'impact de ces pratiques sur l'enseignement, les élèves et leurs apprentissages.

Enfin, il nous semble utile de prendre en compte de façon exhaustive non seulement les travaux anglophones sur ces questions mais aussi la littérature scientifique francophone qui échappe aux moteurs de recherche des revues précédentes (López-Pastor *et al.*, 2013). Cela nous a incités d'abord à rechercher dans la littérature francophone (grâce aux moteurs de recherche Cairn et *Google Scholar*) les publications portant sur l'évaluation en EPS à partir des mots-clés suivants : « évaluation », « évaluation sommative », « évaluation formative », « équité », « arrangements évaluatifs ». Puis, dans un deuxième temps, nous avons appliqué la même méthode (grâce aux moteurs de recherche *Web of Science*, *Science Direct*, *Taylor & Francis*, *Google Scholar*) pour la littérature anglophone à partir des mots-clés suivants : « *evaluation* », « *assessment* », « *physical education* ». Notre choix temporel s'est porté sur les 12 dernières années marquées par de multiples réformes en Europe (France, Norvège, Suède, Angleterre, Irlande et Turquie) ainsi que sur d'autres continents (États-Unis, Australie). En discussion, nous avons néanmoins eu recours à des auteurs *princeps* plus anciens pour éclairer notre synthèse. Cette démarche nous a permis d'identifier et analyser 95 articles et 4 ouvrages et rapports relatifs au monde francophone (38) anglophone (61). Pour des raisons de format, nous avons retenu dans cet article les 55 travaux qui nous semblaient les plus représentatifs de l'ensemble des problématiques du sujet, ainsi que 12 travaux non spécifiques à l'EPS, touchant à l'évaluation scolaire et constituant des apports importants pour notre réflexion.

Après avoir rappelé les différentes acceptions du mot « évaluation », nous pointerons dans un deuxième temps la persistance des tensions entre les différentes fonctions évaluatives en EPS, notamment l'omniprésence des aspects certificatifs dans les pratiques évaluatives. Dans un troisième temps, nous montrerons que les critiques récentes relatives aux pratiques dominantes s'accompagnent de travaux pour analyser ou expérimenter des formes alternatives d'évaluation en EPS. Ces trois temps nous conduiront enfin à proposer une synthèse de ces travaux et à proposer de nouvelles pistes de recherche.

LES DIFFÉRENTES FONCTIONS DE L'ÉVALUATION

L'évaluation fait partie du processus enseignement-apprentissage et il est impossible qu'un apprentissage ait lieu sans évaluation. Gérard (2013) analyse l'étymologie du mot « évaluer » dont les racines indo-européennes et latines lui donnent le sens de « *faire sortir la valeur de ce qu'on évalue, en montrer la force et la puissance* ». Evaluer consiste donc à apprécier la valeur d'une chose, en l'occurrence d'une prestation d'élève en EPS. Cette « appréciation » prend la forme d'une « information en retour » mais, derrière cette définition, les différentes pratiques qui sont rassemblées sous le vocable d'évaluation sont diverses selon les destinataires de cette « information en retour » et selon la fonction que l'on cherche à lui faire jouer. Cette diversité et cette hétérogénéité des pratiques sont liées à ces différentes fonctions possibles. De Ketele (2010) distingue au moins trois fonctions principales de l'évaluation :

- la fonction d'orientation : évaluer pour orienter une nouvelle action à entreprendre ;
- la fonction de régulation : évaluer pour améliorer une action en cours. Cette fonction est mieux connue dans le monde scolaire sous l'appellation d'évaluation formative, concept créé par Scriven en 1967 ;
- la fonction de certification : évaluer pour certifier socialement une action considérée comme terminée.

Il semblerait donc que la confusion entre évaluation certificative et évaluation formative persiste depuis une cinquantaine d'années (elle était repérée déjà par Scriven (1967) et Bloom (1971)). Les travaux de recherche en docimologie ont ainsi porté essentiellement sur les notes comme échelles de mesure des performances. Ceci a eu pour conséquence qu'aux yeux de nombre de praticiens, « l'évaluation se résumait à attribuer une note fiable pour sanctionner la réussite des apprentissages » (De Ketele, 2010). Cette conception n'est pas sans risque car elle consiste à réduire l'évaluation à la seule fonction de « sanction ». De plus, si tous les auteurs s'accordent sur le fait que le processus évaluatif consiste à confronter un référent à un référé, ils divergent sur la finalité du processus : pour les uns, il s'agit de « porter un jugement sur » ; pour les autres, de « fonder une prise de décision ».

LES ENJEUX CONTRADICTOIRES DE L'ÉVALUATION

La littérature reconnaît généralement l'importance de l'évaluation en EPS dans un but d'efficacité et d'équité de l'enseignement. Mais des études européennes mettent en lumière la confusion qui règne au sein de ces pratiques évaluatives, entre évaluation sommative et formative (Annerstedt & Larsson, 2010 ; Mougenot, 2013 ; 2016). Si la langue française utilise le même concept pour les deux types d'évaluation, les anglophones parlent d'*assessment* lorsque l'évaluation se rapporte aux apprentissages, et d'*evaluation*, lorsqu'elle porte sur les dispositifs (Allal, 2012 ; Collier, 2011). La relation entre formatif et sommatif n'est toutefois pas clairement définie dans la littérature de recherche (Stobart, 2011). Les débats entre les fonctions régulatrices de l'évaluation et celles normatives du contrôle se perpétuent.

La différence entre les intentions et les évaluations

Dans les textes officiels, les procédures d'évaluation sont censées ne pas avoir la même fonction ni être mises en œuvre au même moment (Lentillon-

Kaestner, Allain, Deriaz, & Voisard, 2018). Mais dans la pratique, ces distinctions ne sont pas toujours aussi clairement observées. Il existe un décalage entre les intentions des enseignants et ce qu'ils mettent en œuvre (Bourghouts, Slingerland & Haerens, 2017). Selon Mougenot et Dugas (2014), si 97 % des enseignants d'EPS interrogés disent utiliser l'évaluation formative, régulièrement ou non, seuls 11 % d'entre eux la proposent lors de chaque cycle d'apprentissage sans jamais la noter. Certains notent des évaluations qu'ils présentent pourtant comme formatives (Mougenot, 2013). Apparaissent donc régulièrement des tensions entre évaluation sommative et formative, entre une logique de *contrôle* permettant d'assurer un principe d'équité et une régulation *dialogique* s'inscrivant dans une logique de développement et d'apprentissage (Mottier Lopez, 2009). Dans la littérature internationale, l'évaluation formative en EPS est souvent présentée comme plus utile pour aider les élèves à progresser que l'évaluation sommative, par exemple 70 % des enseignants français (Mougenot & Dugas, 2014 : 72). La relation entre évaluation sommative et formative donne lieu à des avis contradictoires : si certains auteurs prétendent qu'une distinction entre ces deux types d'évaluation doit être maintenue, d'autres estiment que l'évaluation sommative peut aussi favoriser l'apprentissage (Leirhaug, 2016). Selon Lau (2016), il est temps de dépasser le clivage entre le « mauvais » sommatif et le « bon » formatif. Mottier Lopez (2009) insiste de son côté sur « la légitimité de chaque logique, y compris celle de contrôle » et précise que l'enjeu est la cohérence du sens et des utilisations faites des différentes formes d'évaluation. Les problèmes semblent donc liés à un manque de clarté quant à ce qu'on souhaite évaluer et pourquoi une telle évaluation est retenue.

La question de l'objet des évaluations en EPS interfère dans ce débat, notamment depuis la généralisation du concept de compétence (Genelot, Gardes, Mansanti, & Pinsard, 2016). En effet, définie comme un savoir complexe intégrant des connaissances, des attitudes, des savoir-faire à mobiliser, la compétence ne peut plus être appréhendée par des tâches simplistes (De Ketele, 2010). En EPS, l'évaluation de compétences dépasse donc largement la seule maîtrise d'habiletés motrices et implique la prise en compte du temps de l'apprentissage et du caractère généralisable du savoir au travers de multiples situations, ce qui accentue encore la confusion entre les différentes fonctions de l'évaluation.

Des notes pour évaluer les apprentissages

Evaluer consiste à apprécier la valeur d'une prestation d'élèves en EPS dans plusieurs situations, ce qui rend inadéquate l'évaluation des compétences en EPS par une note (Collier, 2011) car celle-ci ne permet pas de prendre en compte tous les mécanismes d'apprentissage des élèves. Et en EPS, les travaux en didactologie ont porté essentiellement sur les notes comme échelle de mesure des performances, ce qui a eu pour conséquence qu'aux yeux de nombreux praticiens, l'évaluation se résumait à attribuer une note fiable pour sanctionner la réussite des apprentissages. Or, cette conception réductrice du processus évaluatif n'est pas sans risque et le travail scientifique pointe les risques d'arbitraire et de subjectivité dans la notation en EPS (Annerstedt & Larsson, 2010 ; Zhu, 2015). Autrement dit, l'évaluation objective n'existe pas en EPS (Lentillon-Kaestner, 2008). De plus, la plupart des études montre que la note, en tant que source de motivation extrinsèque, ne suscite pas l'envie de réussir mieux que les autres et limite l'investissement et les performances en jouant négativement sur les buts d'accomplissement (Pulfrey, Buchs & Butera, 2011). En France, des actions en vue de supprimer la note se sont d'ailleurs multipliées cette dernière décennie (Genelot *et al.*, 2016).

Ainsi, lors d'une évaluation en EPS expurgée de sa dimension certificative, les élèves ressentent davantage l'aide de l'enseignant, leurs relations s'améliorent, ceux qui sont plus faibles se sentent moins discriminés (Leirhaug & MacPhail, 2015). Sur ce point, Allain, Deriaz, Voisard et Lentillon – Kaestner (2016) se sont intéressés au cas particulier de l'EPS dans le Canton de Vaud en Suisse, où les élèves ne sont pas notés. Ces auteurs confirment que, dans ces conditions, les élèves, surtout les moins performants, s'y sentent davantage soutenus par leurs enseignants que dans le Canton de Genève (où l'EPS est notée) (Allain, Deriaz, Voisard & Lentillon–Kaestner, 2016).

L'omniprésence du certificatif

Des critiques émergent régulièrement à l'égard des évaluations traditionnelles en EPS, à l'instar des tests de condition physique, qui ne répondraient plus aux exigences pédagogiques actuelles visant l'éducation de l'enfant dans son entier (*holistic approach*) (Dyson, 2014 ; Lopez–Pastor *et al.*, 2013). Un enseignement de qualité impliquerait une évaluation inscrite dans une perspective socio-culturelle critique (*social-critical perspective*). Or, malgré cette littérature étendue (Butera *et al.*, 2011), force est de constater que l'évaluation sommative reste dominante (Braun-Antony & Hubert, 2014 ; Cogérino & Mnaffakh, 2008). De nombreuses études révèlent que les enseignants d'EPS recourent plus fréquemment à des évaluations traditionnelles basées sur des connaissances ou des performances, plus faciles à noter et laissant apparemment moins de place à la subjectivité (Constantinou, 2017 ; Mougenot, 2013). Au Royaume-Uni par exemple, près de 90 % des enseignants utilisent les *Physical Fitness Tests (PFTs)* dans leur programme (López-Pastor *et al.*, 2013).

Par ailleurs, l'analyse des pratiques évaluatives en EPS montrent que la plupart des évaluateurs de terrain confondent l'évaluation sommative et certificative (De Ketele, 2010). Cette dernière fait référence au jugement d'un produit ou d'une opération à partir d'informations et est utilisée pour certifier socialement ou sélectionner (Richard, 2016). Il s'agit donc d'une fonction en réponse à une finalité permettant de définir le « pourquoi » de l'évaluation. Le concept d'évaluation sommative se réfère plutôt au « comment », relevant d'une démarche qui consiste à attester de la maîtrise ou non des acquis des élèves à un moment précis (Lentillon–Kaestner *et al.*, 2018). Elle n'induit donc pas toujours la mise en place d'un système de notation, réduit à la seule fonction de « sanction » dans le but de porter un jugement. Ce malentendu entre certificatif et sommatif incite à opposer « évaluation sommative » à « évaluation formative » (Mougenot, 2013 ; Mougenot & Dugas, 2014). Mais au vu de ce qui précède, une évaluation sommative pourrait être réalisée sous forme d'un contrôle en cours d'apprentissage et répondre à des fonctions d'orientation (préparer une nouvelle action) ou de régulation. Ce raisonnement apparaît déterminant dans le contexte de l'évaluation en EPS lorsque certains pays proposent des évaluations sommatives non certificatives (Allain *et al.*, 2016). L'association abusive et fréquente entre évaluation sommative et certificative met le doigt sur un des nœuds du système et permet peut-être de mieux comprendre les contradictions chez les élèves et les enseignants (Cogérino, 2009). Cette confusion est souvent mise en rapport chez les auteurs avec les dilemmes ressentis par les enseignants lorsqu'ils doivent mettre une note en EPS (Braun-Antony & Hubert, 2014 ; Brown & Penney, 2016 ; Cogérino, 2009 ; Cogérino & Mnaffakh, 2008). Là encore, les dispositifs semblent dépendre de l'objet à évaluer. Et les auteurs soulignent l'importance d'une réflexion en amont afin d'identifier la fonction d'une évaluation (De Ketele, 2010).

Les limites des évaluations « traditionnelles » en EPS

Des auteurs soulignent qu'une EPS de qualité doit articuler *curriculum, pedagogy and assessment* (contenu, pédagogie et évaluation) (Penney, Brooker, Hay & Gillepsie, 2009). Or, cet alignement n'est pas toujours effectif et les pratiques évaluatives des enseignants ne sont pas toujours en accord avec les programmes scolaires en vigueur (Nazari, Jafari, Nasr & Marandi, 2017 ; Oh & Graber, 2017 ; Redelius & Hay, 2012).

Sur ce point, la question de la prise en compte ou non de facteurs sociaux et affectifs dans l'évaluation en EPS est au cœur des débats (Collier, 2011 ; Glennon, Hart & Foley, 2015). Les spécificités de cette discipline font naître des avis divergents au sein de la communauté pédagogique et scientifique : les enseignants d'EPS se préoccuperaient davantage de l'équité (justice distributive) tenant compte des différences entre élèves que d'une égalité (justice procédurale) préconisée dans les textes officiels (Cogérino & Mnaffakh, 2008), même si l'étude de Lentillon-Kaestner (2014) sur les injustices perçues par les élèves en EPS nuance ce constat en pointant que les élèves sont aussi sensibles à l'égalité. De même, pour certains enseignants d'EPS, les facteurs tels que le comportement, l'effort, la participation et les progrès ne peuvent être considérés comme des objectifs d'apprentissage évalués et notés (Melograno, 2007). D'autres avancent au contraire que l'évaluation en EPS ne doit pas se limiter à des facteurs cognitifs et moteurs, pour assurer l'égalité des chances face aux pratiques physiques scolaires fortement sexuées (Glennon *et al.*, 2015 ; Lentillon-Kaestner, 2014), aux différences morphologiques entre élèves et aux écarts d'activités physiques extra-scolaires (Collier, 2011). L'EPS apparaît comme bien davantage qu'une activité physique, elle permet la transmission de valeurs, par le recours au jeu notamment (Dyson, 2014). Ainsi, les standards américains *National Standards & Grade-Level Outcomes for K – 12 Physical Education* définis par *Society of Health And Physical Educators* en 2014 (SHAPE America, 2014) incluent le domaine affectif et émotionnel. Une récente étude américaine montre toutefois les dérives liés à la prise en compte de ces standards à propos desquels les élèves perçoivent que la notation est influencée davantage par la participation ou le comportement que par les aptitudes physiques (Zhu, 2015). Cette étude met en lumière le manque de connaissances des élèves en ce qui concerne les attentes fixées par les normes en EPS, ce qui peut engendrer une démotivation pour l'apprentissage. Cela plaide encore en faveur d'un alignement entre pratiques évaluatives et standards nationaux, grâce à des outils adaptés (Holt/Hale & Persse, 2015 ; Lorson & Mitchell, 2016 ; Martin, Horton & Tarr, 2015 ; Oh & Graber, 2017).

ÉVALUER AVEC DES APPROCHES ALTERNATIVES

Dans ce contexte international homogène, l'intérêt de la recherche pour de nouvelles approches évaluatives va croissant (Lorente-Catalan & Kirk, 2014). Aux États-Unis, en Norvège, Grande-Bretagne, Irlande, Australie et Espagne, ces tentatives « alternatives » ouvrent la voie à une évaluation plus formative, voire « formatrice » dans laquelle les élèves sont davantage responsabilisés et impliqués dans leurs apprentissages (Ni Chròinin & Cosgrave, 2013 ; O'Connor, Jeanes & Alfrey, 2016).

Un faisceau de tentatives alternatives

Le choix de *Assessment for learning (AfL)* (évaluation pour les apprentissages) est présenté comme un moyen primordial pour améliorer les apprentissages

des élèves, ainsi que les pratiques évaluatives des enseignants d'EPS (Leirhaug & MacPhail, 2015). En contexte francophone, cette orientation prend de l'importance dans les travaux de l'Association pour le Développement des Méthodologies d'Évaluation en Éducation (ADMEE), mais semble peu répandue dans les pratiques scolaires, contrairement aux pays anglo-saxons (De Ketele, 2012). Dans cette approche basée sur les théories constructivistes, les élèves deviennent acteurs de leur apprentissage (MacPhail & Halbert, 2010). Ils sont censés construire leurs connaissances dans l'interaction avec les pairs par le biais de la co-évaluation et de l'auto-évaluation (Arslan, Erturan Ilker & Demirhan, 2013). Dans une démarche globale visant leur autonomie, ils seraient plus impliqués et assimileraient mieux les critères d'évaluation. Les travaux de Black et William en 1998 ont mis en évidence les effets positifs de l'*AfL* sur les apprentissages des élèves. L'*AfL* n'exclut pas les évaluations sommatives, ni la notation dans la mesure où celle-ci est valide et socialement juste (Hay & Penney, 2009). Une étude réalisée en Norvège sur les effets positifs attendus de cette approche sur les notes des élèves vient corroborer cette affirmation : en l'absence d'alignement entre les pratiques formatives et sommatives (notation), la mise en place de l'*AfL* n'influence pas les notes des élèves (Leirhaug, 2016).

Bien que de nombreux chercheurs reconnaissent que cette approche peut améliorer significativement les apprentissages des élèves, son application pratique peut générer des difficultés pour les enseignants et un certain nombre de prérequis sont nécessaires (Arslan *et al.*, 2013). D'autres études américaines et néerlandaises soulignent les changements philosophiques et pédagogiques en cours, ainsi que le besoin de mettre en place des moyens d'évaluation fiables qui correspondent aux standards nationaux et concluent à la nécessité d'un accompagnement des enseignants (Avery, 2012 ; Collier, 2011 ; Rink, Jones, Kirby, Mitchell & Douthett, 2007). Mais les exemples et les descriptions sur la manière de mettre en place une *AfL* sur le terrain sont encore rares (Leirhaug & MacPhail, 2015).

Inspirée par le modèle *Sport Education* de Siedentop, une autre orientation alternative en matière d'évaluation en EPS est *Authentic assessment* (Évaluation authentique) (Wallhead & O'sullivan, 2005). Le but est ici de développer la responsabilisation des élèves, leur confiance en soi et leur coopération par un jeu d'interactions. Une récente étude américano-portugaise dans le cadre d'un programme en volleyball sur trois saisons révèle qu'il est préférable d'implanter des séquences de longue durée dans une approche authentique (Araújo, Hastie, Lohse, Bessa & Mesquita, 2017). Ainsi en respectant les caractéristiques authentiques d'un sport, les évaluations associées à cette approche sont supposées agir sur le plaisir de l'élève (Collier, 2011 ; Dyson, 2014), ce qui aurait un effet positif sur l'engagement et la motivation intrinsèque plus durable des élèves (Hay & Penney, 2009 ; Leirhaug & MacPhail, 2015 ; Spittle & Byrne, 2009), en les orientant vers des buts de maîtrise plutôt que vers des buts de performance ou de reconnaissance sociale. Le modèle *Teaching Games for Understanding model (FGfU)*, développé au Royaume-Uni, est un exemple de dispositif qui s'inscrit dans cette approche. Une étude australienne met en évidence l'impact sur l'intérêt des élèves d'une méthode de travail appelée *Take Action*, basée sur la collaboration et supposée favoriser l'engagement critique des élèves selon une approche authentique (O'Connor *et al.*, 2016).

Basées sur l'idée que l'évaluation doit être intégrée au processus enseignement-apprentissage, ces initiatives trouvent leurs racines dans les travaux américains, mais aussi québécois (Desrosiers) et français (Amade-Escot) des années quatre-vingt-dix qu'on peut regrouper sous le vocable d'*Integrated Assessment*

(évaluation intégrée) (Hay & Penney, 2009). Ils introduisent le principe d'une *Assessment culture* qui remplace la *Test culture* qui se référait davantage à la mesure et la notation des connaissances et des habiletés motrices.

Limites de ces tentatives alternatives

Ces différentes orientations alternatives convergent vers le projet d'une évaluation motivante, empreinte de sens et « responsabilisante ». Fortes de leur orientation formative, voire formatrice, ces évaluations ouvrent des perspectives dans le but d'améliorer l'enseignement de l'EPS, même si leur implantation dans les pratiques rencontre un certain nombre d'obstacles.

En effet, les limites des évaluations alternatives en EPS sont relevées par deux études australiennes : Hay et Penney (2009) soulignent que, pour être efficace, une évaluation en EPS doit être non seulement juste et en accord avec les programmes, mais également orientée vers les apprentissages, réalisée dans un contexte global et validée par des outils fiables. L'interdépendance de ces quatre facteurs est donc fondamentale (*ibid.* : 400). Les réformes de ces dernières années dans de nombreux pays sont certainement apparues en réponse à ces préoccupations. De même, Georgakis & Wilson (2012) relèvent des divergences entre les objectifs d'apprentissages en EPS et les points de vue des enseignants en ce qui concerne l'évaluation. Les auteurs affirment que malgré les changements réalisés depuis les années quatre-vingt-dix dans le domaine de l'évaluation en éducation, peu de recherches se sont intéressées à observer les répercussions de ces innovations sur les pratiques d'évaluation en EPS. Les résultats de cette étude montrent que le manque d'alignement entre les programmes et les pratiques limite fortement les possibilités d'exploiter les avantages potentiels de l'*AfL*. L'implantation d'une évaluation alternative réussie exige du temps et des ressources et les enseignants doivent y être préparés et accompagnés (López-Pastor *et al.*, 2013).

Ces limites sont particulièrement visibles pour ce qui concerne l'usage des nouvelles technologies de l'information et de la communication (NTIC) pour évaluer en EPS. En effet, les travaux sur ce sujet sont nombreux, notamment sur l'utilisation des tablettes numériques qui permettent à l'élève d'obtenir une information en retour, une trace, à chaud sur sa prestation ou celle d'un autre élève (Harvey & Gittins, 2014 ; Palao, Hastie, Cruz & Ortega, 2015). Ils soulignent le potentiel de ces évaluations alternatives pour qu'il apprenne, pour augmenter son engagement en l'impliquant dans des auto et des co-évaluations (Casey & Jones, 2011 ; Lu-Ho, Iwen, & Gwo-Jen, 2016). Mais ces travaux évaluent peu les effets réels de ces NTIC et ne sont pas prolongés par une meilleure connaissance de ces NTIC par les enseignants (Kretschmann, 2015) ni par leur massification dans les pratiques de terrain (Cain, 2010).

HYPOTHÈSES À PROPOS D'UN CONSERVATISME

On réalise à la lecture de ces travaux récents sur l'évaluation en EPS que les pratiques traditionnelles restent bien présentes malgré la masse d'études tendant à montrer leurs effets non souhaitables. Il y a donc lieu de se demander pourquoi ces approches n'ont pas réussi à gagner l'ensemble de la communauté internationale de l'EPS. De notre point de vue, la zone d'ombre qui reste à éclairer concerne « l'activité de l'enseignant qui évalue ».

Le relatif conservatisme des pratiques évaluatives est d'abord à mettre en lien avec la complexité fortement sous-estimée d'une pratique scolaire pour laquelle une formation est nécessaire. L'ambiguïté de la posture de l'enseignant, à

la fois « contrôleur » et « conseiller », implique d'identifier clairement les domaines de compétences professionnelles qu'il doit acquérir en vue d'une évaluation utile en termes d'apprentissage. Dinan Thompson et Penney (2015) soulignent les lacunes en ce qui concerne les connaissances de l'évaluation face aux difficultés des pratiques évaluatives pour les enseignants. Les auteurs insistent dès lors sur le besoin de former les enseignants dans ce domaine et de développer leur esprit critique (Carnus & Terrisse, 2006). On parvient à un constat analogue à propos des enseignements et de l'évaluation en musique et en langues où sont juxtaposées des incitations de chercheurs en faveur d'approches communicationnelles et basées sur l'autoscopie ou la co-évaluation et des pratiques qui peinent à intégrer ces recommandations (Aubin, 2010 ; Simard *et al.*, 2010).

A ce sujet, il est frappant de constater que de nombreux enseignants d'EPS déclarent donner la priorité à l'évaluation-régulation alors qu'en fait, le sommatif est souvent omniprésent dans leur démarche. Tout se passe comme si le discours argumentatif de l'évaluation pour l'apprentissage, formative, authentique, était intégré mais ne pouvait surmonter les difficultés de l'action. Si l'on considère cette hypothèse d'une continuité des pratiques évaluatives du fait finalement de la complexité professionnelle à évaluer « pour les apprentissages », la discipline EPS apporte des contraintes surajoutées, au moins sur trois points. En effet, comme en musique, en technologie et en langues à l'oral, la prestation de l'élève dans cette discipline ne laisse pas souvent de trace (dans le meilleur des cas, une empreinte dans le sable du sautoir, un temps sur chronomètre en natation). Dans ces conditions, mettre en place une évaluation « pour apprendre » est sensiblement plus difficile que lors de la résolution d'un problème mathématique dont le résultat reste inscrit sur un papier que l'on peut analyser et faire analyser *a posteriori* : en EPS, si l'élève chute après avoir réalisé une figure gymnique, l'analyse doit se faire à chaud à partir du souvenir de la figure réalisée. L'arsenal à mettre en œuvre pour que celle-ci soit possible devient compliqué (définir un ou deux critères accessibles à l'élève ou à l'élève-observateur, traduire ces critères en images).

La complexité est redoublée par le fait que les prestations d'élèves sont en EPS souvent dépendantes les unes des autres : concrètement, quel est l'élève responsable d'une passe manquée en sports collectifs ? Le passeur ou le réceptionneur ? Enfin, troisième spécificité de l'EPS qui augmente encore la complexité de l'évaluation « pour les apprentissages » : ceux-ci portent souvent sur des activités de performance ou compétitives. Dès lors, l'enseignant peut toujours minimiser les comparaisons entre élèves ou par rapport à des barèmes, « accorder un statut positif à l'erreur » pour accompagner les apprentissages, il n'empêchera pas les élèves de chercher à courir plus vite que leurs camarades ou de donner la priorité au résultat du match. Paradoxalement, plus l'évaluation sera « authentique » dans ces activités, plus ces effets pervers risqueront de se multiplier, réintroduisant la comparaison à une norme interne et externe.

Le relatif immobilisme des pratiques évaluatives en EPS trouve aussi sans doute son origine chez l'enseignant dans les dilemmes liés à deux impératifs simultanés et difficilement compatibles : en interne, l'obligation d'accompagner l'élève, de susciter l'engagement de la classe ; en externe, l'obligation de rendre des comptes. Et il n'est finalement pas étonnant que les enseignants soient focalisés sur les performances et la certification, dans une école ancrée dans un *accountability system* (Dinan Thompson & Penney, 2015) et dans un contexte social qui prône la compétitivité. Obligés de « rendre des comptes », ils sont traversés néanmoins par des tensions et des incertitudes face aux possibles injustices liées à ce

type d'évaluation – contrôle et à ses conséquences négatives sur l'engagement et l'apprentissage des élèves.

À ce sujet, la question de la subjectivité devient le point d'achoppement de l'activité évaluative. Elle ne peut être évitée à propos d'activités motrices complexes et permet en outre une compréhension fine des situations. Mais les éventuelles défaillances du jugement des enseignants liées à cette subjectivité peuvent aussi créer un sentiment d'injustice. Un phénomène relevé par la littérature francophone attire l'attention sur ce point car il tend à étayer l'hypothèse d'un dilemme fondamental lorsqu'un enseignant évalue en EPS : les « arrangements évaluatifs » (Cogérino, 2009). Il s'agit de pratiques évaluatives consistant à réévaluer *a posteriori* et de façon clandestine une note donnée à des élèves à des épreuves d'évaluation. Concrètement, les enseignants jugeant que tel élève peu engagé ou perturbateur obtient de trop bons résultats baissent sa note à l'insu de tous. À l'inverse, ils augmentent de la même manière la note d'élèves peu performants mais engagés. Si ce type de pratiques injustes et même illégales peut être jugé sévèrement, notamment pour leur subjectivité qui confine à l'arbitraire, on peut noter qu'elles émergent des tiraillements intenses ressentis face à des évaluations certificatives ponctuelles, rendant seulement compte de différences de niveaux, indépendamment des efforts consentis par l'élève, des progrès réalisés. Le conflit de valeurs sous-jacent aux évaluations formatives et certificatives en EPS apparaît ici à son point culminant.

CONCLUSION

Comme l'indiquent à juste titre les recherches portant sur toutes les disciplines scolaires, l'évaluation endosse une multitude d'exigences dont certaines sont inconciliables (Allal, 2012) : rendre compte des acquis d'un élève à un moment donné, comparer ces acquis à une norme externe, rendre justice à tel élève en fonction de ses performances ponctuelles, mais aussi selon ses progrès, ses efforts, favoriser ses apprentissages, susciter son engagement, etc. L'ampleur des enjeux à relever au quotidien est très importante, d'autant plus que les options dans ce domaine ne sont pas seulement indexées à des problèmes techniques, à l'opérationnalité supposée de dispositifs ou de gestes professionnels. L'action évaluative relève aussi d'orientations de valeurs, de choix philosophiques. Parce qu'elle est intégrée à l'enseignement-apprentissage, parce qu'elle se déploie à l'interface du processus individuel d'apprentissage et du versant social, normé, de l'enseignement, elle constitue un carrefour spécifiquement encombré, congestionné, par toutes les problématiques éducatives.

Et l'EPS n'échappe pas à ce nœud problématique qui rend si difficile l'établissement d'un paradigme consensuel (De Ketele, 2012). Et sans doute même, l'évaluation dans cette discipline constitue-t-elle un « analyseur » particulièrement éclairant : l'importance dans les prestations évaluées d'élèves des facteurs internes et stables de la performance motrice, liés au gabarit, à la physiologie d'une part, liés aux apprentissages extrascolaires d'autre part. De plus, comme nous l'avons montré, les enseignants sont confrontés aux difficultés liées à l'absence de traces lors des apprentissages d'élèves, à l'interdépendance des élèves et à des supports d'apprentissages compétitifs. Enfin, l'implication récurrente des élèves dans des situations collectives est aussi source de perturbations en EPS (Mougenot, 2013).

Cette synthèse de la littérature sur les pratiques évaluatives en EPS nous incite à suggérer des pistes de recherches qui permettraient de gagner en compréhension à propos des problématiques d'évaluation. Une de ces pistes consisterait à

sonder les résistances récurrentes des enseignants, à analyser de façon clinique leur activité au moment où ils évaluent, leurs motifs simultanés, leur activité empêchée, les opérations qu'ils maîtrisent et celles qu'ils ne maîtrisent pas.

Annabelle GRANDCHAMP
Jacques MEARD
Grégory QUIN
Haute Ecole Pédagogique
du Canton de Vaud
Université de Lausanne
Institut des Sciences du Sport

Abstract : Evaluation practices in physical education (PE) have been widely analyzed and discussed in the educational issues of the industrialized countries. Far from reaching a conceptual consensus, contradictions and tensions persist, especially between theoretical concepts and teacher practices. This literature review first tries to clarify the ambiguities intrinsically related to the evaluation in PE by proposing a reminder of the definitions and different objects of the evaluation. In a second step, this paper presents alternative assessments appeared in some countries in response to a dissatisfaction in terms of learning traditional assessments. However, we will see that these more formative orientations are struggling to generalize. Finally, we will try to find some hypotheses to explain the relative continuity of evaluation practices in EPS.

Keywords : physical education, summative assessment, formative assessment, evaluation, reform.

Bibliographie

- Allain M., Deriaz D., Voisard N. & Lentillon-Kaestner V. (2016) « Perceptions des élèves au niveau du soutien de l'enseignant et de leur compétence dans trois contextes au mode de communication de l'évaluation sommative différent en éducation physique et sportive » – *Staps* 113 (73-87).
- Allal L. (2012) « Les exigences inconciliables des activités évaluatives et de leurs paradigmes de référence » – in : L. Mottier Lopez et G. Figari (éds.) *Modélisations de l'évaluation en éducation* (181-194). Bruxelles : De Boeck.
- Annerstedt C. & Larsson H. (2010) « "I have my own picture of what the demands are..." : Grading in Swedish PEH – problems of validity, comparability and fairness » – *European Physical Education Review* 16, 2 (97-115).
- Araújo R., Hastie P., Lohse K. R., Bessa C. & Mesquita I. (2017) « The long-term development of volleyball game play performance using Sport Education and the Step-Game-Approach model » – *European Physical Education Review* 21 (1-16).
- Arslan Y., Erturan Ilker G. & Demirhan G. (2013) « The Impact of a Developed Measurement and Evaluation Development Program on Pre-service Physical Education Teachers' Perceptions Related to Measurement and Evaluation » – *Educational Sciences: Theory & Practice* 13, 2 (1119-1124).
- Aubin S. (2010) « De la terminologie musicale pour la didactique du FLE : ouvertures et repérages » – *Anales de Filologia Francesa* 18 (17-29).

- Avery M. (2012) « Web-Based Assessment of Physical Education Standards » – *Journal of Physical Education, Recreation & Dance* 83, 5 (27-34).
- Bourghouts L. B., Slingerland M. & Haerens L. (2017) « Assessment quality and practices in secondary PE in the Netherlands » – *Physical Education and Sport Pedagogy* 22, 5 (473-489).
- Brau-Antony S. & Hubert B. (2014) « Curriculum en Education Physique et Sportive et évaluation certificative au baccalauréat » – *Questions Vives* 22 (67-82).
- Brown T. D. & Penney D. (2017) « Interpretation and enactment of Senior Secondary Physical Education : pedagogic realities and the expression of Arnoldian dimensions of movement » – *Physical Education and Sport Pedagogy* 22 (121-136).
- Cain D. I. (2010) « Beliefs and practices about implementing technology in physical education » – *Dissertations Abstracts International, AI-A* 71, 06, December.
- Carnus M.-F. & Terrisse A. (2006) « La compétence à évaluer : une compétence révélatrice de la professionnalité des enseignants d'EPS. Étude de cas et perspectives pour la formation » – *Savoirs* 3, 12 (55-74).
- Casey A. & Jones B. (2011) « Using digital technology to enhance student engagement in physical education » – *Asia-Pacific Journal of Health, Sport and Physical Education* 2, 2 (51-66).
- Cogérino G. (2009) « Enseignement d'EPS et évaluation certificative : enjeux implicites relatifs à la justice et l'équité » – *Spirale* 43 (9-19).
<http://spirale-edu-revue.fr/spip.php?article862>
- Cogérino G. & Mnaffakh H. (2008) « Évaluation, équité de la note en éducation physique et norme d'effort » – *RFP* 164 (111-122).
- Collier D. (2011) « Increasing the Value of Physical Education » – *Journal of Physical Education, Recreation & Dance* 82, 7 (38-41).
- Constantinou P. (2017) « Instructional Assessment Strategies for Health and Physical Education » – *Strategies* 30, 3 (3-9).
- De Ketele J.-M. (2010) « Ne pas se tromper d'évaluation » – *Revue Française de Linguistique Appliquée* 15, 1 (25-37).
- De Ketele J.-M. (2012) « À la recherche de paradigmes unificateurs ou de paradigmes intégrateurs ? » – in : L. Mottier Lopez et G. Figari (éds.) *Modélisations de l'évaluation en éducation*. Bruxelles : De Boeck.
- Dinan Thompson M. & Penney D. (2015) « Assessment literacy in primary physical education » – *European Physical Education Review* 21, 4 (485-503).
- Dyson B. (2014) « Quality Physical Education : A Commentary on Effective Physical Education Teaching » – *Research Quarterly for Exercise and Sport* 85, 2 (144-152).
- Genelot S., Gardes D., Mansanti J. & Pinsard N. (2016) « Évaluer par compétences ou évaluer sans noter ? » – *Éducation permanente* 208, 3 (89-100).
- Georgakis S. & Wilson R. (2012) « Australian Physical Education and School Sport: An Exploration into Contemporary Assessment » – *Asian Journal of Exercise & Sports Science* 9, 1.
- Gerard J.-M. (2013) « L'évaluation au service de la régulation des apprentissages : enjeux, nécessités et difficultés » – *Revue Française de Linguistique Appliquée* 153, 1 (75-92).

- Glennon W., Hart A. & Foley J. T. (2015) « Developing Effective Affective Assessment Practices » – *Journal of Physical Education, Recreation & Dance* 86, 6 (40-44).
- Harvey S. & Gittins C. (2014) « Effects of integrating video-based feedback into a teaching games for understanding soccer unit » – *Agora for physical and sport* 16, 3 (271-290).
- Hay P. & Penney D. (2009) « Proposing conditions for assessment efficacy in physical education » – *European Physical Education Review* 15, 3 (389-405).
- Holt/Hale S. A. & Persse D. (2015) « The National Physical Education Standards and Grade-level Outcomes : The Future of Elementary Physical Education » – *Journal of Physical Education, Recreation & Dance* 86, 7 (14-16).
- Kretschmann R. (2015) « Effect of Physical Education Teachers' Computer Literacy on Technology Use in Physical Education » – *Physical Educator* 72, 5 (261-277).
- Larsson H. & Nyberg G. (2017) « “It doesn't matter how they move really, as long as they move.” Physical education teachers on developing their students' movement capabilities » – *Physical Education and Sport Pedagogy* 22, 2 (137-149).
- Lau A. M. S. (2016) « Formative good, summative bad ? » – *Journal of Further and Higher Education* 40, 4 (509-525).
- Leirhaug P. E. (2016) « Exploring the relationship between student grades and assessment for learning in Norwegian physical education » – *European Physical Education Review* 22, 3 (298-314).
- Leirhaug P. E. & Annerstedt C. (2015) « Assessing with new eyes? Assessment for learning in Norwegian physical education » – *Physical Education and Sport Pedagogy* 21, 6 (616-631).
- Leirhaug P. E. & MacPhail A. (2015) « “It's the other assessment that is the key” : three Norwegian physical education teachers' engagement (or not) with assessment for learning » – *Sport, Education and Society* 20, 5 (624-640).
- Leirhaug P. E., MacPhail A. & Annerstedt C. (2016) « “The grade alone provides no learning” : investigating assessment literacy among Norwegian physical education teachers » – *Asia-Pacific Journal of Health, Sport and Physical Education* 7, 1 (21-36).
- Lentillon-Kaestner V. (2008) « Les élèves du second degré face à l'évaluation en éducation physique et sportive » – *Staps* 79 (49-66).
- Lentillon-Kaestner, V. (2014) « Notation en éducation physique et perceptions d'injustice chez les élèves : les apports d'une approche mixte » – *Formation et Pratiques d'Enseignement en Questions* 17 (89-108).
- Lentillon-Kaestner V., Allain M., Deriaz D. & Voisard N. (2018) *Noter en éducation physique ? Incidences sur l'enseignement et les élèves*. Louvain-la-Neuve : Éditions EME.
- López-Pastor V. M., Kirk D., Lorente-Catalan E., MacPhail A. & Macdonald D. (2013) « Alternative assessment in physical education: a review of international literature » – *Sport, Education and Society* 18, 1 (57-76).
- Lorente-Catalan E. & Kirk D. (2014) « Making the case for democratic assessment practices within a critical pedagogy of physical education tea-

- cher education » – *European Physical Education Review* 20, 1 (104-119).
- Lorson K. & Mitchell S. (2016) « Advocacy, Assessment and Accountability: Using Policy to Impact Practice in Ohio » – *Journal of Physical Education, Recreation & Dance* 87, 3 (44-49).
- Lu-Ho H., Iwen H. & Gwo-Jen H. (2016) « Effects of different online peer-feedback approaches on students' performance skills, motivation and self-efficacy in a dance course » – *Computers and Education* 96 (55-71).
- MacPhail A. & Halbert J. (2010) « “We had to do intelligent thinking during recent PE” : students' and teachers' experiences of assessment for learning in post-primary physical education » – *Assessment in Education : Principles, Policy & Practice* 17, 1 (23-39).
- MacPhail A. & Murphy F. (2017) « Too much freedom and autonomy in the enactment of assessment? Assessment in physical education in Ireland » – *Irish Educational Studies* 36, 2 (237-252).
- Martin C., Horton M. L. & Tarr S. J. (2015) « Building Assessment Tools Aligned with Grade-level Outcomes » – *Journal of Physical Education, Recreation & Dance* 86, 7 (28-34).
- Méard J. & Grandchamp A. (2018) « Les différentes formes d'évaluation » – in : M. Travert et O. Rey (éds.) *L'engagement de l'élève en EPS*. Paris : Revue EPS (171-177).
- Melograno V. J. (2007) « Grading and Report Cards for Standards-Based Physical Education » – *Journal of Physical Education, Recreation & Dance* 78, 6 (45-53)
- Mottier Lopez L. (2009) « Introduction. L'évaluation en éducation : des tensions aux controverses » – in : L. Mottier Lopez et M. Crahay (éds.), *Évaluations en tension* (7-25). Bruxelles : De Boeck.
- Mougenot L. (2013) *L'évaluation scolaire en question : de l'activité des enseignants aux conduites des élèves. Impact des évaluations proposées et des modalités de groupement sur les conduites des élèves du secondaire en éducation physique et sportive*. Université René Descartes, Paris V, Thèse de Doctorat.
- Mougenot L. (2016) « Régulation différée et apprentissage en sport collectif » – *Staps* 111 (29-41).
- Mougenot L. & Dugas E. (2014) « Formation et pratique des enseignants : Regard porté sur l'évaluation en EPS » – *Spiral-E – Revue de Recherches en Éducation* 53 (67-83).
<http://spirale-edu-revue.fr/spip.php?article1198>
- Nazari H., Jafari E. M., Nasr A. R. & Marandi S. M. (2017) « School Physical Education Curriculum of Iran from Experts' Perspective: “What It Is and Should Be” » – *International Journal of Environmental & Science Education* 12, 5 (971-984).
- Ni Chròinìn D. & Cosgrave C. (2013) « Implementing formative assessment in primary physical education: teacher perspectives and experiences » – *Physical Education and Sport Pedagogy* 18, 2 (219-233).
- O'Connor J., Jeanes R. & Alfrey L. (2016) « Authentic inquiry-based learning in health and physical education: a case study of 'r/evolutionary' practice » – *Physical Education and Sport Pedagogy* 21, 2 (201-216).
- Oh J. & Graber K. C. (2017) « National Curriculum for Physical Education in the United States » – *Quest* 69, 2 (220-235).

- Palao J.-M., Hastie P.-A., Cruz P.-G. & Ortega E. (2015) « The impact of video technology on student performance in physical education » – *Technology, Pedagogy and Education* 24, 1 (51-63).
- Penney D., Brooker R., Hay P. & 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).
- Pulfrey C., Buchs C. & Butera F. (2011) « Why Grades Engender Performance-Avoidance Goals : The Mediating Role of Autonomous Motivation » – *Journal of Educational Psychology* 103, 3 (683-700).
- Redelius K. & Hay P. J. (2012) « Student views on criterion-referenced assessment and grading in Swedish physical education » – *Physical Education and Sport Pedagogy* 17, 2 (211-225).
- Richard E. (2016) « Former à évaluer : enjeux, tensions, solutions » – *Education permanente* 208, 3 (101-114).
- Rink J., Jones L., Kirby K., Mitchell M. & Doutis P. (2007) « Teacher perceptions of Physical Education Statewide Assessment Program » – *Quarterly for Exercise and Sport* 78, 3 (204-215).
- Simard C., Dufays J.-L., Dolz-Mestre J. & Garcia-Debanc C. (2010) *Didactique du français langue première*. Bruxelles : De Boeck.
- Society of Health and Physical Educators (2014) « National standards and grade-level outcomes for K-12 physical education » – Champaign, IL: Human Kinetics.
- Spittle M. & Byrne K. (2009) « The influence of Sport Education on student motivation in physical education » – *Physical Education and Sport Pedagogy* 14, 3 (253-266).
- Stobart, G. (2011) « L'évaluation pour les apprentissages : d'une expérimentation locale à une politique nationale » – *RFP* 174 (41-48).
- Wallhead T. & O'sullivan M. (2005) « Sport Education: physical education for the new millennium? » – *Physical Education and Sport Pedagogy* 10, 2 (181-210).
- Zhu X. (2015) « Student perspectives of grading in physical education » – *European Physical Education Review* 21, 4 (409-420).