

As provas de aferição e a epistemologia da Educação Física

RESUMO

Em Portugal, as provas de aferição foram introduzidas na disciplina de Educação Física com o objetivo de fornecer informações que levem a melhores aprendizagens, o que se traduz em conhecimento novo posto à disposição dos agentes e, por isso, tem implicações na epistemologia da disciplina. Contudo, esse processo está a decorrer sem escrutínio científico, o que é relevante, porque sabemos que as avaliações externas das aprendizagens podem ter efeitos imprevistos e levar a discursos de autoridade ideológica sobre a escola. Assim sendo, importa questionar sobre quais são os discursos que justificam as provas de aferição, qual é a validade e confiabilidade das provas; e quais são as interações entre o conhecimento oferecido pelas provas e as práticas letivas. Para responder, foi implementada uma investigação, inscrita no paradigma interpretativo, na modalidade de estudo de caso instrumental e com abordagem qualitativa, com recurso a análise documental e a análise de conteúdo. Os resultados sugerem que as provas de aferição são implementadas, também, com objetivos de controlo das práticas de Educação Física no 1.º ciclo, de valorização social da disciplina e medem uma parte diminuída das aprendizagens dos alunos, através de processos muito pouco confiáveis. Ainda assim, os resultados são usados para influenciar a generalidade do processo de ensino, o que traduz encobrimento curricular. Conclui-se que as provas suportam a metamorfose epistemológica da Educação Física, do cuidado para o clínico, do sentido para o dever, do jogo para a repetição e da coevolução para a linearidade.

Palavras-chave: Educação Física; Avaliação externa das aprendizagens; Provas de aferição; Epistemologia.

1. INTRODUÇÃO

Em Portugal, as Provas de Aferição (PA) foram implementadas na disciplina de Educação Física (EF) em 2017, na expectativa de oferecer informações válidas a serem usadas por alunos, professores, escolas e famílias para melhorar as aprendizagens. Nessa medida, traduzem expectativas acerca do que é o conhecimento dos professores e têm potencial para influenciar a ação docente (Gess-Newsome, 2015). Logo, importa perguntar se elas são coerentes com o perfil epistémico da disciplina, tanto mais que o escasso escrutínio científico sobre as PA não tem ponderado sobre as qualidades

Nuno Miranda e Silva ⁱ
Universidade de Évora,
Portugal.

Sónia Pereira Dinis ⁱⁱ
Universidade de Évora,
Portugal.

avaliativas nem sobre a relação entre os propósitos e a realidade – os estudos que conhecemos dedicam-se à percepção de professores e encarregados de educação do 1º ciclo sobre a pertinência das PA (e.g., Neves et al., 2023, 2024).

Esse escrutínio é relevante porque sabemos que a testagem massiva e uniformizada tem efeitos que nem sempre são os esperados (Madaus et al., 2009) e que o conhecimento que produzem pode suportar discursos autoritários e de imposição política sobre a escola (Berkovich & Benoliel, 2020; Bart, 2015; Lopo, 2021). A isto, juntam-se três situações específicas à EF, que se seguem.

Primeiro, os resultados das provas de 2022 apontam a melhoria de algumas aprendizagens relativamente a 2019 (Instituto de Avaliação Educativa [IAVE], 2022b), devendo-se lembrar que esse intervalo assistiu a uma pandemia e a literatura revela que, nesse período, as práticas letivas foram radicalmente afetadas e as aprendizagens dos alunos limitadas (González-Calvo et al., 2022; Silva & Dinis, 2023; Varea & González-Calvo, 2021). Assim, este contraste leva à necessidade de debater a validade e confiabilidade do que é medido nas PA.

A segunda situação a mencionar relaciona-se com dinâmicas emergentes na comunidade docente de EF, que tendem a focar-se na proteção da desvalorização social da disciplina, podendo estar na origem da mimetização de estratégias que, sendo socialmente valorizadas, asseguram a paridade da disciplina perante outras (Silva & Dinis, 2022), mas podem não se adequar à estrutura específica da EF. Isso justifica que se analisem os discursos que fundamentam as PA.

Por fim, a terceira assenta no facto de a disciplina estar estruturada em ciclos sistemáticos de diagnóstico, avaliação e revisão do planeamento (Jacinto et al., 2001), a partir de referenciais de avaliação nacionalmente uniformizados. Portanto, aqui importa conhecer as interações entre as PA, que uniformizam as tarefas de avaliação num momento do ano, e as práticas letivas, permanentes e assentes na diferenciação.

Neste contexto, o presente artigo apresenta uma investigação destinada a produzir conhecimento sobre estas três áreas de necessidade e a debater as influências que têm no perfil epistémico da EF.

2. PROBLEMÁTICA

O debate central deste artigo é o conhecimento na disciplina de EF, que está subjacente aos documentos orientadores e é partilhado pelos alunos e professores portugueses, bem como o conhecimento que as PA acrescentam. Por isso, começa-se por fazer o enquadramento bibliográfico da epistemologia da disciplina e das PA.

2.1 A EPISTEMOLOGIA DA EDUCAÇÃO FÍSICA

Na EF, a educação assenta na proximidade, no movimento e no contacto físico (Öhman & Quennerstedt, 2017) e a disciplina pode ser entendida como

a prática pedagógica que influencia as condutas motoras. Estas expressam a organização significativa do comportamento motor e compreendem os elementos físico-motor, afetivo-emocional, cognitivo, social-relacional e expressivo, que traduzem a simultânea globalidade e particularidade do ser humano (Parlebas, 2001).

São estes meios que constroem um conhecimento específico (Gess-Newsome, 2015) que alimenta as práticas letivas. Uma vez que o conjunto de orientações curriculares dificilmente cabe no espaço de um artigo, optamos por traduzi-lo em metáforas (elas têm mostrado capacidade para representar e dar sentido à realidade e guiar a reflexão) (Jensen, 2016), que caracterizam dimensões da epistemologia da EF: o cuidado, o sentido, o jogo e o movimento e a coevolução.

O conhecimento na EF traduz as necessidades de cuidado de cada um dos alunos e de todos eles (Gilligan, 2016), em função das características únicas do contexto (e.g., condições materiais) e das crianças e jovens (e.g., o seu historial desportivo e as motivações).

Para isso, o ensino começa pela identificação das potencialidades e necessidades de cada aluno nas primeiras semanas do ano letivo (a avaliação inicial), o que está na origem do planeamento do ano letivo (Fialho & Fernandes, 2011; Jacinto et al., 2001) e leva a que ele seja diferente em cada turma. Depois, esse planeamento é revisto em três momentos, sempre no sentido de acolher as aprendizagens desenvolvidas. É a partir desse conhecimento gerúndio que o professor organiza os grupos, prevê as tarefas adequadas a cada aluno, pondera a sua própria intervenção (no sentido de estar mais próximo dos alunos com mais necessidades de apoio e de autonomizar aqueles que não o necessitam) e negocia objetivos de desempenho de modo que os alunos reorganizem as suas expectativas e participação.

O conhecimento estrutura-se, também, no questionamento sobre o sentido (Berger, 2009) que cada atividade tem para os alunos e para a turma e, por isso, a disciplina funciona para que a diversidade seja acolhida.

Primeiro, o sucesso dos alunos pode ser conseguido através de percursos muito diferenciados e um determinado aluno pode ter sucesso educativo nos jogos desportivos coletivos por desenvolver as aprendizagens essenciais em basquetebol e voleibol e outro por fazê-lo no futebol e andebol (Direção-Geral de Educação [DGE], 2023a). Mesmo na aptidão física, sujeita a testes padronizados, a avaliação está orientada para a saúde do aluno, na perspetiva das suas condições específicas e dos valores culturais e sociais que o podem ajudar a tornar-se um cidadão feliz e saudável.

Nessa medida, a EF afasta-se de médias aritméticas e da classificação e recorre a níveis de desempenho que integram (sem dividir) os vários domínios do desenvolvimento humano (e.g., cognitivo, emocional, motor) (Jacinto et al., 2001; Parlebas, 2001) – assim, o aluno é interpretado na sua integralidade. Essa dimensão do sentido é revelada, por exemplo, quando os professores podem estabelecer objetivos mais exigentes do que os previstos, tendo em vista alunos extraordinários; ou a inclusão de matérias que traduzam potencialidades locais (e.g., a natação, caso haja acesso a piscinas) (Jacinto et al., 2001).

O conhecimento também está enraizado no jogo e no movimento. As competências e aprendizagens são imediatamente desafiadas nas interações entre os alunos (e.g., no jogo de badmínton ou na coreografia de dança) (Méndez et al., 2010), e a avaliação prossegue por instantes irrepetíveis, o que se traduz em informação adaptativa. Nesse sentido, assenta em descritores de ação que estão aferidos para todo o país (organizados em rubricas de avaliação de grau de complexidade crescente, denominadas “introdução”, “elementar” e “avançado”). O que se torna variável são as circunstâncias do aluno/turma, e a realidade e a aprendizagem são protagonizadas pelo jogo (não pela competição) de ir sempre desafiando limites. Nessa medida, a resposta dos alunos proporciona novos níveis de exigência e o que a aprendizagem verdadeiramente permite é um estado de prontidão para aprender mais, em vez de objetivos definitivos e únicos.

O conhecimento que emerge na EF também assume a coevolução, devido a relações de interdependência, em que a alteração das condições dos agentes influencia as circunstâncias dos outros (Silva, 2019). Os alunos influenciam ativamente a aprendizagem e a avaliação dos colegas (e.g., ajudam na ginástica, ou são colegas de equipa no basquetebol) e uma das aprendizagens da EF é a da colaboração independentemente do nível de desempenho (DGE, 2023a, 2023b; Jacinto et al., 2001), o que leva a que se evitem situações em que os melhores alunos estejam sistematicamente juntos – por vezes, eles são convocados para auxiliar o professor (Jacinto et al., 2001).

Além disso, a avaliação sumativa é realizada a partir das atividades em que os alunos têm melhor desempenho (e.g., a avaliação de um aluno pode integrar a ginástica de solo e de outro a ginástica de aparelhos), o que significa que os alunos têm opções e escolhas respeitadoras das suas vivências e motivações – talvez por isso, a EF apresente níveis de sucesso consistentemente elevados, entre os 96% e os 99% (Direção-Geral de Estatísticas da Educação, 2022).

Tudo isto traduz que a avaliação, a aprendizagem e o ensino são heterárquicos (Silva, 2019), porque os alunos tomam decisões e têm poderes influenciadores da ação dos professores – aliás, nos 11^o e 12^o anos, escolhem as matérias que serão lecionadas (DGE, 2023a; Jacinto et al., 2001).

2.2 AVALIAÇÃO, AVALIAÇÃO EXTERNA E PROVAS DE AFERIÇÃO

A literatura tem identificado dimensões que permitem interpretar a avaliação, em função dos seus objetivos (avaliação da aprendizagem, para a aprendizagem e como aprendizagem), da regulação que propicia (expressar, comparar ou interpretar as aprendizagens), da ética social que persegue (standardização, equidade ou diferenciação) e da teorização sobre a própria avaliação (confiabilidade de consistência entre avaliações e validade da prova medir o que se propõe) (DeLuca et al., 2016).

A expressão destas dimensões decorre de indicadores como as funções, natureza e utilidade da avaliação; a articulação com os processos de ensino e de aprendizagem; e o papel que os alunos e os professores têm no processo e o recurso a estratégias específicas (e.g., a auto e heteroavaliação e a distribuição e tipologia do feedback) (Borrvalho et al., 2015) (Figura 1).

Figura 1
Caracterização teórica da prática de avaliação

Dimensões	Elementos	Expressão
Objetivos	Avaliar as aprendizagens Avaliar para as aprendizagens Avaliar como aprendizagem	Tarefas de avaliação mais utilizadas Funções e natureza da avaliação Utilidade da avaliação Integração/articulação entre os processos de ensino, avaliação e aprendizagem Natureza, frequência e distribuição de feedback Papel dos professores e dos alunos Recurso a estratégias de auto e heteroavaliação
Potencial de regulação	Expressão das aprendizagens face ao previsto Comparação das aprendizagens Interpretação das aprendizagens	
Componente ética	Standartização Equidade (necessidades de grupos) Diferenciação (caraterísticas de cada aluno)	
Teoria	Confiabilidade (consistência entre avaliações) Viabilidade (medição do objeto) Mista	

Fonte: Elaborado a partir de DeLuca et al. (2016) e de Borralho et al. (2015).

As PA constituem-se como avaliação externa, pois são elaboradas, administradas e corrigidas por uma instituição exterior à escola (Fernandes, 2019), o Instituto de Avaliação Educativa (IAVE). Este tipo de avaliação, de carácter externo, apresenta potencialidades, mas também acrescenta desequilíbrios ao sistema.

Quando as provas têm impactos na classificação dos alunos, por um lado, permitem diluir o efeito das escolas na classificação final e ordenar os alunos nos concursos de acesso ao ensino superior, mas, por outro, estão na origem do afunilamento do currículo, em que os professores ensinam o que esperam que seja avaliado (Fernandes, 2019).

Quando a avaliação externa não tem impactos na classificação dos alunos, como é o caso das PA, o problema emergente é da validade dos resultados, porque os alunos podem entender que as provas não são importantes e limitar o esforço e o empenho que colocam nas tarefas (Debeer et al., 2014; Pools & Monseur, 2021; Schlosser et al., 2019). De qualquer forma, as avaliações externas, de carácter massivo, têm estado na origem de discursos de autoridade sobre a escola, a partir de ideologias políticas e de senso comum (Berkovich & Benoliel, 2020; Bart, 2015; Lopo, 2021).

Em Portugal, as PA têm os objetivos de:

Acompanhar o desenvolvimento do currículo nas diferentes áreas, fornecer informações detalhadas às escolas, aos professores, aos encarregados de educação e aos alunos sobre o desempenho destes últimos e potenciar uma intervenção pedagógica atempada, dirigida às dificuldades específicas de cada aluno. (IAVE, 2022a, p. 1)

Para isso, asseguram a cobertura integral do currículo (n.º 2 do artigo 26.º da Portaria n.º 223-A/2018), dirigem-se a toda a população escolar portuguesa dos 2.º, 5.º e 8.º anos e estão coordenadas com os ciclos de ensino, de tal forma que foram planeadas para oferecer informações a meio de cada ciclo, com vista a eventuais correções (Tabela 1).

Tabela 1

Visão geral dos ciclos de ensino, do currículo de Educação Física e do enquadramento das provas de aferição

Ciclo	Ano	Docente	Currículo
1.º ciclo	1.º	Professor único que leciona todas as áreas do currículo, podendo ser coadjuvado se as escolas tiverem recursos humanos disponíveis.	Perícias e manipulações; Deslocamentos e equilíbrios; Jogos.
	2.º		
	3.º		Ginástica; Jogos;
	4.º		Exploração da natureza ou patinagem ou natação.
2.º ciclo	5.º	Professor especialista, com formação para lecionar o 5.º e 6.º anos, podendo coadjuvar o 1.º ciclo.	Conhecimentos; Aptidão física; Jogos (apenas no 5.º ano); Jogos desportivos coletivos;
	6.º		
3.º ciclo	7.º	Professor especialista, com formação para lecionar do 7.º ano ao 12.º ano, podendo lecionar ao 5.º e 6.º ano e coadjuvar o 1.º ciclo.	Ginástica; Atletismo;
	8.º		
Ensino Secundário	9.º		Danças; Outras.
	10.º		
	11.º		
	12.º		

Nota: Assinalados a cinzento os anos letivos em que se realizam as provas de aferição.

Os resultados das PA de EF revelam que, entre 2019 e 2022, o número de alunos do 2.º ano a conseguir realizar as tarefas de deslocamentos e equilíbrios e de perícias e manipulações duplicou. No entanto, constata-se que houve menos alunos a conseguir realizar as tarefas de jogos infantis. Quanto ao 3.º ciclo, os resultados também apontam melhorias: em 2022, a percentagem de alunos que consegue realizar as tarefas de ginástica é de 18,7% (10,1% em 2019) e nos jogos desportivos coletivos é de 29% (19,1% em 2019). Os níveis de realização nas tarefas das PA de 2022 que não têm comparação com as PA de 2019 é de 32% nas atividades rítmicas expressivas e 71,2% no atletismo. Os restantes alunos são avaliados com “consegue, mas”, “revelaram dificuldades” e “não conseguiram/não responderam” (IAVE, 2022b).

3. METODOLOGIA

A investigação assume que a EF tem um perfil epistémico assente no cuidado, no sentido, na coevolução e no jogo e no movimento, propondo-se a debater a influência das PA nesse perfil. Para isso, analisa documentos e interações entre os agentes educativos da EF mobilizados por entidades representativas da EF para responder às perguntas: (a) Que discursos sustentam as PA?; (b) Como se caracteriza a validade e confiabilidade do conhecimento inerente às PA?; e (c) Qual é a interação entre o conhecimento das PA e as práticas letivas?

Face às questões estabelecidas, a investigação inscreve-se no paradigma interpretativo, uma vez que os resultados assentam na interpretação de documentos produzidos por agentes educativos (Guba & Lincoln, 1998). Não obstante, assume uma abordagem qualitativa (Merriam & Tisdell, 2016), pois mobiliza documentos e percepções de agentes para compreender o fenómeno e enquadra-se na modalidade de estudo de caso, de carácter instrumental (Stake, 2009), já que pretende compreender as PA para debater a epistemologia da EF (Figura 2).

Figura 2

Desenho da investigação

Perguntas	Objetos	Recolha de dados	Tratamento de dados
Que discursos sustentam as provas de aferição e que expectativas de conhecimento emergem?	Objetivos das provas Potencial de regulação Perfil ético	Documentos de entidades da EF	
Como se caracteriza a validade e confiabilidade do conhecimento inerente às provas de aferição?	Tarefas Materiais Critérios de correção Áreas avaliadas	Provas de aferição Aprendizagens essenciais Orientações programáticas	Análise de conteúdo por categorias apriorísticas e por categorias induzidas
Qual é a interação entre o conhecimento das provas de aferição e as práticas letivas?	Integração/articulação dos resultados com os processos de ensino, avaliação e aprendizagem	Documentos de entidades da EF Programas de EF	

↓

Debate sobre a influência das provas de aferição na epistemologia da EF

A recolha de dados ocorreu por análise documental. Os documentos são registos permanentes e de carácter não intrusivo, o que os liberta, numa certa medida, da interpretação dos investigadores (Hatch, 2002; Merriam & Tisdell, 2016); e o facto de persistirem no tempo e de serem de acesso público favorece a replicação da investigação e permite aceder às dinâmicas sociais que os agentes assumem.

Os dados recolhidos tiveram em conta as questões de investigação estabelecidas (Tabela 1), o que esteve na origem da análise: dos documentos públicos emanados pelo Conselho Nacional de Associações de Profissionais de Educação Física e Desporto (CNAPEF) sobre as PA, porque se entendeu que esta organização reúne e sintetiza o pensamento dos agentes da área,

com a vantagem de se mostrar capaz de mobilizar atores estatais (no caso, representantes do Ministério da Educação e do IAVE que, respetivamente, decidem sobre as PA e as produzem, administram e corrigem) – num total de 28 documentos relevantes, entre os quais, há a salientar o registo vídeo de um seminário dedicado às boas práticas de três escolas, localizadas na região de Lisboa (traduzidas em apresentações por parte de um dirigente de uma Associação de Professores de EF e de três professores representantes de grupos de EF de agrupamentos de escolas), quanto à organização das PA e à utilização dos resultados, tendo, também, contado com a participação de dirigentes das entidades representativas da EF (para além do CNAPEF, da Sociedade Portuguesa de Educação Física – SPEF), um dirigente do IAVE e um dirigente do Ministério da Educação diretamente ligados à implementação das PA; dos guiões e critérios de correção das PA de EF (33 documentos); e das orientações curriculares, aprendizagens essenciais e respetivos anexos, que esclarecem os objetivos de aprendizagem de cada ano de escolaridade (7 documentos).

O tratamento dos dados decorreu por análise de conteúdo, técnica que permite extrair-lhes o sentido e o significado (Bardin, 2011; Peräkylä, 2005). Parte do tratamento recorreu à análise de conteúdo de verificação (Grawitz, 1993), com recurso a uma matriz de categorização (disponível como material complementar, Silva, 2024), uma vez que se mobilizaram as estruturas teóricas que permitiram estabelecer, aprioristicamente, categorias (Tabela 2). Outra parte ocorreu por indução, de forma a constituir categorias emergentes e não constantes do enquadramento teórico apriorístico, tendo-se usado a técnica da comparação constante, o que favorece a organização de categorias em função do que os dados sugerem e assegura a relação entre as categorias induzidas e os dados (Silva, 2022).

O tratamento foi auxiliado pelo programa QDA Miner, em que os investigadores assumiram papéis independentes no tratamento dos dados a partir de categorias apriorísticas. Um investigador assegurou o tratamento dos dados e outro fez a análise crítica do tratamento, o que levou a ajustamentos por consenso.

Tabela 2

Relação entre categorias apriorísticas de interpretação dos dados e os objetivos da investigação

Objetivo	Objetos	Dimensões
Sustentação das PA e expectativas de conhecimento.	Objetivos da avaliação	Avaliação da aprendizagem; para a aprendizagem; como aprendizagem.
	Regulação	Expressar; Comparar; Interpretar as aprendizagens.
	Ética social	Standardização; Equidade; Diferenciação.
Validade e confiabilidade do conhecimento das PA. Interação entre as PA e as práticas letivas?	Teoria da avaliação	Validade; Confiabilidade.
	Práticas letivas	Ensino; Avaliação; Aprendizagem.

Os dados que sugeriam categorias emergentes foram analisados indutivamente por um dos investigadores. Primeiro, foram codificados por um processo *in vivo*, linha por linha. Em seguida, os códigos foram agrupados em categorias que agregavam aqueles que eram compatíveis, num processo monitorizado por comparação constante (Silva, 2022) e sujeitos à crítica da segunda investigadora.

Além disso, a credibilidade e a confiabilidade da pesquisa foram asseguradas por procedimentos de análise de casos negativos que envolveram a discussão dos dados que aparentavam não estar suportados nas categorias apriorísticas ou posteriorísticas.

Tratando-se de um estudo interpretativo, em abono da credibilidade e de confiabilidade da pesquisa, esclarecemos que os autores têm experiência em pesquisa de abordagem qualitativa, um tem formação inicial na área da EF e todos os dados e documentos da pesquisa estão disponíveis (Silva, 2024).

4. RESULTADOS

A apresentação dos resultados segue o questionamento da investigação. Começamos pelos resultados que permitem conhecer os argumentos que justificam as PA; depois especificam-se os que caracterizam a validade e confiabilidade das mesmas e, finalmente, os resultados acerca da interação entre as práticas letivas e as PA.

4.1 A SUSTENTAÇÃO E AS EXPECTATIVAS DAS PROVAS DE AFERIÇÃO

Os resultados apontam que o discurso dos agentes acompanha a narrativa oficial acerca dos objetivos e funções das PA. Fica claro que a sua função central é a de avaliar para a aprendizagem, através da oferta de informações específicas às famílias e aos alunos, às escolas e aos professores, a partir de processos que agregam a medição das aprendizagens dos alunos face aos objetivos previstos e a comunicação dos resultados, para que possam ser interpretados pelos vários agentes envolvidos. Isto traduz expectativas de regulação por comparação entre o esperado e o desempenhado, numa lógica ética de standardização.

Mas, além destes processos, os discursos revelam dois objetivos não escritos: a sindicância do 1º ciclo e a valorização social da EF.

Quanto ao 1º ciclo, as PA são justificadas pela expectativa de produzirem conhecimento de influência e controle do cumprimento do currículo. Na base deste objetivo implícito, estão percepções de que o currículo da EF não é desenvolvido no 1º ciclo, que alguns encarregados de educação não valorizam a EF e que as direções dos agrupamentos incumprem, regularmente, no apetrechamento das condições que permitem o ensino de EF no 1º ciclo:

Em termos de resultados, muito interessante saber que no primeiro ciclo, e começamos em 2017, os resultados nacionais das provas de aferição, o documento do IAVE, nos mostra que as provas das expressões são aquelas que têm mais sucesso, o que é para nós muito positivo, mesmo tendo consciência de que em muitos casos a Educação Física não é lecionada, mas isto é um indicador também que vem [...] do segundo ano de escolaridade, portanto, quando a falta da sua leção ainda não [...] se concretizou em lacunas. Por outro lado, quando chegamos à prova de 2018 [8º ano], nós vemos que, se calhar aqui sim, a falta de pré-requisitos, a não leção generalizada da Educação Física começa a evidenciar, ou a notar-se pela falta de capacidade dos alunos para desenvolverem e cumprirem aquilo que são as aprendizagens essenciais. (Dirigente de Associação de Professores de EF, Seminário: Provas de aferição de Educação Física)

Nessa medida, as provas emergem como resolução desses desvios, ora porque estimulam à aquisição de materiais, ora porque fazem a monitorização do trabalho que ocorre no 1º ciclo. De qualquer forma, o período é entendido como a fonte das aprendizagens menores em ciclos posteriores e, por isso, os agentes controlam a qualidade através da intervenção direta em coadjuvação:

No caso particular do 1º Ciclo de Ensino Básico, as provas constituíram-se como um estímulo para a criação de condições para que todos os alunos tenham acesso, na sua plenitude, ao currículo definido para o 1º ciclo do ensino básico e à possibilidade de garantir uma monitorização sistemática da Educação Física neste ciclo. (Comunicado CNAPEF e SPEF)

O segundo objetivo emergente é o do prestígio da própria disciplina. Aqui, as provas parecem ganhar relevo, porque equiparam a importância da disciplina à de outras – já que também é testada – e, uma vez que as PA são consideradas inovadoras, são usadas para valorizar a EF portuguesa em fóruns internacionais:

Resumidamente, eu julgo [...] que este seminário foi importante para nós conhecermos realmente as boas práticas associadas ao aproveitamento dos resultados das provas de aferição [...]. Acho que também serviu claramente para reforçar a autoridade das provas de aferição de Educação Física como instrumento que valoriza a Educação Física. (Dirigente CNAPEF, Seminário: Provas de aferição de Educação Física)

Nesse processo, alguns discursos dos agentes da EF parecem censurar (ou caracterizam com carga negativa) as eventuais atitudes de resistência, na medida em que são contraproducentes ao ganho de prestígio social:

Esta é outra questão [...] que também já foi afluída pelos, pelos colegas que, que falaram antes de mim. [...] tem a ver [...] com alguma dificuldade de mobilização dos professores, não sei se são os mais velhos, são os que têm a cabeça mais antiga, diria eu, porque também há, há professores novos de idade que têm a cabeça muito velha, portanto, não sei se são os professores mais velhos que têm dificuldade. (Dirigente SPEF, Seminário: Provas de aferição de Educação Física)

4.2 VALIDADE E CONFIABILIDADE

Neste ponto, apresentamos os resultados que caracterizam a capacidade de as provas avaliarem o currículo (validade) e a consistência das avaliações entre anos, escolas e avaliadores (confiabilidade).

Quanto à validade, os dados sugerem que as provas não são adequadas a medir o desenvolvimento do currículo, porque estão longe de fazer a sua cobertura integral e estruturam-se em tarefas clínicas e com pouca ligação à realidade das aprendizagens.

Primeiro, os resultados apontam que as provas se organizam por um processo de redução curricular. As PA do 2º ano abordam todas as áreas de aprendizagens essenciais (perícias e manipulações, deslocamentos e equilíbrio e jogos), mas isso ocorre por cobertura de apenas um quarto das habilidades que os alunos devem desenvolver.

Por exemplo, a versão 1 da prova de 2019 (IAVE, 2019, pp. 6, 8-9), na área das perícias e manipulações, solicita tarefas de: rolar o arco (tarefa 3); saltar à corda (tarefa 4); toques de sustentação com raquete (tarefa 7); drible da bola (tarefa 9); passe e recepção da bola (tarefa 10) – apesar de a indicação que o aluno recebe seja de pontapear a bola e não de passar.

Aqui, é significativo que as tarefas sejam realizadas individualmente, quando os objetivos curriculares preveem, também, situações de estafeta e pares.

Já os objetivos programáticos contemplam, também, por exemplo e resumidamente:

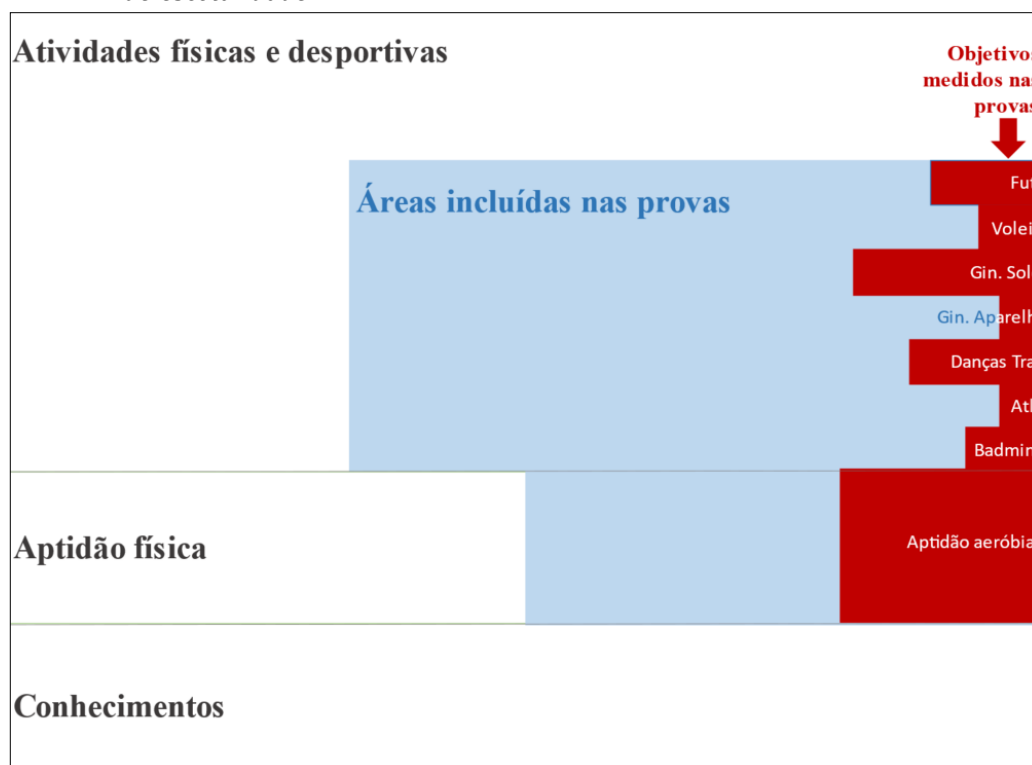
Lançar a bola com uma mão e com as mãos; lançar uma bola para cima da cabeça e recebê-la acima da cabeça e perto do solo; rolar a bola nos membros superiores unidos e em extensão; fazer toque de sustentação numa bola-balão, com a cabeça e com os membros superiores; lançar o arco na vertical; passar por dentro de um arco a rolar; conduzir a bola com os pés; receber uma bola com as duas mãos; passar a bola com o pé ao companheiro; impulsionar a bola com a raquete a várias alturas e com e sem ressalto no solo. (DGE, 2023b, pp. 3-5)

Nos 5º e 8º anos (Figura 3), a relação entre as áreas-subáreas-objetivos do currículo e aqueles avaliados é mais deficitária: são avaliadas duas das três áreas (os conhecimentos não constam das provas); na área

da aptidão física, é avaliada uma das duas sub-áreas e 20% dos objetivos (a resistência); e, na área das atividades físicas, as tarefas cobrem entre 50% a 67% das sub-áreas de aprendizagens essenciais e abordam cinco de 17 matérias no 5º ano e cinco ou seis de 16 no 8º ano, com o pormenor de a maioria das matérias não se repetir entre PA de anos diferentes.

Figura 3

Relação entre o currículo e a cobertura do currículo das PA no 8.º ano de escolaridade



A observação mais fina revela que as provas apenas permitem aceder entre 7% a 39% dos objetivos de aprendizagem das sub-áreas. Por exemplo, na ginástica de aparelhos são avaliados dois objetivos de 27 (apenas objetivos no aparelho boque/plinto e nenhum da trave, do minitrampolim e da barra fixa). A tarefa da prova do 8º ano de 2022 (IAVE, 2022c, p.4) é a seguinte:

Após corrida de balanço e com receção equilibrada no colchão:

- salta ao eixo no boque;
- salta entre mãos no plinto ou no boque.

Já as Aprendizagens Essenciais do 8º ano (DGE, 2023a) e os objetivos que lhe estão inerentes consideram (resumidamente) que o aluno alcança o nível introdutório de desempenho (o que satisfaz a condição de sucesso na sub-área da ginástica) quando:

1. Coopera com os companheiros nas ajudas, paradas e correções que favoreçam a melhoria das suas prestações, garantindo condições de segurança, pessoal e dos companheiros, e colabora na preparação, arrumação e preservação do material.

2. Realiza, após corrida de balanço e chamada a pés juntos [...] Salto de eixo no boque, [...] Salto entre-mãos no boque ou plinto transversal, [...] Cambalhota à frente no plinto longitudinal [...].
3. No mini-trampolim, realiza [...] Salto em extensão.
4. Em equilíbrio elevado (na trave baixa), realiza um encadeamento das [...] habilidades, Marcha à frente e atrás [...] Marcha na ponta dos pés [...] Meia volta, em apoio nas pontas dos pés [...] Salto a pés juntos [...].
5. Na barra fixa, realiza [...] Subida para apoio ventral [...] seguida de rolamento à frente, [...] Balanço atrás e à frente [...] Balanços laterais. (DGE, 2023c, pp. 13-14)

Segundo, os resultados sugerem que as condições e as tarefas das provas têm pouca ligação com a realidade da aprendizagem.

Em algumas escolas, as provas avaliam as tarefas na certeza de que os alunos puderam aprender com acesso aos materiais e aos espaços desportivos adequados, mas, noutras, as provas avaliam tarefas com recurso a condições materiais e espaciais a que se sabe que os alunos não tiveram acesso prévio – assim, o que é aferido é a capacidade de os alunos realizarem as tarefas como novidade:

Houve muitas escolas que não estavam apetrechadas e houve uma necessidade de aquisição de material. Isso obrigou a fazer aqui um levantamento rápido dos materiais necessários para a aplicação da prova. Nessa altura, havia muitos materiais em muitos fornecedores que estavam esgotados, portanto, foi até à última mesmo que chegaram as encomendas para nós conseguirmos suprir a necessidade de materiais. (Representante Agrupamento C)

A própria natureza do que é medido vai mudando, em função de estratégias diferentes, de escola para escola, quanto ao contacto anterior com as tarefas previsíveis. Algumas escolas promovem a experimentação (e a aprendizagem intensiva) das tarefas, enquanto outras escolas a evitam:

A questão que foi colocada [...] em relação à, à situação de se permitir o treino da tarefa [...] no primeiro ciclo, a mim parece-me ser uma situação em que o resultado final pode realmente ser alterado, porque são tarefas de complexidade muitas vezes reduzida [...] e a experimentação da tarefa várias vezes pode levar a que os alunos realmente consigam uma prestação diferente e que não reflete o trabalho realizado pela escola e sim o adestramento prévio. (Dirigente de Associação de Professores de EF, Seminário: Provas de aferição de Educação Física)

Encontra-se, também, a sensibilidade das provas à participação dos alunos, já que alguns dados sugerem que estes se recusam a realizar, em particular, a tarefa do vai-vem:

Aptidão aeróbica, alunos que não fazem ou que se recusam a fazer, que é um fenómeno, ou que não fazem, digamos, o que lhes é exigido relativamente ao teste de aptidão aeróbica, para a sua idade e sexo: 21%. (Representante Agrupamento A)

Além disso, é preciso ter em consideração que o(a) docente da turma é responsável pela formação de grupos quando as tarefas são coletivas e que isso influencia o desempenho dos alunos e das alunas, com algumas evidências a sugerirem que essa seleção pode levar a grupos de nível.

No conjunto, está expressa a sensibilidade da medição a fatores que parecem ir metamorfoseando o significado do que é avaliado (Tabela 3).

Tabela 3

Sensibilidades emergentes nas PA em relação às condições de aprendizagem

Condição	Aprendizagem	Prova
Material	Sem algum material (em algumas escolas).	Com todo o material.
Espaço	Não específico (em algumas escolas).	Sala de ginástica e pavilhão.
Contacto com a tarefa	No processo.	Como treino/produto em algumas escolas.
Corpo discente	Todos os alunos.	Alguns alunos recusam participar.
Grupos	Grupos para a aprendizagem.	Eventuais grupos de nível.
Apoio aos alunos	Diferenciação por medidas universais, adicionais e seletivas.	Sem diferenciação.

Em terceiro lugar, as próprias tarefas das provas parecem divergir daquelas que expressam a aprendizagem, ora porque não há correspondência com os objetivos curriculares (e.g., as provas não avaliam o nível elementar, que se aplica ao 8º ano, e recusam situações de exercício que estão previstas nas aprendizagens essenciais), ora porque vários dos objetivos curriculares permitem que os alunos usem criatividade e opção individual, podendo escolher alguns elementos, mas as provas não o permitem (e.g., todos os alunos realizam a mesma sequência de ginástica), ora porque excluem aprendizagens (e.g., não avaliam a função de guarda-redes), ora, ainda, porque algumas tarefas são executadas em função das escolhas dos professores, o que significa que avaliam a aprendizagem que estes preferem evidenciar (e.g., na prova do 8º ano de 2022, na dança, é o docente que seleciona a tarefa – ou danças tradicionais ou sociais – e, depois, uma dança específica).

Além disso, as tarefas decorrem em ambiente quase clínico, com observação de alunos num tempo determinado, sem a intervenção de outros alunos nas tarefas individuais (o que retira aprendizagens que devem ser evidenciadas, como as ajudas, o manuseamento do material e a arbitragem), os quais, a menos que estejam em prova, devem manter-se sentados e reservados (na procura de um ambiente estéril que dificilmente existe nas aulas de EF).

Quanto à confiabilidade, os resultados sugerem que a observação pelos classificadores e as condições de observação estão sujeitas a oscilações que diminuem a confiabilidade.

Os critérios de correção usam expressões que podem ser interpretadas e que deixam parte da avaliação ao cuidado da interação social entre os observadores. Por exemplo, nos jogos desportivos coletivos, a atribuição do código 20 em vez do 30 assenta em situações em que, “ocasionalmente”, o aluno não tem o desempenho esperado, e o código 10 é atribuído pelo critério “frequentemente”, sem que haja descritores para o significado de “ocasionalmente” e “frequentemente”.

Além disso, os critérios de correção assentam na expectativa de conhecimento tácito e imediato dos professores-classificadores. Um exemplo recorrente é o critério “Realiza todas as habilidades gímnicas da sequência de acordo com o padrão global de execução, combinando-as com fluidez” (IAVE, 2022d, p.3).

Isso significa que, quanto à prova do 2º ano, os professores corretores devem dominar integralmente o conhecimento e a interpretação de 48 objetivos. Quanto ao 8º, só nos jogos desportivos coletivos, o valor é de 156 objetivos.

Na organização do processo, os professores conhecem os objetivos a avaliar com muito pouca antecedência antes do início da prova, e os procedimentos de avaliação obrigam a que dois professores-classificadores concordem com a avaliação. Isto parece estar na origem de dificuldades que os próprios docentes consideram limitar a confiabilidade dos resultados:

Tivemos conhecimento das provas na sexta-feira anterior à realização das provas [que se realizaram na segunda-feira seguinte] [...]. Isto impediu que houvesse, naturalmente, uma aferição prévia de critérios e descritores e processos de codificação. [...] Houve uma completa insensibilidade [no secretariado de exames] [...], porque isto não pode ser informado vinte minutos antes de começar as provas! Vinte minutos antes de começar as provas eu tenho na minha mão quais são as notações e quais são as fichas. Isto não pode ser. (Representante Agrupamento A)

Este conjunto de situações junta-se ao carácter instantâneo das tarefas e leva a que os professores e agentes da EF tenham um forte sentimento de necessidade de formação ou proponham o registo dos desempenhos em vídeo e que surjam algumas dúvidas quanto à possibilidade de se usar os dados recolhidos para a tomada de decisão:

Não meto, nem nós, nenhum de nós em termos de departamento mete, as mãos no fogo relativamente à validade destes resultados. Dou-vos uma ideia, eu estou sentado com a minha colega a vermos um conjunto de destrezas que é pedido numa determinada tarefa e nós, em cima da hora, estamos a ajustar, com produtos que não são permanentes e que mesmo, repete um, repete duas vezes,

com grande falibilidade relativamente, digamos, ao tipo de produto que é feito. (Representante Agrupamento A)

Ainda assim, alguns agentes salientam que algumas escolas procuraram antecipar estes problemas, através de treino dos professores, com base em provas anteriores, e que o IAVE, reconhecendo as dificuldades e limitações da avaliação de produtos irrepetíveis, considera que as observações têm qualidade, porque usou processos estatísticos que o atestam.

4.3 AS PROVAS E AS PRÁTICAS LETIVAS

Os resultados sugerem que as PA são vistas como momentos fiáveis de diagnóstico que são usados para guiar os processos de ensino, nomeadamente influenciando a perceção dos professores acerca do estado da aprendizagem dos alunos e levando a medidas de planeamento e à implementação de estratégias e atividades. Além disso, as provas são valorizadas por se mostrarem úteis para gerar e justificar processos colaborativos entre professores.

As PA parecem ganhar primazia como referência de planeamento. Apesar de, numa das escolas, serem vistas como informação complementar, a ser coordenada com as conferências curriculares internas, os discursos evidenciam que as provas estão na origem de várias decisões pedagógicas, didáticas e curriculares que se sobrepõem a outros momentos de decisão:

Os resultados das provas de aferição permitem perceber quais são as matérias em que os alunos obtiveram melhores ou piores resultados, sendo que isso deverá ajudar os professores a definirem as prioridades de desenvolvimento dos alunos; permitem estabelecer uma comparação dos resultados dos alunos, com os resultados da avaliação inicial nas matérias em causa, interpretar a evolução dos alunos e refletir sobre a necessidade de melhorar as práticas pedagógicas e as aprendizagens. Permitem perceber se há resultados discrepantes que merecem uma atenção particular e que podem estar relacionados pelas opções de cada professor ou com as condições de realização das atividades. Permitem perceber se se justifica investir mais numa escola do que noutra, no mesmo agrupamento. (Representante Agrupamento B)

Aliás, essa sobreposição também emerge dissonante entre o que o IAVE e o Ministério da Educação aconselham, no sentido em que as provas são informações que complementam as dinâmicas da EF:

Agora vamos lá em conjunto, os professores do grupo desta escola, ver como é que nós utilizamos estas fotografias. Não é para substituir as fotografias que nós já tínhamos, é sim para enriquecer as fotografias que nós já tínhamos [...]. (Representante do Ministério da Educação)

Não obstante, também complementam outro tipo de aconselhamento, em que as práticas letivas devem ser guiadas muito diretamente pelas informações das PA:

Há muitas formas de trabalhar isto, mas, normalmente, damos os RIPA aos alunos e eles marcam, tipo com a tinta vermelha o “não conseguiu”, com tinta verde o “conseguiu” [...] e suponham que olharam para o vosso REPA turma, a vossa turma tem dificuldades na ginástica [...], o que vocês vão fazer é: descobrem no REPA turma que há dificuldades específicas na ginástica, que até está dividida por subdomínios, até podem saber exatamente quais são os pontos que devem trabalhar e, depois, vão ver nos alunos, nos RIPA ligados à ginástica, quais são as dificuldades específicas de cada aluno. (Representante do IAVE)

Por outro lado, os discursos vincam a expectativa de que as provas se constituam como informação formativa que ajude os encarregados de educação e os alunos a intervir sobre a sua realidade e, nesse sentido, aparentam pretender papéis ativos para estes agentes.

Contudo, no 5º e 8º anos, é relevante notar que a linguagem que poderia respeitar essas expectativas, parece impedir qualquer leitura correspondente entre os resultados e os objetivos de aprendizagem, porque a designação e o significado da avaliação das provas não encontram qualquer paridade com as Aprendizagens Essenciais (DGE, 2023a).

As provas recorrem às expressões “consegue”, quando o aluno realiza a tarefa sem falhas e “consegue, mas”, quando há pequenas variações/falhas relativamente ao esperado. Já as aprendizagens recorrem às rubricas “introdução” e “elementar” para designar o desempenho dos alunos.

Além do uso de palavras diferentes, o seu significado não tem qualquer correspondência, uma vez que as provas designam conjuntos de objetivos que não têm paralelo em qualquer nível de desempenho (exemplo na Tabela 4).

Tabela 4

Relação de linguagem e significado das PA do 5.º e 8.º anos e do currículo

Critérios de classificação das Provas de aferição 2018 8.º ano		Currículo
Designação	Consegue:	Introdução:
Significado	Faz batimentos em clear de acordo com o padrão global de execução, garantindo a continuidade da jogada e mantendo: o volante ao alcance do companheiro; trajetórias altas e compridas, a uma distância de 6 a 8 metros do companheiro. (IAVE, 2018 p. 3)	O aluno é capaz de cooperar com os companheiros para alcançar êxito, conhece o jogo, as regras e conhece e sabe usar os batimentos clear, lob, serviço curto e serviço comprido, usa a posição base e sabe deslocar-se em função do jogo e da deslocação do volante. (DGE, 2023c, pp. 33-34)

Como resultado, as informações das PA não têm conexão evidente com o sucesso na EF, e o facto de um determinado aluno ter sido avaliado com “Consegue” em todas as tarefas não significa que tenha sucesso na disciplina (e.g., como referido anteriormente, mesmo que obtenha sucesso na tarefa de ginástica de aparelhos da PA tem, em aula, de evidenciar aprendizagens em outros três aparelhos para ter sucesso nessa sub-área).

5. DISCUSSÃO

Os resultados sugerem que as PA em EF também servem propósitos não escritos de prestígio da disciplina e sindicância das práticas letivas do 1º ciclo. Além disso, apontam que as PA dificilmente medem o que se propõem (o que significa que as intenções formativas que são explicitadas estão em causa), mas, ainda assim, são usadas para orientar as práticas de ensino. Nessa medida, há que pôr estes resultados em contraste com as dimensões epistemológicas da EF.

Primeiro, as provas emergem enquanto discurso consensual de autoridade dos professores de EF relativamente ao 1º ciclo e como evidência de que esse ciclo é a raiz dos problemas de aprendizagem da EF. Através desse consenso, as provas dissuadem o debate. Esse mesmo potencial autoritário havia sido identificado acerca de outras avaliações externas da aprendizagem (Berkovich & Benoliel, 2020; Bart, 2015; Lopo, 2021).

De qualquer forma, esta investigação levanta questões sobre esse autoritarismo. Repare-se que os resultados nacionais das PA indicam níveis elevados de sucesso no 1º ciclo (IAVE, 2022b). Por outro lado, todas as escolas consideradas promotoras de boas práticas implementaram projetos em que os 3º e 4º anos de escolaridade são assistidos por um professor de EF. E os resultados nacionais das PA do 5º e 8º anos revelam que as aprendizagens dos alunos decrescem ao longo dos ciclos.

Então, as aprendizagens menores surgem, nas três escolas, na presença de professores especialistas e, ainda assim, os agentes da EF sustentam que o 1º ciclo é um período em que a EF não é ensinada e é a raiz dos problemas vindouros de aprendizagem (também evidente em Silva & Dinis, 2022). Pois bem, é esta autoridade que transforma as PA em conhecimento ideológico, em vez de pedagógico – e é isso que define a sua função autoritária, por cedência da sua função crítica e reflexiva.

Depois, este estudo oferece evidências de um fenómeno novo. Se alguns estudos sugerem que as avaliações externas das aprendizagens podem levar ao afunilamento do currículo, em que os professores ensinam o que esperam que seja avaliado (Fernandes, 2019), o que aqui emerge é um processo de encobrimento curricular, já que a avaliação de um excerto do currículo parece levar a decisões pedagógicas acerca da integralidade do mesmo – a que se junta a falta de correspondência entre o sucesso nas provas e o sucesso na EF. Portanto, o conhecimento parece adquirir uma forte vertente de exame, de virtual e de previsão.

Quanto ao planeamento, note-se que a EF funciona na expectativa de caducidade do conhecimento acerca das aprendizagens dos alunos (e.g., o período de verão pode levar à perda de competência e alguns saltos desenvolvimentais a impactos no desempenho), e as aprendizagens que, entretanto, os alunos desenvolvem nas aulas são pré-requisitos para níveis de desempenho mais complexos. É por isso que implementa o período de avaliação inicial (em setembro e outubro de cada ano letivo), para recolher informações sobre todas as matérias nucleares que os alunos devem aprender e, a partir daí, planejar o ano letivo. Também por esse motivo, realizam-se conferências curriculares ao longo do ano, para ajustar o planeamento – sempre na base de níveis de desempenho e critérios de interpretação semelhantes para todas as escolas.

Ora, este estudo sugere que apesar das PA devolverem resultados antigos (relativos a maio do ano letivo anterior), que chegam aos professores depois de estes avaliarem formativamente todas as matérias (em outubro), essas informações são consideradas muito relevantes para organizar o ano letivo e alterar os currículos, o que corresponde a um boicote da essência da EF, para promover a afirmação da disciplina através de processos de utilidade pedagógica menor, mas que, sendo socialmente valorizados, podem contribuir para o prestígio social. Assim, o conhecimento que as provas oferecem adquire um carácter hierarquizado e de dever.

Portanto, surpreendentemente, a irrelevância pedagógica das provas transforma-se em intencionalidade educativa. Aqui importa salientar que esta investigação sugere que as PA não têm validade, porque, além de recorrerem a ambientes estéreis e a tarefas clínicas, podem estar a avaliar as condições de aprendizagem (os materiais, o espaço de aula, a experimentação prévia das tarefas) ou as opções dos professores no dia da prova (as tarefas que os docentes selecionam para serem avaliadas), em vez das aprendizagens em si.

Esta surpresa é aumentada quando se considera a confiabilidade das provas. Ou os resultados das PA são confiáveis, tal como defende o representante do IAVE e, portanto, os professores de EF são capazes de ter um olhar avaliativo semelhante, ainda que não tenham treinado e coordenado a sua ação, e isso significa que as PA são irrelevantes, porque o que os docentes observam em aula é tão ou mais válido do que o que observam nas provas; ou os resultados não são confiáveis – é isso que este estudo sugere – e o seu uso perde qualquer capacidade pedagógica.

Neste contexto, deve-se, pelo menos, considerar que os dados que sugerem que algumas aprendizagens dos alunos melhoraram entre 2019 e 2022 (IAVE, 2022b), apesar da pandemia que metamorfoseou radicalmente a disciplina (González-Calvo et al., 2022; Silva & Dinis, 2023; Varea & González-Calvo, 2021), são apenas mais um indicador da validade e confiabilidade diminuídas das PA.

Veja-se o caso específico da tarefa de aptidão física (5º e 8º anos). O teste de desempenho aeróbio vai-vem terá sido realizado por todos os alunos, no âmbito da sua escolaridade obrigatória, durante o mês de maio ou junho, de acordo com as orientações curriculares (Jacinto et al., 2001). Aliás, esses resultados podem ser introduzidos numa base de dados

(FitEscola) que permite aferir com os resultados nacionais. Por exemplo, em maio de 2022, os alunos de 14 anos da zona norte tiveram níveis de sucesso de 86% (dos quais, 13% relevaram desempenho de perfil atlético, nível que as PA não medem). Ora, os níveis de sucesso das PA do 8º ano (realizadas em maio) foram de 74,5%, o que sugere que se limitam a conhecer o que já se conhece, embora com acuidade diminuída e que esta repetição ineficaz justifica o grau de recusa dos alunos em realizar a tarefa – eis, provavelmente, uma situação em que o esforço dos alunos não está colocado na tarefa (para quê fazê-lo, se o terão feito recentemente em aula?). Em conjunto, isso sugere que o conhecimento que advém das PA está orientado para o dever e para as dificuldades que os alunos apresentam.

Em suma, emerge um traço hegemónico, também reconhecível em outras iniciativas de avaliação externa e massiva das aprendizagens, e o discurso autoritário que é usado pelos agentes da EF parece estar dirigido mais à comunidade de decisores que têm uma linguagem neo-liberal (comparação, objetividade, responsabilização), para que a disciplina se prestigie por aquilo que é valorizado por esses agentes dominantes, do que verdadeiramente aos alunos, às famílias e aos professores – o que parece alavancar a descaracterização epistémica da disciplina.

6. CONCLUSÃO

Propusemo-nos debater a epistemologia da EF, a partir do racional de que as práticas letivas assentam nas dimensões de conhecimento do cuidado, do sentido, da coevolução, do jogo e do movimento e de que as PA, determinadas externamente, podem influenciar essa estrutura epistémica.

Pois bem, este estudo sugere que as PA estão na origem de metamorfoses da epistemologia da EF. A disciplina pode estar a passar de uma dimensão de cuidado para uma de cuidado clínico, em que as debilidades dos alunos são atendidas em função de excertos curriculares; de uma dimensão do sentido para uma do dever, focada na garantia de comportamentos motores mínimos; de uma dimensão do jogo e do movimento para uma da repetição da tarefa, com limitação da criatividade e desarranjo das aprendizagens curriculares; e de uma dimensão da coevolução para uma de linearidade previsível – nesse sentido, poderão estar a não traduzir a integralidade do significado das condutas motoras (Parlebas, 2001).

Naturalmente, os resultados em que estas conclusões se suportam merecem ser olhados cuidadosamente, já que expressam o produto de uma comunidade institucionalizada e de escolas que foram consideradas como boas práticas, sem que, contudo, estejam claros os critérios que permitiram essa categorização. Outras comunidades de EF, menos formais, e outras escolas, com práticas divergentes, podem levar a perfis epistémicos diferentes dos que aqui emergiram.

Também é preciso ter em conta que o recurso que aqui fizemos às metáforas, apesar de contribuir para a constituição de um mapa cognitivo de reflexão, não deixa de ser uma síntese da realidade, que pode não captar toda a riqueza da disciplina.

Ainda assim, este estudo sugere que há um grau de alteração epistémica da EF, orientada para o que é considerado prestigiante, e que o conhecimento que alimenta as interações pedagógicas pode estar a mudar, talvez de forma pouco coerente com o pensamento que esteve na origem da estrutura da disciplina – nesse sentido, as PA apropriam-se dos recursos internos de cada aluno para responder a interesses ideológicos e, nesse processo, a EF auto-neo-liberaliza-se. Isto parece traduzir uma regressão epistémica, devendo notar-se que a disciplina já integra processos que permitem aferir, de forma coerente e coordenada, a aprendizagem dos alunos e as dinâmicas formais de reconstrução curricular (as conferências).

Assim, a pergunta que se justifica é a de saber porque é que esses processos não são socialmente valorizados e têm de ser ultrapassados por processos menores.

Se a intenção é a de fornecer informações ao corpo discente, às famílias e às escolas, para que haja intervenção pedagógica atempada, pode-se, então, recorrer ao processo de avaliação inicial (em que o corpo docente recolhe informações acerca do desempenho de todos os alunos nas matérias essenciais em todos os anos letivos e por referência aos níveis de desempenho nacionalmente uniformizados). Nesse sentido, basta transformar os registos dos docentes em informação útil aos agentes – com a vantagem de se usar uma linguagem significativa e comum e de se dirigir tanto às dificuldades como às potencialidades dos alunos (e não prioritariamente às dificuldades, como consta na definição de objetivos das provas) (IAVE, 2022a).

Se, além disso, existe a necessidade de compreender e comparar a situação nacional ou regional das aprendizagens, a disciplina pode, também aí, recorrer aos seus processos, considerando a uniformização nacional dos critérios de sucesso e dos níveis de desempenho. Os protocolos de avaliação inicial são construídos ao nível das escolas e, por isso, não se pode esperar que sejam semelhantes. Contudo, os objetos que avaliam são os mesmos e as condições de avaliação também se encontram determinadas nas orientações curriculares (por exemplo, concurso a pares, prova, jogo). Ainda assim, pode criar-se um guião que assessore os grupos de EF para que, transversalmente, as tarefas protocoladas permitam avaliar coerentemente os níveis de desempenho.

Os dados recolhidos poderiam ser depositados e disponibilizados numa plataforma centralizada, o que permitiria o conhecimento nacional quanto às aprendizagens na disciplina de EF – como já acontece em muitas escolas no caso da avaliação da aptidão física –, o que permitiria a comparação da situação das escolas com o panorama nacional, havendo a vantagem de incluir os 12 anos de escolaridade obrigatória.

Tais iniciativas não transtornam o funcionamento epistémico da disciplina e, pelo contrário, podem constituir-se como pontos de valorização de práticas eficazes, aferidas e transversais – o que satisfaz a necessidade de reconhecimento social da EF.

Há, ainda, que considerar, face à aparente necessidade de resolver questões relacionadas com as condições materiais e espaciais (também em Silva & Dinis, 2023), que será melhor adotar procedimentos diretos

de mapeamento da situação e definir as condições mínimas de funcionamento da EF. Que instalações desportivas existem em cada escola? Quais são as condições mínimas de espaço e materiais para que a EF se cumpra? Que escolas asseguram coadjuvação no 1º ciclo? Esse conhecimento pode assistir a decisão política (aos níveis micro, meso e macro) acerca do investimento público. Além disso, também neste caso seria possível recolher dados em todas as escolas, em colaboração com os professores de EF, a partir de orientações que coordenassem o levantamento de informação.

Esse ponto da situação será condição para libertar as escolas de comparações sobre aprendizagens perante circunstâncias diferentes e, eventualmente, iníquas – por exemplo, os resultados seriam relacionados com o acesso a espaços, materiais e coadjuvação.

A efetivação destas propostas carece de recursos. Porém, oferece uma perspetiva comunitarista, assente em instrumentos de coordenação de agentes que já se encontram em campo e que já utilizam parte do seu tempo para as PA. Assim, a sugestão é que todos os recursos gastos nas PA devem ser orientados para a melhoria dos processos internos já existentes – esses parecem ser verdadeiramente inovadores, apesar de existirem há mais de 20 anos – e para o estabelecimento de novos que completem a informação, o que permite retomar o perfil epistémico da EF.

REFERÊNCIAS

- Bardin, L. (2011). *Análise de conteúdo* (L. A. Reto & A. Pinheiro, Trads.). Edições 70.
- Bart, D. (2015). Le discours de la recherche dans le programme international de suivi des acquis des élèves: Un mode d'exposition pour un effet d'imposition?. *Revue Française de Pédagogie*, 191(2), 89-100. <https://doi.org/10.4000/rfp.4764>
- Berger, G. (2009). A investigação em educação: Modelos sócioepistemológicos de inserção de inserção institucional. *Educação, Sociedade & Culturas*, (28), 175-192.
- Berkovich, I., & Benoliel, P. (2020). The educational aims of the OECD in its TALIS insight and lesson reports: Exploring societal orientations. *Critical Studies in Education*, 61(2), 166-179. <https://doi.org/10.1080/17508487.2017.1370428>
- Borrvalho, A. M., Fialho, I., & Cid, M. (2015). A triangulação sustentada de dados como condição fundamental para a investigação qualitativa. *Revista Lusófona de Educação*, 29(29), 53-69. <https://revistas.ulusofona.pt/index.php/rleducacao/article/view/5094>
- Debeer, D., Buchholz, J., Hartig, J., & Janssen, R. (2014). Student, school, and country differences in sustained test-taking effort in the 2009 PISA reading assessment. *Journal of Educational and Behavioral Statistics*, 39(6), 502-523. <https://doi.org/10.3102/1076998614558485>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, 21(4), 248-266. <https://doi.org/10.1080/10627197.2016.1236677>

Direção-Geral de Educação [DGE] (2023a). *Aprendizagens Essenciais: Educação Física (8.º ano)* [Currículo escolar]. Direção-Geral de Educação. https://www.dge.mec.pt/sites/default/files/Curriculo/Aprendizagens_Essenciais/3_ciclo/educacao_fisica_3c_8a_ff.pdf

Direção-Geral de Educação [DGE] (2023b). *Aprendizagens Essenciais: Educação Física (Anexo I)* [Currículo do Ensino Básico]. Direção-Geral de Educação. https://www.dge.mec.pt/sites/default/files/Curriculo/Aprendizagens_Essenciais/1_ciclo/anexo1_ef.pdf

Direção-Geral de Educação [DGE] (2023c). *Aprendizagens Essenciais: Educação Física (Anexo III)* [Currículo do Ensino Básico e do Ensino Secundário]. Direção-Geral de Educação. https://www.dge.mec.pt/sites/default/files/Curriculo/Aprendizagens_Essenciais/3_ciclo/anexo3_ef.pdf

Direção-Geral de Estatísticas da Educação. (2022). *Resultados escolares por disciplina: 3.º Ciclo do ensino público geral (Portugal continental – 2011/12-2020/21)*. Direção-Geral de Estatísticas da Educação e Ciência. [https://www.dgeec.mec.pt/np4/413/%7B\\$clientServletPath%7D/?newsId=899&fileName=DGEEC_Resultados_por_disciplina_3ciclo_s3.pdf](https://www.dgeec.mec.pt/np4/413/%7B$clientServletPath%7D/?newsId=899&fileName=DGEEC_Resultados_por_disciplina_3ciclo_s3.pdf)

Fernandes, D. (2019). Avaliações externas e aprendizagens dos alunos: Uma reflexão crítica. *Linhas Críticas*, 25, e24579. <https://doi.org/10.26512/lc.v25.2019.24579>

Fialho, N., & Fernandes, D. (2011). Práticas de professores de Educação Física no contexto da avaliação inicial numa escola secundária do concelho de Lisboa. *Meta: Avaliação*, 3(8)168-191.

Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK Summit. In A. Berry, A. Frieddrichsen & J. Loughran (Eds.), *Reexamining pedagogical content knowledge in science education* (pp. 28-42). Routledge.

Gilligan, C. (2016). *In a different voice: Psychological theory and women's development*. Harvard University Press.

González-Calvo, G., Barba-Martín, R. A., Bores-García, D., & Hortigüela-Alcalá, D. (2022). The (virtual) teaching of Physical Education in times of pandemic. *European Physical Education Review*, 28(1), 205-224. <https://doi.org/10.1177/13563336X211031533>

Grawitz, M. (1993). *Méthodes des Sciences Sociales* (9e Éd.). Éditions Dalloz.

Guba, E. G., & Lincoln, Y. S. (1998). Competing paradigms in qualitative research. In N. K. Denzin, & Y. S. Lincoln, *The landscape of qualitative research: Theories and issues* (pp. 195-219). SAGE Publications.

Hatch, J. A. (2002). *Doing qualitative research in education settings*. State University of New York Press.

Instituto de Avaliação Educativa. (2018). *Prova de Aferição de Educação Física – Critérios de Classificação*. Instituto de Avaliação Educativa. <https://iave.pt/wp-content/uploads/2020/04/PA-EF84-2018-CC-V1.pdf>

Instituto de Avaliação Educativa. (2019). *Prova de Aferição de Expressões Físico-Motoras*. Instituto de Avaliação Educativa. <https://iave.pt/wp-content/uploads/2020/01/PA-EFM28-2019-GUIAO-V1.pdf>

Instituto de Avaliação Educativa [IAVE] (2022a). *Informação-prova: 5.º ano de escolaridade*. Instituto de Avaliação Educativa. <https://iave.pt/wp-content/uploads/2022/11/IP-PA-5-2023-2.pdf>

Instituto de Avaliação Educativa [IAVE] (2022b). *Provas de Aferição do Ensino Básico 2022: Resultados Nacionais*. Instituto de Avaliação Educativa. https://iave.pt/wp-content/uploads/2022/12/Relatorio-Provas-de-Afericao_Resultados-Nacionais_2022_Final.pdf

Instituto de Avaliação Educativa [IAVE] (2022c). *Prova de Aferição de Educação Física*. Instituto de Avaliação Educativa. https://iave.pt/wp-content/uploads/2022/06/PA-EF84-2022-GUIAO-V1_net.pdf

Instituto de Avaliação Educativa [IAVE] (2022d). *Prova de Aferição de Educação Física – Critérios de Classificação*. Instituto de Avaliação Educativa. https://iave.pt/wp-content/uploads/2022/06/PA-EF84-2022-CC_net.pdf

Jacinto, J., Carvalho, L., Comédias, J., & Mira, J. (2001). *Programa de Educação Física: 10.º, 11.º e 12.º* [Programa de formação geral dos cursos científico-humanísticos]. Ministério da Educação. https://www.dge.mec.pt/sites/default/files/Secundario/Documentos/Documentos_Disciplinas_novo/Curso_Ciencias_Tecnologias/Educacao_Fisica/ed_fisica_10_11_12.pdf

Jensen, D. (2016). Metaphors as a bridge to understanding educational and social contexts. *International Journal of Qualitative Methods*, 5(1), 36-54. <https://doi.org/10.1177/160940690600500104>

Lopo, T. T. (2021). The political decision on Portugal's entry into PISA: A research note. *Policy Futures in Education*, 19(6), 723-729. <https://doi.org/10.1177/1478210320971537>

Madaus, G., Russell, M., & Higgins, J. (2009). *The paradoxes of high stakes testing: How they affect students and their parents, teachers, principals, schools and society*. Information Age Publishing.

Méndez, A., Valero, A., & Casey, A. (2010). What are we being told about how to teach games? A three-dimensional analysis of comparative research into different instructional studies in Physical Education and School Sports. *Revista Internacional de Ciencias del Deporte*, VI(18), 37-56. <https://doi.org/10.5232/ricyde2010.01803>

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th Ed.). Jossey-Bass.

Neves, R., Almeida, D. M., & Capela, A. J. (2023). Educação Física no 1.º Ciclo do Ensino Básico (1.º CEB), Provas de Aferição e pais: Perceções globais. *Motrivivência*, 35(66), 1-13. <https://doi.org/10.5007/2175-8042.2023.e93994>

Neves, R., Morato, D., & Capela, A. (2024). Os professores do 1.º CEB e as Provas de Aferição em Educação Física: Importância, preparação e expectativas. *Conhecimento & Diversidade*, 16(41), 1-16. <https://doi.org/10.18316/rcd.v16i41.11363>

Öhman, M., & Quennerstedt, A. (2017). Questioning the no-touch discourse in Physical Education from a children's rights perspective. *Sport, Education and Society*, 22(3), 305-320. <https://doi.org/10.1080/13573322.2015.1030384>

Parlebas, P. (2001). *Juegos, deporte y sociedad: Léxico de praxiología motriz* (F. G. Campo Román, Trad.). Editorial Paidotribo.

Peräkylä, A. (2005). Analyzing talk and text. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (pp. 869-886). SAGE Publications.

Pools, E., & Monseur, C. (2021). Student test-taking effort in low-stakes assessments: Evidence from the English version of the PISA 2015 science test. *Large-scale Assessment in Education*, 9(10). <https://doi.org/10.1186/s40536-021-00104-6>

Portaria n.º 223-A/2018, de 3 de agosto. Diário da República, 1.ª série – N.º 149 (1.º suplemento). <https://diariodarepublica.pt/dr/detalhe/portaria/223-a-2018-115886163>

Schlosser, A., Neeman, Z., & Attali, Y. (2019). Differential performance in high versus low stakes tests: Evidence from the Gre Test. *The Economic Journal*, 129(623), 2916-2948. <https://doi.org/10.1093/ej/uez015>

Silva, N. M. (2019). *Liderar organizações complexas: O caso das escolas*. Chiado Books.

Silva, N. M. (2022). Grounded Theory para iniciantes: Contributo para a investigação em educação. *Cadernos de Pesquisa*, 52, e08563. <https://doi.org/10.1590/198053148563>

Silva, N. M. (2024). *Material complementar: As provas de aferição e a epistemologia da Educação Física*. Harvard Dataverse. <https://doi.org/10.7910/DVN/GXT4EX>

Silva, N. M., & Dinis, S. P. (2022). A Educação Física nos tempos de crise: Que estatuto emerge?. *Revista Portuguesa de Investigação Educacional*, (23), 1-33. <https://doi.org/10.34632/investigacaoeducacional.2022.11400>

Silva, N. M., & Dinis, S. P. (2023). As práticas letivas na educação física durante a pandemia: Expectativas e realidades. *Educação e Pesquisa*, 49, e234016. <https://doi.org/10.1590/S1678-4634202349262256>

Stake, R. E. (2009). *A arte da investigação com estudos de caso* (A. M. Chaves, Trad., 2.ª Ed.). Fundação Calouste Gulbenkian.

Varea, V., & González-Calvo, G. (2021). Touchless classes and absent bodies: Teaching Physical Education in times of COVID-19. *Sport, Education and Society*, 26(8), 831-845. <https://doi.org/10.1080/13573322.2020.1791814>

i Centro de Investigação em Educação e Psicologia, Universidade de Évora, Portugal.
nuno.silva@uevora.pt
<https://orcid.org/0000-0002-3574-7679>

ii Centro de Investigação em Educação e Psicologia, Universidade de Évora, Portugal.
soniapsidinis@hotmail.com
<https://orcid.org/0000-0002-4492-0322>

Toda a correspondência relativa a este artigo deve ser enviada para:

Nuno Miranda e Silva
nuno.silva@uevora.pt

Recebido em 06 de Abril de 2023
Aceite para publicação em 12 de abril de 2024
Publicado em 30 de setembro de 2024

The low-stakes tests and the epistemology of Physical Education

ABSTRACT

In Portugal, low-stakes testing has been introduced in the discipline of Physical Education with the purpose of providing information that leads to better learning, which translates into new knowledge made available to the agents and, therefore, has implications for the epistemology of the discipline. However, this process is taking place without scientific scrutiny, which is relevant because we know that external assessments of learning can have unforeseen effects and lead to discourses of ideological authority over schools. Therefore, it is important to ask what discourses justify the low-stakes tests, what the validity and reliability of the tests are, and what the interactions between the knowledge offered by the tests and the teaching practices are. To answer this question, an investigation was implemented within the interpretative paradigm in the form of an instrumental case study and with a qualitative approach, using document analysis and content analysis. The results suggest that the low-stakes tests are also implemented with the objectives of controlling Physical Education practices in the first four years of compulsory education and of social valorization of Physical Education, measuring a small part of the school curriculum through very unreliable processes. Yet, the results are used to influence the generality of the teaching process, which translates to curriculum concealment. It is concluded that the evidence supports the epistemological metamorphosis of Physical Education from care to clinical, from meaning to duty, from game to repetition, and from coevolution to linearity.

Keywords: Physical Education; External assessment of learning; Low-stakes tests; Epistemology.

Las pruebas de evaluación externa y la epistemología de la Educación Física

RESUMEN

En Portugal, las pruebas de evaluación externa se introdujeron en la disciplina de Educación Física con el propósito de proporcionar información que conduzca a un mejor aprendizaje, lo que se traduce en nuevos conocimientos puestos a disposición de los agentes y, por lo tanto, tiene implicaciones en la epistemología de la disciplina. Sin embargo, este proceso ocurre sin escrutinio científico, lo que es relevante porque sabemos que las evaluaciones externas del aprendizaje pueden tener efectos imprevistos y llevar a discursos de autoridad ideológica sobre las escuelas. Por lo tanto, es importante preguntarse sobre cuáles son los discursos que justifican las pruebas de evaluación externa, cuál es la validez y confiabilidad de las pruebas y cuáles son las interacciones entre los conocimientos ofrecidos por las pruebas y las prácticas de enseñanza. Para responder a esta pregunta, se implementó una investigación, dentro del paradigma interpretativo, en la modalidad de estudio de caso instrumental y con enfoque cualitativo, utilizando análisis documental y análisis de contenido. Los resultados sugieren que las pruebas de evaluación externa también se implementan con objetivos de control de las prácticas de Educación Física en los cuatro primeros años de la enseñanza obligatoria y de valorización social de la Educación Física, miden una pequeña parte del currículo escolar, a través de procesos muy poco fiables. Aun así, los resultados se utilizan para influir en la generalidad del proceso de enseñanza, lo que se traduce en ocultación curricular. Concluimos que las evidencias apoyan la metamorfosis epistemológica de la Educación Física, del cuidado a la clínica, del significado al deber, del juego a la repetición y de la coevolución a la linealidad.

Palabras clave: Educación Física; Evaluación externa del aprendizaje; Pruebas de evaluación; Epistemología.