

O OLHAR DE QUEM AVALIA: SUPERVISÃO PEDAGÓGICA, GÊNERO E AVALIAÇÃO DO DESEMPENHO DOCENTE EM PORTUGAL

THE EVALUATOR'S GAZE: PEDAGOGICAL SUPERVISION, GENDER AND
TEACHER PERFORMANCE APPRAISAL IN PORTUGAL

Ciências Humanas • 27/10/2026

REGISTRO DOI: [10.70773/revistatopicos/792989312](https://doi.org/10.70773/revistatopicos/792989312)

Andressa Leoni Leandro da Silva Borges¹

RESUMO

O artigo discute como expectativas de gênero podem interferir na observação de aulas e no *feedback* da avaliação do desempenho docente no ensino não superior português. Nesse regime, um avaliador externo observa aulas e atribui uma classificação com efeitos na carreira, numa profissão em que as mulheres são maioria. O objetivo é discutir essas expectativas e propor um critério que ajude o supervisor-avaliador a examinar o próprio juízo. O estudo é teórico-conceitual, de abordagem qualitativa, natureza básica e objetivo exploratório. Desenvolve-se por revisão narrativa, com análise interpretativa da legislação portuguesa e da literatura sobre supervisão pedagógica, gênero e avaliação de professores. A análise percorre os atributos julgados numa aula: a autoridade e a gestão da turma, o cuidado e a relação pedagógica, o domínio científico e a passagem do registro ao *feedback*. Os estudos internacionais encontram diferenças de gênero em direções opostas, conforme o contexto, o avaliador e o atributo. Por isso, a direção do efeito é tratada como questão em aberto. Sustenta-se que o risco está nas expectativas sobre o que cada professor deve mostrar. Propõe-se a dupla leitura do juízo avaliativo: descrever primeiro o que aconteceu na aula e depois perguntar se o registro seria o mesmo caso a aula tivesse sido dada por um colega de outro gênero. A proposta ainda não foi testada empiricamente e serve à reflexão do supervisor e à conversa com o professor antes da classificação.

Palavras-chave: Supervisão pedagógica; Avaliação do desempenho docente; Gênero; Observação de aulas; Feedback.

ABSTRACT

This article discusses how gender expectations may affect classroom observation and feedback in teacher performance appraisal in Portuguese schools. In this system, an external evaluator observes

lessons and assigns a rating with career consequences, in a profession where women are the majority. The aim is to discuss these expectations and to propose a criterion that helps the supervisor-evaluator examine their own judgment. The study is theoretical and conceptual, with a qualitative approach, basic in nature and exploratory in aim. It is developed as a narrative review, with an interpretive analysis of Portuguese legislation and of the literature on pedagogical supervision, gender and teacher evaluation. The analysis follows the attributes judged in a lesson: authority and classroom management, care and the pedagogical relationship, subject mastery, and the passage from the written record to feedback. International studies find gender differences in opposite directions, depending on the context, the evaluator and the attribute. For this reason, the direction of the effect is treated as an open question. The article argues that the risk lies in the expectations about what each teacher should display. It proposes a double reading of the evaluative judgment: first describe what happened in the lesson, then ask whether the record would be the same had the lesson been taught by a colleague of another gender. The proposal has not yet been tested empirically and is meant to support the supervisor's reflection and the conversation with the teacher before the rating.

Keywords: Pedagogical supervision; Teacher performance appraisal; Gender; Classroom observation; Feedback.

1. INTRODUÇÃO

Imagine-se uma aula de Ciências Naturais numa turma do 3.º ciclo do ensino básico, observada por um avaliador externo. No meio da atividade, a professora interrompe o trabalho para pedir silêncio, repete o pedido em voz mais alta e retoma a explicação quando a

turma se acalma. No registro da observação, o avaliador pode anotar que a docente *teve dificuldade em manter a turma em ordem* ou que *impôs as regras com firmeza*. As duas anotações descrevem o mesmo episódio, mas a primeira aponta um problema e a segunda, uma competência. Se quem estivesse diante da turma fosse um professor, com os mesmos gestos, a escolha poderia ser outra. A cena é hipotética e serve para situar o problema deste estudo: o que o avaliador vê e escreve depende da aula, e pode depender também do que ele espera de quem a dá.

Em Portugal, a avaliação do desempenho docente (ADD) dos educadores de infância e dos professores dos ensinos básico e secundário é regulada pelo Decreto Regulamentar n.º 26/2012, que a página oficial do Diário da República indicava como em vigor em outubro de 2026. Em situações definidas, a avaliação inclui a observação de aulas por um avaliador externo, docente que integra uma bolsa de avaliadores, pertence ao mesmo grupo de recrutamento do avaliado (a área disciplinar e o nível em que o professor é recrutado) e está em escalão igual ou superior ao dele. Escalão é cada um dos dez patamares da carreira, a que corresponde um índice remuneratório. A classificação final, de 1 a 10, é convertida em menções, as classificações qualitativas Excelente, Muito Bom, Bom, Regular e Insuficiente. Só obtém Excelente quem fica no percentil 95 ou acima, e Muito Bom quem fica no percentil 75 ou acima, entre outros requisitos, e essas duas menções antecipam a progressão na carreira (Portugal, 2012a). O regime pode mudar. Em março de 2026, a revisão do Estatuto da Carreira Docente (ECD) estava em negociação entre o ministério e os sindicatos, organizada em sete temas, e o modelo de avaliação do desempenho era o último deles, ainda por discutir (Carreira, 2026).

Com isso, a observação de aulas fica num ponto de tensão bem conhecido da supervisão pedagógica. Obras de referência da supervisão em Portugal, como as de Alarcão e Tavares (2003) e de Vieira e Moreira (2011), tratam a supervisão como acompanhamento orientado para o desenvolvimento e a aprendizagem do professor. Na ADD, a mesma visita que poderia servir a esse acompanhamento termina numa nota com efeitos na carreira. Quem redigiu o decreto reconheceu a tensão: o texto declara a intenção de potenciar a dimensão formativa da avaliação e afirma que, com o novo regime, “minimizam-se conflitos entre avaliadores e avaliados” (Portugal, 2012a, p. 855). Seabra et al. (2021a), num questionário aplicado a educadores de infância da rede pública portuguesa, descrevem uma experiência de supervisão muito ligada à ADD, ao lado de alguma abertura a formas de supervisão mais horizontais.

O juízo do avaliador recai sobre uma profissão majoritariamente feminina. Segundo a Direção-Geral de Estatísticas da Educação e Ciência (DGEEC), no ano letivo 2023/2024 as mulheres eram 99,0% dos docentes da educação pré-escolar em Portugal Continental, 88,4% no 1.º ciclo, 73,2% no 2.º ciclo e 71,5% no 3.º ciclo e no ensino secundário (DGEEC, [s. d.]). No inquérito da Organização para a Cooperação e Desenvolvimento Económico (OCDE) sobre professores e diretores (Teaching and Learning International Survey, TALIS) de 2024, 75% dos professores portugueses do 3.º ciclo eram mulheres, acima da média de 70% da OCDE (OCDE, 2025). Na direção das escolas, a proporção era outra: na edição de 2018 do mesmo inquérito, 43% dos diretores de escolas com 3.º ciclo eram mulheres, para 74% de professoras (OCDE, 2019). Quanto à bolsa de avaliadores externos, as fontes consultadas não informam a sua composição por sexo.

A pergunta pedagógica é se, e como, expectativas de gênero sobre o que é ser bom professor podem entrar no que o avaliador vê, registra e devolve ao professor. Gênero é entendido aqui, em primeiro lugar, na formulação de Scott (1986), para quem a categoria designa um elemento das relações sociais fundado nas diferenças percebidas entre os sexos e uma forma primeira de dar sentido a relações de poder. Butler (2003), em obra publicada originalmente em 1990, recusa que o gênero seja expressão de uma essência anterior a ele: o gênero se produz pela repetição de atos e gestos que as normas sociais reconhecem como próprios de homens ou de mulheres. Nenhuma das duas autoras tratou da avaliação de professores, e a extensão das suas categorias a esse objeto é feita por este estudo. Dela decorre uma consequência prática. O foco recai sobre as expectativas a respeito de professoras e professores, e não sobre supostas diferenças no modo como uns e outras ensinam. Se o gênero se faz na repetição de condutas reconhecidas, o registro do avaliador, ao premiar umas e estranhar outras, participa dessas normas, mesmo sem intenção.

Nos Estados Unidos, Grissom e Bartanen (2022) encontraram notas de observação mais altas para professoras do que para professores, e Campbell e Ronfeldt (2018), notas mais baixas para os professores homens. Beg, Fitzpatrick e Lucas (2021) encontraram, em Gana, país da África Ocidental, diretores de escolas primárias que subavaliavam as professoras. Nenhum desses estudos é português, e nenhum permite dizer o que acontece na ADD. Chama-se aqui viés a uma diferença de juízo que a prática observada na aula não explica, o que significa que uma diferença de notas entre grupos, sozinha, não basta para caracterizá-lo. Como a evidência internacional aponta em direções diferentes, ela não oferece uma correção que se possa

aplicar de fora, e o problema volta ao próprio avaliador e à supervisão.

Pergunta-se, então: como as expectativas de gênero sobre o trabalho docente podem influenciar o que o avaliador vê e diz na observação de aulas e no *feedback*, e o que a supervisão pedagógica pode fazer para que o professor avaliado seja julgado pelo que fez na aula? O objetivo é discutir como expectativas de gênero podem atravessar a observação de aulas e o *feedback* na avaliação do desempenho docente em Portugal e propor um critério que ajude o supervisor-avaliador, o docente que observa e classifica e que também deveria apoiar, a examinar o próprio juízo. Para isso, o estudo caracteriza a observação de aulas na ADD como momento em que apoio e controle se encontram na mesma pessoa, examina como três atributos valorizados na observação podem ser lidos de modo diferente conforme o gênero de quem dá a aula, discute como essas leituras passam ao registro e ao *feedback* e propõe um critério de análise do juízo avaliativo para uso na supervisão.

Sustenta-se que a observação de aulas e o *feedback* dependem do juízo de quem observa e que esse juízo pode incorporar expectativas de gênero sobre a autoridade, o cuidado e o domínio da matéria. Conforme o atributo julgado e o contexto da escola e da disciplina, essas expectativas podem desfavorecer professoras ou professores. Uma supervisão orientada para o desenvolvimento oferece meios para torná-las visíveis e discutíveis, desde que o supervisor-avaliador trate o próprio juízo como objeto de reflexão.

2. METODOLOGIA

Trata-se de um estudo teórico-conceitual, desenvolvido por revisão narrativa. Na classificação de Prodanov e Freitas (2013), a sua natureza é básica, porque procura gerar conhecimento sobre um problema sem aplicação prática prevista: aqui, compreender como o juízo do avaliador pode incorporar expectativas de gênero e formular uma proposição a esse respeito, sem aplicá-la a um caso. A abordagem é qualitativa no sentido que os mesmos autores dão ao termo, centrada na interpretação dos fenômenos e na atribuição de significados, e não na medida. Quanto ao objetivo, o estudo é exploratório. Gil (2002) reserva essa designação às pesquisas que buscam maior familiaridade com um problema, para torná-lo mais explícito ou para construir hipóteses, e é essa a situação de um tema que, no regime português, ainda pede mais perguntas bem formuladas do que verificações. O procedimento é bibliográfico, ou seja, apoiado em material já elaborado, sobretudo livros e artigos científicos (Gil, 2002), e a análise segue uma perspectiva reflexivo-analítica.

O mesmo autor adverte que a pesquisa apoiada em fontes secundárias tende a reproduzir os erros delas e recomenda cotejar fontes diversas e examinar as condições em que cada dado foi obtido (Gil, 2002). Por isso, a legislação foi lida no texto publicado no Diário da República, e os números atribuídos a cada estudo foram tirados do próprio estudo, e não de quem o cita. Três conjuntos de fontes foram analisados: a legislação portuguesa da ADD, em especial o Decreto Regulamentar n.º 26/2012 e o Despacho n.º 13981/2012, que fixa os parâmetros nacionais da observação de aulas e os modelos de registro, acompanhada de dados estatísticos da DGEEC e da OCDE; estudos sobre supervisão pedagógica e sobre a ADD em Portugal, incluindo teses e dissertações; e estudos internacionais sobre gênero na avaliação de professores e em outras

avaliações de desempenho, usados como ilustração, sempre com o país, o nível de ensino e o tipo de avaliador indicados.

Os estudos foram localizados em outubro de 2026 no Repositório Científico de Acesso Aberto de Portugal (RCAAP), que reúne artigos, dissertações e teses de repositórios portugueses, na SciELO, no Education Resources Information Center (ERIC) e no Crossref, além de buscas gerais na web. Usaram-se descritores em português e em inglês, como *avaliação do desempenho docente*, *observação de aulas*, *supervisão pedagógica*, *género*, *teacher evaluation*, *classroom observation* e *gender bias*. A busca serviu para orientar a leitura e não teve pretensão de revisão sistemática ou exaustiva.

A análise foi organizada pelos atributos que o avaliador julga numa aula, derivados dos parâmetros científico e pedagógico da observação (Portugal, 2012b): a autoridade e a gestão da turma, o cuidado e a relação pedagógica, o domínio científico e o rigor, e a passagem do registro ao *feedback*. Cada atributo foi examinado com perguntas transversais, vindas da supervisão orientada para o desenvolvimento, do gênero como categoria de análise, da história da feminização do magistério, da teoria da congruência de papel, que vem da psicologia social, e das exigências de justiça de uma avaliação com efeitos na carreira. A perspectiva reflexivo-analítica consistiu em confrontar a literatura com situações típicas de supervisão, apresentadas como hipotéticas, para examinar o que cada achado permite dizer sobre a ADD. Estudos correlacionais não foram lidos como prova de causa, e estudos com outros avaliadores (diretores, estudantes, gestores, médicos) ou de outros países não foram tomados como descrição do que ocorre em Portugal.

Nas bases consultadas, não se localizaram estudos que analisem, com dados portugueses, diferenças de gênero nas classificações, na observação de aulas ou no *feedback* da avaliação do desempenho docente. Os trabalhos que incluem o gênero tratam-no como característica dos respondentes, como o de Alves et al. (2018), sobre as percepções de 240 docentes acerca da ADD, em que as associações com o sexo dos respondentes não foram estatisticamente significativas.

3. O QUE SE JULGA NUMA AULA OBSERVADA

Na ADD, a observação de aulas não abrange todos os docentes: o Decreto Regulamentar n.º 26/2012 torna-a obrigatória em período probatório, nos 2.º e 4.º escalões, para quem pretende a menção de Excelente e para quem recebeu Insuficiente, e exclui os docentes contratados a termo (Portugal, 2012a). Quando ocorre, a observação soma 180 minutos, distribuídos por pelo menos dois momentos, e responde por 70% da dimensão científica e pedagógica, que por sua vez vale 60% da classificação final. Os parâmetros são nacionais. O científico e o pedagógico têm o mesmo peso, e o pedagógico divide-se em elementos didáticos e relacionais. O avaliador registra o que observa num guião de caráter indicativo, com colunas para registros positivos e negativos, e preenche depois uma grelha de observação, o instrumento em que atribui a classificação de 1 a 10 e que tem campos para uma apreciação global e para recomendações (Portugal, 2012b).

Segundo Machado et al. (2012), que reconstroem o percurso normativo da ADD, até 1974 a avaliação cabia a um inspetor, exterior à escola e à classe docente, e entre 1974 e 2008 prevaleceu a autoavaliação. A partir de 2008, passou a ser feita por pares da

própria escola, o que levantou dúvidas sobre a capacidade de um colega avaliar outro com objetividade e sem envolvimento pessoal. O modelo de 2012 respondeu com o avaliador externo, que os autores descrevem como uma exterioridade mitigada: continua a ser um par, mas vem de fora. Duas das perguntas que deixam em aberto interessam aqui, a de saber se a avaliação pelos pares é compatível com quotas e classificações e a do seu lugar em modelos de supervisão.

Mais de dez anos depois, Guedes (2023) inquiriu 527 avaliadores externos, 111 diretores de agrupamentos de escolas (unidades de gestão que reúnem várias escolas sob uma direção comum) e de escolas não agrupadas, e 49 diretores de centros de formação de associação de escolas (CFAE), estruturas que organizam a formação contínua dos professores e acolhem as bolsas de avaliadores. A autora conclui que o modelo não mostrou efeito na melhoria do ensino nem no desenvolvimento profissional, que o sigilo levado ao extremo tornou o sistema opaco e que a avaliação funcionou como instrumento de regulação da progressão, gerador de arbitrariedades. As perguntas de Machado et al. continuam, portanto, sem resposta.

Outros trabalhos portugueses olham para o instrumento e para quem o usa. Ferreira (2016) pediu a 68 observadores que aplicassem duas vezes o guião legal a quatro aulas gravadas de Educação Física e encontrou, na maioria dos casos, acordo insignificante ou mediano entre eles, que atribui à subjetividade do instrumento. Marques (2023), num estudo de caso num CFAE, relata que a formação em avaliação externa perdeu 12 horas a partir de 2019 e que a observação passou a ser operada como um conjunto de procedimentos pronto a aplicar. Cavalheiro (2024), que inquiriu 52

avaliadores externos das escolas associadas ao CFAE Lezíria do Tejo, parte da constatação de que a contestação à ADD se deve, em parte, à falta de formação de quem avalia. Fora de Portugal, num sistema de avaliação de professores em larga escala, Wind, Jones e Bergin (2021) mostram que diferenças de severidade entre diretores bastam para mudar a categoria de desempenho atribuída aos professores.

Nenhum desses trabalhos trata do gênero, mas todos indicam por onde ele poderia entrar. Numa avaliação com efeitos na carreira, a justiça do resultado depende de que pessoas diferentes, diante da mesma aula, cheguem a juízos parecidos. Quando a classificação fica nas mãos de uma única pessoa, com um instrumento de descritores amplos e uma formação curta, cresce o espaço em que o juízo pode ser influenciado pelo que não está na aula. As subseções seguintes examinam esse espaço atributo por atributo.

3.1. Firmeza ou Rigidez: A Autoridade e a Gestão da Turma

Antes de examinar cada atributo, lembramos que a autoridade, o cuidado e o domínio da matéria são qualidades esperadas de qualquer professor, homem ou mulher. O que se discute daqui em diante são expectativas a respeito de professoras e de professores, capazes de levar o avaliador a ver e a valorizar de forma diferente a mesma conduta. Associar a autoridade ao masculino e o cuidado ao feminino é o que se quer pôr em exame, não uma descrição do que uns e outras fazem em sala de aula.

O Despacho n.º 13981/2012 inclui entre os elementos relacionais observados o “funcionamento da aula com base em regras que acautelem a disciplina” (Portugal, 2012b, p. 35375), ao lado do envolvimento dos alunos e dos estímulos à melhoria da

aprendizagem. Esses elementos valem pouco na conta, 10% do parâmetro pedagógico. O vocabulário dos níveis de desempenho, porém, alcança o parâmetro inteiro. Para o nível Excelente, o descritor pedagógico é “Segurança inequívoca tanto em termos relacionais como didáticos” (Portugal, 2012b, p. 35376). Segurança pode designar uma conduta observável, como dar uma instrução uma única vez e ser seguido, ou uma impressão sobre a pessoa, e é na passagem de uma à outra que as expectativas podem agir.

A teoria da congruência de papel, proposta por Eagly e Karau (2002) para explicar o preconceito contra mulheres em posições de liderança, permite dar forma a esse risco. Segundo os autores, quando o papel de gênero atribuído a um grupo parece incongruente com o papel exercido, os membros desse grupo tendem a ser vistos como menos aptos para ele, e a mesma conduta tende a ser avaliada de forma menos favorável quando é exibida por eles. Eagly e Karau trataram de liderança em organizações, não de docência. Aplicar a teoria à autoridade em sala de aula é extensão deste estudo, que a usa como hipótese de leitura e não como explicação demonstrada.

Correll et al. (2020) analisaram avaliações escritas de desempenho numa empresa de tecnologia dos Estados Unidos e encontraram mulheres descritas como agressivas demais e homens como brandos demais, além de um comportamento, o de assumir o comando, associado às notas mais altas só no caso dos homens. Em grupos focais, Billick, Rassos e Ginsburg (2022) ouviram 22 médicos residentes de medicina interna na Universidade de Toronto, no Canadá. As residentes relataram *feedback* contraditório sobre confiança e assertividade, que ora lhes pedia mais assertividade, ora menos, algo raro nos relatos dos homens. A professora da cena

inicial, que elevou a voz para retomar a turma, estaria sujeita a uma leitura desse tipo, em que a fronteira entre firmeza e excesso muda conforme quem age.

Expectativas desse tipo também pesam sobre os homens. Jóhannesson, Ottesen e Bjarnadóttir (2024) entrevistaram cinco vezes, ao longo dos dois primeiros anos de profissão, sete professores homens recém-formados em escolas de ensino obrigatório da Islândia. A escola e as famílias esperavam deles que fossem disciplinadores naturais, a quem os alunos respeitariam por serem homens. A expectativa deu a alguns uma autoridade de partida, mas pesou sobre eles como iniciantes, e nem todos se encaixavam nela. Para os autores, a gestão da turma pede cuidado e atenção ao detalhe, qualquer que seja o gênero de quem ensina. Levada para a observação, a mesma expectativa pode fazer o avaliador cobrar de um professor uma autoridade que não cobraria de uma colega, ou ler como fraqueza nele uma condução da turma mais negociada.

Esse raciocínio enfrenta uma objeção séria: a gestão da turma depende em parte da turma, e não apenas de quem ensina. Campbell e Ronfeldt (2018), em análise secundária dos dados do projeto Measures of Effective Teaching, nos Estados Unidos, encontraram notas de observação mais baixas para professores homens e para docentes com turmas com muitos alunos negros, hispânicos, rapazes e de baixo desempenho. Com uma subamostra de professores distribuídos por sorteio entre as turmas, concluíram que essas diferenças dificilmente refletem diferenças reais de qualidade. Delgado e Sartain (2026), comparando o mesmo professor de um ano para outro, mostram que turmas com melhor desempenho e menos indisciplina rendem notas sistematicamente

mais altas, dadas tanto por avaliadores como por estudantes. Parte do que o avaliador lê como autoridade pode ser, portanto, a turma que coube ao professor. A questão de gênero não desaparece com isso, mas passa a envolver duas perguntas. A primeira é quem recebe quais turmas, o que pode ter dimensão de gênero e, na ADD, é hipótese sem dados. A segunda é como o avaliador interpreta a mesma agitação conforme quem está diante dela.

O resultado de Campbell e Ronfeldt contraria, aliás, o estereótipo do disciplinador, já que as notas mais baixas foram dos homens, num instrumento que também julga a gestão da turma. O dado é norte-americano e nada diz sobre Portugal, mas impede que se trate a autoridade como atributo em que a desvantagem cairia sempre sobre as mulheres.

3.2. O Cuidado Esperado e o Cuidado Notado

Um aluno com dificuldade pede ajuda no fim de uma tarefa. Quem ensina agacha-se junto à carteira, refaz o exemplo com outros números e espera que o aluno termine sozinho. O avaliador pode registrar o gesto como atenção à diferenciação pedagógica, pode tomá-lo como algo que qualquer bom professor faz e nem o anotar, ou pode notá-lo como gesto excepcional. Resta saber se a escolha entre essas leituras varia com o gênero de quem ensina.

Nos parâmetros da observação, o cuidado não aparece com esse nome, mas entra por outras portas. O Despacho n.º 13981/2012 pede que o professor crie um ambiente educativo assente em valores reconhecidos e trate os alunos com dignidade, e inclui nos elementos relacionais o envolvimento dos alunos e os estímulos para que melhorem (Portugal, 2012b). Entra também pela

apreciação global e pelas recomendações da grelha, onde o avaliador escreve com as próprias palavras.

A história de uma profissão majoritariamente feminina ajuda a perguntar que qualidades a escola aprendeu a esperar de professoras e de professores. A feminização da docência, fenômeno de origem social que persiste no Brasil e em Portugal, ainda é compreendida de forma frágil pelos futuros professores dos dois países (Silva; Leal da Costa, 2025). Em Portugal, como se viu, a proporção de mulheres cai à medida que se avança nos níveis de ensino, de quase a totalidade na educação pré-escolar para cerca de sete em cada dez no 3.º ciclo e no secundário (DGEEC, [s. d.]). Como o avaliador externo pertence ao mesmo grupo de recrutamento do avaliado, na educação pré-escolar a observação é, com grande probabilidade, feita por uma mulher a outra mulher, enquanto nos grupos do secundário a composição dos pares de avaliador e avaliado é mais variada. Os efeitos dessa diferença sobre o juízo não foram estudados.

Pela teoria da congruência de papel, o cuidado estaria sujeito a uma assimetria. Se é esperado das professoras, pode passar despercebido nelas e ser notado como mérito num professor, e a sua falta pode ser lida com mais severidade numa professora, por romper uma expectativa. Não se localizou estudo sobre a ADD que teste essa hipótese, mas há estudos no ensino superior, com estudantes no papel de avaliadores, que a tornam plausível e que também mostram o seu limite. Evans et al. (2024) analisaram avaliações de monitores de doutorado em Engenharia da Universidade Cornell, nos Estados Unidos. Os monitores homens tinham maior probabilidade de receber notas melhores em todas as questões, e cada gênero parecia ser valorizado por comportamentos

congruentes com o papel que a sociedade lhe atribui. As monitoras recebiam mais comentários positivos sobre comunicação e mais comentários sobre apoio aos estudantes. Parker e Savage (2026) encontraram, nos comentários escritos de estudantes, docentes mulheres descritas em linguagem relacional e de mentoria e docentes homens em termos de autoridade, rigor e desafio intelectual. Como nesse segundo trabalho quatro docentes negros de uma universidade do Sudeste dos Estados Unidos analisaram os comentários que eles próprios receberam num ano letivo, ele vale aqui só como ilustração.

O limite dessa leitura aparece no experimento de Khokhlova, Lamba e Kishore (2023), com 671 estudantes universitários, numa amostra que os autores associam ao Sul da Ásia e ao Oriente Médio. A mesma aula foi dada por avatares que só diferiam no gênero, e os masculinos foram vistos como mais entusiasmados e expressivos. Se o calor humano fosse sempre atribuído às mulheres, o resultado seria outro. As expectativas de gênero não formam um repertório fixo, e o que o avaliador valoriza num professor e numa professora pode variar de um atributo e de um contexto para outro.

Como nenhum desses estudos envolve um professor que avalia um colega, o que eles oferecem à ADD é uma pergunta: o avaliador registra da mesma maneira, com o mesmo peso, o gesto de cuidado de uma professora e o de um professor? Para uma supervisão orientada para o desenvolvimento, o cuidado tem ainda outro estatuto. É parte do que o supervisor deveria praticar com o professor que observa, e não só uma qualidade a pontuar nele. Quando o mesmo docente acompanha e classifica, o cuidado que a supervisão lhe pede convive com a distância que a classificação impõe, e é aí que a tensão entre controle e apoio se torna concreta.

3.3. Domínio da Matéria: De Onde Vem a Confiança do Avaliador

No parâmetro científico, que vale metade da observação, o avaliador julga o domínio dos conteúdos disciplinares e, com peso menor, os conhecimentos de língua portuguesa que enquadram a aprendizagem desses conteúdos. O nível mais alto descreve um domínio pleno (Portugal, 2012b). A exigência de que o avaliador pertença ao mesmo grupo de recrutamento garante que ele tenha competência para esse julgamento. Tem também uma consequência menos visível, porque as expectativas sobre quem domina que matéria circulam dentro do próprio grupo, entre colegas que compartilham a mesma formação e as mesmas referências.

O exemplo mais claro de desvantagem feminina no julgamento da competência vem de Gana, país da África Ocidental. Em escolas primárias de 20 distritos ganeses, Beg, Fitzpatrick e Lucas (2021) compararam a avaliação que os diretores faziam dos professores com a autoavaliação deles e com medidas de eficácia baseadas na aprendizagem dos alunos. Professoras e professores avaliavam-se de forma equivalente, e as professoras eram, pelas medidas objetivas, mais eficazes. Mesmo assim, os diretores tinham cerca de 11 pontos percentuais menos de probabilidade de classificá-las como mais eficazes, e um experimento com questionário confirmou a tendência. O próprio artigo traz dados que pedem três cautelas. A estimativa principal é significativa apenas ao nível de 10%, a medida objetiva de eficácia inclui o efeito da escola, e o contexto é o inverso do português, porque na amostra ganesa as mulheres eram 22% dos professores e 15% dos diretores.

Grissom e Bartanen (2022) encontraram a direção oposta nos Estados Unidos. Nos primeiros oito anos do sistema de avaliação de professores do Tennessee, as professoras tiveram, em média, notas de observação cerca de 0,30 desvio-padrão mais altas. A diferença persistiu quando os autores consideraram outras medidas de eficácia, como os ganhos de aprendizagem dos alunos e a sua assiduidade, o que é compatível com viés. Afora diferenças associadas ao ano e à disciplina lecionados, encontraram poucas explicações para essa vantagem, e nenhum efeito de avaliador do mesmo gênero. Neste último ponto, Chi (2023), com dados de um grande distrito do Sudeste dos Estados Unidos, chegou a outro resultado: os professores tinham pequenos ganhos na nota quando compartilhavam o gênero, ou a raça, com quem os observava. Mesmo onde há dados, portanto, o efeito do gênero do avaliador é controverso.

As avaliações que estudantes do ensino superior fazem dos docentes envolvem outro avaliador e outro nível de ensino, mas ajudam a ver em que condições a competência de quem ensina é julgada de modo diferente. A revisão sistemática de Stoesz et al. (2023), com 63 artigos, encontrou forte apoio à persistência de um viés favorável aos docentes homens. Mengel, Sauermann e Zölitz (2019), num contexto em que os estudantes eram distribuídos por sorteio entre docentes, encontraram notas mais baixas para as mulheres, sem diferença nas notas ou nas horas de estudo dos alunos, e um viés maior em disciplinas matemáticas. Já a meta-análise de Felkey e Batz-Barbarich (2021) localizou a diferença a favor dos homens só em Economia, sem evidência nas demais ciências sociais. Binderkrantz e Bisgaard (2024), em cerca de 125 mil avaliações de uma universidade dinamarquesa, encontraram pouca diferença global e uma afinidade de gênero, com estudantes

avaliando melhor os docentes do próprio gênero. LaPaglia, Miller e Protexter (2022) acrescentam um dado sobre o erro. Num experimento com vídeo de uma aula, a docente foi avaliada abaixo do docente apenas quando a exposição era pouco fluente. O conjunto não sustenta uma regra geral, mas sugere onde olhar: nas áreas em que a competência é associada a um gênero e nos momentos de falha, em que um tropeço pode ser lido como incompetência num caso e como acidente noutro.

Essa variação não surpreende quando se toma o gênero como categoria relacional. Na leitura de Scott (1986), o gênero não atua isolado, mas combinado com outras relações de poder, e Campbell (2023), na Carolina do Norte, mostra uma dessas combinações: com eficácia semelhante, os administradores deram notas de observação mais baixas a professoras negras do que a professoras brancas. Na ADD, o gênero combina-se ao menos com o grupo de recrutamento e com o nível de ensino. Separar o efeito de cada um exigiria dados que, nas fontes consultadas, não estão publicados. Para o supervisor-avaliador, a consequência prática é outra. Diante de uma explicação imprecisa ou de uma resposta hesitante a uma pergunta imprevista, cabe perguntar se a confiança, ou a desconfiança, no domínio do professor veio da aula ou de quem a deu.

3.4. Do Registro Ao *Feedback*: O Que Chega Ao Professor

O guião de observação anexo ao Despacho n.º 13981/2012 determina que, tendo em conta as circunstâncias concretas de ensino, “os registos derivados da observação devem incidir nos comportamentos do docente avaliado” (Portugal, 2012b, p. 35376). A regra separa duas coisas que costumam vir juntas no registro: a descrição do que o professor fez e a apreciação de quem ele parece

ser. Escrever que *a professora repetiu a instrução e a turma iniciou a tarefa ao fim de dois minutos* é descrever um comportamento. Escrever que *a professora revela insegurança* é atribuir uma qualidade à pessoa. A segunda frase é mais curta e cabe melhor numa grelha, e é também a mais exposta às expectativas.

O modelo que Correll et al. (2020) propõem a partir das avaliações escritas da empresa já referida ajuda a ver como isso acontece. Segundo os autores, as crenças de gênero podem afetar dois momentos do juízo, o que o avaliador nota e o que ele valoriza. A habilidade técnica era descrita da mesma forma para homens e mulheres; a diferença estava em outros traços e no peso que tinham na nota. Na educação médica dos Estados Unidos, Heath et al. (2024) fizeram uma análise secundária de um ensaio sobre formação de avaliadores em que 77 médicos observaram residentes simulados, com desempenho roteirizado e igual para homens e mulheres. As residentes receberam notas mais baixas, e os comentários sobre o que tinham feito mal usaram com mais frequência termos considerados femininos. O gênero do avaliador não se associou à nota. É o caso mais próximo da observação de aulas, porque alguém observa diretamente um desempenho e o registra, mas trata-se de medicina e de desempenho encenado.

Também o que se diz ao professor, e não só a nota, pode ser moldado por expectativas. Jampol e Zayas (2021) verificaram, em experimentos, que os participantes inflavam o retorno dado a mulheres em relação à avaliação que tinham feito sem saber o gênero do autor do trabalho, com comentários mais positivos. Um *feedback* mais gentil e menos verdadeiro pode privar quem o recebe de informação útil para melhorar. Somado ao *feedback* contraditório relatado pelas residentes ouvidas por Billick, Rassos e

Ginsburg (2022), esse resultado mostra que o risco não termina quando a nota é atribuída. Zonno (2026), numa revisão sistemática de 54 estudos sobre avaliação de desempenho em locais de trabalho hierárquicos, conclui que avaliações formalmente neutras reproduzem desigualdades e que o viés pode entrar na definição dos critérios, nos instrumentos e nas escalas e na interação entre avaliador e avaliado. A revisão excluiu os contextos educacionais, e a transposição para a escola é deste estudo.

O regime português prevê que o professor conheça a classificação e possa contestá-la. A avaliação final é comunicada por escrito, cabe reclamação em dez dias úteis e, da decisão sobre a reclamação, recurso a uma comissão de três árbitros docentes; aplicam-se aos intervenientes as regras de impedimento e suspeição do procedimento administrativo (Portugal, 2012a). Para contestar, porém, é preciso saber em que se baseou o juízo. Um registro feito de descrições pode ser discutido, porque o professor reconhece ou não a aula que ali está, enquanto um registro feito de adjetivos oferece pouco a que responder. O diagnóstico de opacidade que Guedes (2023) faz do sistema, ligado ao sigilo e à falta de relação entre os intervenientes, ganha aqui um sentido preciso.

Numa supervisão orientada para o desenvolvimento, o *feedback* é mais do que a comunicação de uma nota. Na tradição da supervisão clínica, divulgada em Portugal por autores como Alarcão e Tavares (2003), a observação insere-se num ciclo que inclui um encontro anterior, em que se combina o que será observado, e uma conversa posterior, em que professor e supervisor analisam juntos o que aconteceu. Nessa conversa, o professor pode descrever a aula do seu ponto de vista, e o supervisor pode rever o que registrou. Seabra et al. (2021b) encontraram, nos relatórios da avaliação externa das

escolas portuguesas, a valorização de uma supervisão de matriz colaborativa, considerada ainda não plenamente implementada. A proposição apresentada a seguir apoia-se nesse tipo de *feedback*, aberto e transparente, que mostra ao professor o que foi visto antes de lhe dizer quanto vale.

3.5. A Dupla Leitura do Juízo Avaliativo

Propõe-se que cada juízo do supervisor-avaliador sobre uma aula seja lido duas vezes. Na primeira leitura, ele é tratado como descrição do que aconteceu, isto é, o que o professor fez, o que os alunos fizeram e com que efeito observável, como o guião de observação já pede. Depois, o mesmo juízo é tratado como possível expressão de uma expectativa sobre quem deu a aula. O instrumento dessa segunda leitura é um teste de troca, em que o avaliador se pergunta se registraria a mesma coisa, com as mesmas palavras e o mesmo peso, se a aula tivesse sido dada por um colega de outro gênero, com a mesma turma e a mesma disciplina. Quando as duas leituras coincidem, o registro segue. Quando divergem, o juízo é levado à conversa com o professor avaliado, que pode descrever a aula do seu ponto de vista, antes de entrar na classificação.

O teste de troca é proposta deste estudo e não foi testado empiricamente. Na literatura consultada, não se localizou descrição acadêmica do procedimento como prática de avaliação ou de *feedback*. A ideia inspira-se em dois resultados experimentais obtidos fora da escola. Uhlmann e Cohen (2005), em três estudos sobre contratação, mostraram que os avaliadores redefiniam os critérios de mérito de modo a favorecer as credenciais que o candidato do gênero preferido tinha, e que o compromisso com

critérios fixados antes de se conhecer o gênero do candidato eliminava a discriminação. Bohnet, van Geen e Bazerman (2016) mostraram que, quando os candidatos são avaliados em conjunto e não um de cada vez, as decisões passam a basear-se mais no desempenho individual e menos em estereótipos de grupo. A primeira leitura retoma a ideia de fixar o que conta antes de julgar. O teste de troca, por sua vez, é uma comparação feita na imaginação, para quando não há um segundo caso à mão para pôr ao lado do primeiro.

Como ferramenta de reflexão, a dupla leitura não detecta viés. Duas leituras que coincidem não provam que o juízo é isento, porque o avaliador pode compartilhar a expectativa sem a perceber, e duas leituras que divergem não provam que houve viés, porque podem indicar apenas que o registro era vago. A divergência marca um juízo que merece ser discutido, e nada mais. Num experimento de campo com avaliações de docentes feitas por estudantes, Boring e Philippe (2021) verificaram que um aviso normativo genérico contra a discriminação não teve efeito, enquanto uma informação precisa sobre o viés observado em anos anteriores reduziu a discriminação, sobretudo nas avaliações que os estudantes homens faziam das docentes. O resultado vem de outro tipo de avaliação, mas sugere que voltar a um registro concreto com uma pergunta concreta pode fazer mais do que lembrar ao avaliador que não deve discriminar.

No encontro com o professor, a dupla leitura transforma uma suspeita do avaliador em pergunta compartilhada, feita antes da classificação, e dá ao avaliado a palavra que o regime formal só lhe dá depois, pela reclamação. Na formação dos avaliadores, pode ser exercitada entre pares, como prática de *peer coaching*, o acompanhamento mútuo entre colegas, ou em comunidades de

aprendizagem de avaliadores, em que dois ou mais colegas registram a mesma aula gravada, comparam os registros e aplicam o teste de troca às frases em que discordam. Num processo de *mentoring*, o exercício mostra a avaliadores menos experientes como um registro descritivo se distingue de uma impressão. A baixa concordância entre observadores encontrada por Ferreira (2016) e a redução da formação em avaliação externa relatada por Marques (2023) indicam que esse espaço de formação faz falta independentemente do gênero, e a dupla leitura dá-lhe um conteúdo específico.

O teste de troca simplifica o gênero ao imaginar a passagem entre duas posições. Para Scott, o gênero é uma relação e não um par fixo de categorias, para Butler (2003) ele não se reduz a duas identidades estáveis, e há professores que não se reconhecem em nenhuma delas. O teste é, por isso, um recurso heurístico. Pode estender-se a outras categorias, como a raça, que Campbell (2023) mostra combinar-se com o gênero no juízo dos avaliadores. Não corrige os efeitos da composição da turma, que pedem outro tipo de cuidado, nem dissolve a tensão entre apoio e controle, já que num regime com percentis e efeitos na carreira a conversa formativa continua a acontecer à sombra da nota. Há ainda um limite institucional. Monteiro (2024) analisa resistências institucionais e antigênero à integração da perspectiva de gênero na educação em Portugal, e resistências desse tipo podem surgir também quando se nomeia o gênero na ADD. Butler (2024) examina o movimento que apresenta o próprio conceito de gênero como uma ideologia que ameaçaria a família, o que ajuda a entender por que a palavra, sozinha, pode fechar a conversa. Nesse terreno, a dupla leitura tem a vantagem de poder começar como pergunta sobre um registro, sem exigir que o avaliador adote de partida uma teoria sobre o gênero.

4. CONSIDERAÇÕES FINAIS

O estudo partiu da pergunta sobre como as expectativas de gênero podem influenciar o que o avaliador vê e diz na observação de aulas e no *feedback*, e sobre o que a supervisão pode fazer para que o professor seja julgado pelo que fez na aula. Na ADD portuguesa, a observação reúne na mesma pessoa e na mesma visita o apoio da supervisão e o controle da classificação, e o juízo individual pesa muito: a nota tem efeitos na carreira, os percentis transformam as menções mais altas numa comparação entre colegas, os descritores são amplos e há indícios portugueses de baixa concordância entre observadores e de formação curta para quem avalia.

Nesse espaço, a autoridade pode parecer firmeza num caso e excesso noutro, o cuidado pode passar despercebido quando é esperado e ser notado quando surpreende, e o domínio da matéria pode inspirar mais ou menos confiança conforme a área e quem a ensina. Como a evidência internacional mostra efeitos contra professoras e contra professores, não se pode presumir quem é prejudicado. Essas leituras chegam ao registro e ao *feedback* sobretudo quando o registro se faz de adjetivos sobre a pessoa, e não de descrições da aula, e quando o professor não tem ocasião de discutir o que foi visto antes de receber a nota.

Para a supervisão, ficam duas tarefas. A primeira é exigir de si a descrição antes do juízo, que o próprio guião legal já pede. A segunda é tratar o juízo do avaliador como objeto de reflexão, pela dupla leitura proposta, levando à conversa com o professor os juízos em que a descrição e a expectativa se separam. A proposta não elimina o viés, mas pode torná-lo discutível quando o avaliador ainda decide o que escrever.

Os limites do estudo são os de uma revisão narrativa. Os resultados empíricos discutidos vêm, quase todos, de outros países e de outros avaliadores, e foram usados como ilustração. As extensões feitas, de Scott, Butler e Eagly e Karau para a avaliação de professores e da literatura sobre empresas, medicina e ensino superior para a ADD, são do estudo e foram assinaladas como tal. Nas fontes consultadas, não foram localizados dados públicos das menções da ADD desagregados por sexo.

Esses limites apontam para uma agenda de investigação: comparar, com autorização e salvaguardas, a linguagem dos registros escritos de observação de professoras e professores do mesmo grupo de recrutamento; conhecer a composição por sexo dos pares de avaliador e avaliado em cada nível de ensino; e testar a dupla leitura em ações de formação de avaliadores externos, verificando se ela altera os registros e a conversa pós-observação. A revisão do Estatuto da Carreira Docente, que em 2026 ainda tinha o modelo de avaliação entre os temas por negociar, abre a oportunidade de discutir o que o registro da observação deve conter, por exemplo se o guião deve continuar indicativo ou passar a exigir descrições que o professor possa reconhecer e contestar.

REFERÊNCIAS BIBLIOGRÁFICAS

ALARCÃO, Isabel; TAVARES, José. **Supervisão da prática pedagógica:** uma perspectiva de desenvolvimento e aprendizagem. 2. ed. Coimbra: Almedina, 2003.

ALVES, Maria Piedade; CUNHA, Pedro Fernando; LOURENÇO, Abílio Afonso; MONTEIRO, Ana Paula. Perceção dos professores sobre a avaliação do desempenho docente. **Revista Portuguesa de**

Educação, [S. l.], v. 31, n. 1, p. 61-78, 2018. DOI: <https://doi.org/10.21814/rpe.14082>. Disponível em: <https://revistas.rcaap.pt/rpe/article/view/14082>. Acesso em: 9 out. 2026.

BEG, Sabrin; FITZPATRICK, Anne; LUCAS, Adrienne M. Gender bias in assessments of teacher performance. **AEA Papers and Proceedings**, [S. l.], v. 111, p. 190-195, 2021. DOI: <https://doi.org/10.1257/pandp.20211126>. Disponível em: https://lerner.udel.edu/wp-content/uploads/beg_fitzpatrick_lucas_2021_gender_bias.pdf. Acesso em: 9 out. 2026.

BILLICK, Maxime; RASSOS, James; GINSBURG, Shiphra. Dressing the part: gender differences in residents' experiences of feedback in internal medicine. **Academic Medicine**, [S. l.], v. 97, n. 3, p. 406-413, 2022. DOI: <https://doi.org/10.1097/ACM.0000000000004487>.

BINDERKRANTZ, Anne Skorkjær; BISGAARD, Mette. A gender affinity effect: the role of gender in teaching evaluations at a Danish university. **Higher Education**, [S. l.], v. 87, n. 3, p. 591-610, 2024. DOI: <https://doi.org/10.1007/s10734-023-01025-9>. Disponível em: <https://link.springer.com/article/10.1007/s10734-023-01025-9>. Acesso em: 9 out. 2026.

BOHNET, Iris; VAN GEEN, Alexandra; BAZERMAN, Max. When performance trumps gender bias: joint vs. separate evaluation. **Management Science**, [S. l.], v. 62, n. 5, p. 1225-1234, 2016. DOI: <https://doi.org/10.1287/mnsc.2015.2186>.

BORING, Anne; PHILIPPE, Arnaud. Reducing discrimination in the field: evidence from an awareness raising intervention targeting

gender biases in student evaluations of teaching. **Journal of Public Economics**, [S. /], v. 193, 104323, 2021. DOI: <https://doi.org/10.1016/j.jpubeco.2020.104323>.

BUTLER, Judith. **Problemas de gênero: feminismo e subversão da identidade**. Tradução de Renato Aguiar. Rio de Janeiro: Civilização Brasileira, 2003.

BUTLER, Judith. **Quem tem medo do gênero?** Tradução de Heci Regina Candiani. São Paulo: Boitempo, 2024.

CAMPBELL, Shanyce L. Ratings in black and white: a quantcrit examination of race and gender in teacher evaluation reform. **Race Ethnicity and Education**, [S. /], v. 26, n. 7, p. 815-833, 2023. DOI: <https://doi.org/10.1080/13613324.2020.1842345>.

CAMPBELL, Shanyce L.; RONFELDT, Matthew. Observational evaluation of teachers: measuring more than we bargained for? **American Educational Research Journal**, [S. /], v. 55, n. 6, p. 1233-1267, 2018. DOI: <https://doi.org/10.3102/0002831218776216>.

CARREIRA Docente volta à discussão com manifestação à vista. **Observador**, [S. /], 24 mar. 2026. Disponível em: <https://observador.pt/especiais/professores-negociacoes-do-estatuto-da-carreira-docente-continuam-a-decorrer-em-clima-de-tensao-e-ja-com-protesto-a-vista/>. Acesso em: 9 out. 2026.

CAVALHEIRO, Maria Gabriela Caeiro Fernandes. **A importância da formação dos avaliadores externos no processo de avaliação do desempenho docente: um estudo de caso nas escolas associadas do Centro de Formação de Associação de Escolas Lezíria do Tejo**. 2024. Trabalho de projeto (Mestrado em Educação) – Escola Superior

de Educação de Fafe, Fafe, 2024. Disponível em: <https://comum.rcaap.pt/bitstreams/98cd468f-a481-4f85-bda2-c9cde33717f6/download>. Acesso em: 9 out. 2026.

CHI, Olivia L. A classroom observer like me: the effects of race-congruence and gender-congruence between teachers and raters on observation scores. **Education Finance and Policy**, [S. l.], v. 18, n. 3, p. 442-466, 2023. DOI: https://doi.org/10.1162/edfp_a_00367.

CORRELL, Shelley J.; WEISSHAAR, Katherine R.; WYNN, Alison T.; WEHNER, JoAnne Delfino. Inside the black box of organizational life: the gendered language of performance assessment. **American Sociological Review**, [S. l.], v. 85, n. 6, p. 1022-1050, 2020. DOI: <https://doi.org/10.1177/0003122420962080>.

DELGADO, William; SARTAIN, Lauren. Classroom composition affects teacher performance ratings. **Educational Evaluation and Policy Analysis**, [S. l.], 2026. DOI: <https://doi.org/10.3102/01623737261449250>.

DIREÇÃO-GERAL DE ESTATÍSTICAS DA EDUCAÇÃO E CIÊNCIA (Portugal). **Perfil do Docente 2023/2024**. [S. l.]: DGEEC, [s. d.]. Disponível em: <https://www.dgeec.medu.pt/api/ficheiros/68cd30d69864c806146a83d3>. Acesso em: 9 out. 2026.

EAGLY, Alice H.; KARAU, Steven J. Role congruity theory of prejudice toward female leaders. **Psychological Review**, [S. l.], v. 109, n. 3, p. 573-598, 2002. DOI: <https://doi.org/10.1037/0033-295X.109.3.573>.

EVANS, C. A.; ADLER, K.; YUCALAN, D.; SCHNEIDER-BENTLEY, L. M. Gender patterns in engineering PhD teaching assistant evaluations corroborate role congruity theory. **International Journal of STEM**

Education, [S. l.], v. 11, n. 1, 5, 2024. DOI: <https://doi.org/10.1186/s40594-023-00460-5>.

FELKEY, Amanda J.; BATZ-BARBARICH, Cassondra. Can women teach math (and be promoted)? A meta-analysis of gender differences across student evaluations of teaching. **AEA Papers and Proceedings**, [S. l.], v. 111, p. 184-189, 2021. DOI: <https://doi.org/10.1257/pandp.20211125>.

FERREIRA, Rui. **Avaliação do desempenho docente em Educação Física**: análise da variância dos resultados obtidos através do sistema de observação de aulas. 2016. Tese (Doutorado) – Universidade de Trás-os-Montes e Alto Douro, Vila Real, 2016. Disponível em: <http://hdl.handle.net/10348/7203>. Acesso em: 9 out. 2026.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GRISSOM, Jason A.; BARTANEN, Brendan. Potential race and gender biases in high-stakes teacher observations. **Journal of Policy Analysis and Management**, [S. l.], v. 41, n. 1, p. 131-161, 2022. DOI: <https://doi.org/10.1002/pam.22352>.

GUEDES, Teresa Cristina Melo e. **A avaliação do desempenho docente em Portugal**: contradições e paradoxos. 2023. Tese (Doutoramento em Educação) – Universidade Lusófona, Lisboa, 2023. Disponível em: <http://hdl.handle.net/10437/14518>. Acesso em: 9 out. 2026.

HEATH, Janae K.; KOGAN, Jennifer R.; HOLMBOE, Eric S.; CONFORTI, Lisa; PARK, Yoon Soo; DINE, C. Jessica. Gender differences in work-based assessment scores and narrative comments after direct

observation. **Journal of General Internal Medicine**, [S. l.], v. 39, n. 10, p. 1795-1802, 2024. DOI: <https://doi.org/10.1007/s11606-024-08645-6>. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11282012>. Acesso em: 9 out. 2026.

JAMPOL, Lily; ZAYAS, Vivian. Gendered white lies: women are given inflated performance feedback compared with men. **Personality and Social Psychology Bulletin**, [S. l.], v. 47, n. 1, p. 57-69, 2021. DOI: <https://doi.org/10.1177/0146167220916622>.

JÓHANNESSEN, Ingólfur Ásgeir; OTTESEN, Andri Rafn; BJARNADÓTTIR, Valgerður S. Natural disciplinarians or learning from the job? The first two years of seven male teachers in Icelandic compulsory schools. **Education Inquiry**, [S. l.], v. 15, n. 2, p. 188-202, 2024. DOI: <https://doi.org/10.1080/20004508.2022.2080343>. Disponível em: <https://opinvisindi.is/bitstreams/bb344648-fbb4-430f-911c-aebe3421be7d/download>. Acesso em: 9 out. 2026.

KHOKHLOVA, Olga; LAMBA, Nishtha; KISHORE, Sameer. Evaluating student evaluations: evidence of gender bias against women in higher education based on perceived learning and instructor personality. **Frontiers in Education**, [S. l.], v. 8, 1158132, 2023. DOI: <https://doi.org/10.3389/feduc.2023.1158132>.

LAPAGLIA, Jessica; MILLER, Katelyn; PROTEXTER, Samantha. Gender bias interacts with instructor disfluency to negatively affect student evaluations of teaching. **Frontiers in Education**, [S. l.], v. 7, 817291, 2022. DOI: <https://doi.org/10.3389/feduc.2022.817291>.

MACHADO, Eusébio André; ABELHA, Marta; BARREIRA, Carlos; SALGUEIRO, Ana. Avaliação pelos pares: percurso normativo da

avaliação do desempenho docente em Portugal. **Revista Portuguesa de Pedagogia**, [S. l.], ano 46, n. 1, p. 73-93, 2012.

MARQUES, Eliane Moreira. **Contributos para uma compreensão da formação contínua em avaliação externa do desempenho docente**: um estudo de caso. 2023. Dissertação (Mestrado) – Universidade Aberta, Lisboa, 2023. Disponível em: <http://hdl.handle.net/10400.2/13829>. Acesso em: 9 out. 2026.

MENGEL, Friederike; SAUERMAN, Jan; ZÖLITZ, Ulf. Gender bias in teaching evaluations. **Journal of the European Economic Association**, [S. l.], v. 17, n. 2, p. 535-566, 2019. DOI: <https://doi.org/10.1093/jeea/jvx057>.

MONTEIRO, Rosa Filomena Brás Lopes. Resistências institucionais e antigénero à igualdade de género na educação em Portugal. **Cadernos de Pesquisa**, [S. l.], v. 54, e11410, 2024. DOI: <https://doi.org/10.1590/1980531411410>.

OCDE. **Resultados do TALIS 2018**: Portugal: country note. [S. l.]: OCDE, 2019. Disponível em: https://www.oecd.org/content/dam/oecd/en/about/programmes/edu/talis/talis2018participantnotes/TALIS2018_CN_PRT_prt.pdf. Acesso em: 9 out. 2026.

OCDE. **TALIS 2024 results**: participant note: Portugal. [S. l.]: OECD, 2025. Disponível em: <https://www.dgeec.medu.pt/api/ficheiros/68e5295eceacc0faf4a3991d>. Acesso em: 9 out. 2026.

PARKER, Michele A.; SAVAGE, Shawn S. A troubleshoot test of student evaluations of teaching: role congruity, gendered language,

and educational (in)equalities. **Education Sciences**, [S. /], v. 16, n. 3, 448, 2026. DOI: <https://doi.org/10.3390/educsci16030448>. Disponível em: <https://www.mdpi.com/2227-7102/16/3/448>. Acesso em: 9 out. 2026.

PORTUGAL. Decreto Regulamentar n.º 26/2012, de 21 de fevereiro. Regulamenta o sistema de avaliação do desempenho do pessoal docente estabelecido no Estatuto da Carreira dos Educadores de Infância e dos Professores dos Ensinos Básico e Secundário. **Diário da República**: 1.ª série, [S. /], n. 37, p. 855-861, 21 fev. 2012a. Disponível em:

https://www.rbe.mec.pt/hp4/file/289/decreto_regulamentar_26_2012.pdf. Acesso em: 9 out. 2026.

PORTUGAL. Gabinete do Secretário de Estado do Ensino e da Administração Escolar. Despacho n.º 13981/2012, de 26 de outubro. Estabelece os parâmetros nacionais para a avaliação externa da dimensão científica e pedagógica. **Diário da República**: 2.ª série, [S. /], n. 208, p. 35375-35376, 26 out. 2012b. Disponível em: https://www.dgae.medu.pt/download/gestrehumanos/pessoal_docente/carreira/avaliacao_de_desempenho/docentes/regulamentacao/20121026_Grh_AvDesempenhoDocentesRegul_Despacho_13981.pdf.

Acesso em: 9 out. 2026.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico**: métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Novo Hamburgo: Feevale, 2013.

SCOTT, Joan W. Gender: a useful category of historical analysis. **The American Historical Review**, [S. /], v. 91, n. 5, p. 1053-1075, 1986. DOI:

<https://doi.org/10.1086/ahr/91.5.1053>.

SEABRA, Filipa; ABELHA, Marta; MIRANDA, Helena; MOREIRA, Osvaldo. Supervisão pedagógica: perspectivas de educadores de infância em Portugal. **Práxis Educativa**, [S. l.], v. 16, p. 1-21, 2021a. DOI: <https://doi.org/10.5212/praxeduc.v.16.17961.068>.

SEABRA, Filipa; MOURAZ, Ana; HENRIQUES, Susana; ABELHA, Marta. A supervisão pedagógica na política e na prática educativa: o olhar da Avaliação Externa das Escolas em Portugal. **Education Policy Analysis Archives**, [S. l.], v. 29, 2021b. DOI: <https://doi.org/10.14507/epaa.29.6486>.

SILVA, Márcia Alves da; LEAL DA COSTA, Conceição. Feminização da docência: diálogos entre experiências de formação inicial de professoras no Brasil e Portugal. **Reflexão e Ação**, [S. l.], v. 33, 2025. DOI: <https://doi.org/10.17058/0n25kv23>.

STOESZ, Brenda M.; DE JAEGER, Amy E.; QUESNEL, Matthew; BHOJWANI, Dimple; LOS, Ryan. Bias in student ratings of instruction: a systematic review of research from 2012 to 2021. **Canadian Journal of Educational Administration and Policy**, [S. l.], n. 201, p. 39-62, 2023. DOI: <https://doi.org/10.7202/1095482ar>.

UHLMANN, Eric Luis; COHEN, Geoffrey L. Constructed criteria. **Psychological Science**, [S. l.], v. 16, n. 6, p. 474-480, 2005. DOI: <https://doi.org/10.1111/j.0956-7976.2005.01559.x>.

VIEIRA, Flávia; MOREIRA, Maria Alfredo. **Supervisão e avaliação do desempenho docente**: para uma abordagem de orientação transformadora. Lisboa: Ministério da Educação; Conselho Científico para a Avaliação de Professores, 2011. (Cadernos do CCAP, 1).

WIND, Stefanie A.; JONES, Eli; BERGIN, Christi. Principals' severity affects teacher evaluation: statistical adjustments mitigate effects. **School Effectiveness and School Improvement**, [S. l.], v. 32, n. 3, p. 413-429, 2021. DOI: <https://doi.org/10.1080/09243453.2021.1892773>.

ZONNO, Roberta. Gender inequalities in performance evaluation: a systematic literature review. **Equality, Diversity and Inclusion: An International Journal**, [S. l.], v. 45, n. 9, p. 68-85, 2026. DOI: <https://doi.org/10.1108/edi-04-2025-0289>.

¹ Percurso acadêmico de nível avançado em Educação (Doutoramento e Mestrado pela Universidade Lusófona), sustentado por formação complementar em Pedagogia, Fisioterapia e gestão do ensino superior. Atualmente concentra-se na condução de processos de investigação científica, orientando estudantes de mestrado e especialização desde a concepção do problema de estudo até à defesa final, com foco na consolidação de competências metodológicas e no rigor da escrita académica. No plano institucional, assume responsabilidades de coordenação de programas educativos internacionais, o que inclui ajustar o desenho curricular de disciplinas, capacitar equipas docentes na implementação de metodologias pedagógicas e monitorizar indicadores de aprendizagem para orientar decisões de melhoria contínua. Esta experiência é complementada pela docência direta, em que combina planeamento didático com práticas de avaliação voltadas ao desenvolvimento da autonomia intelectual dos estudantes. A trajetória profissional anterior em contexto clínico-hospitalar contribuiu para uma abordagem de trabalho marcada por decisões fundamentadas, resolução de problemas em cenários complexos e cooperação eficaz em equipas multidisciplinares —

características hoje transferidas para a gestão acadêmica e para o acompanhamento próximo de orientandos ao longo de todo o ciclo investigativo. E-mail: [acesse o artigo original para visualizar o e-mail](#)