

# Métodos de Avaliação das Aprendizagens em Universidades Públicas Portuguesas: Um Estudo Quantitativo

## Assessment Methods in Portuguese Public Universities: A Quantitative Study

Eva Lopes Fernandes \*  
Maria Assunção Flores  
Irene Cadime  
Clara Pereira Coutinho

Centro de Investigação em Estudos da Criança, Universidade do Minho, Portugal

Este artigo enquadra-se num projeto de investigação mais amplo que procura identificar os métodos de avaliação mais valorizados e mais utilizados por professores universitários no contexto português. Os resultados foram recolhidos através de um inquérito por questionário com professores de cinco universidades públicas portuguesas (n=185), de todas as categorias profissionais e ciclos de estudos, das seguintes áreas científicas: Ciências Exatas, Ciências da Engenharia e da Tecnologia, Ciências Médicas e da Saúde, Ciências Sociais e Humanidades. A análise fatorial exploratória indicou uma organização das escalas sobre a valorização e frequência de utilização de métodos de avaliação em três fatores: métodos coletivos, portefólios e reflexões, e métodos individuais, tendo-se procedido ao tratamento dos testes e exames escritos como variável observável. Os resultados das estatísticas descritivas identificam uma maior valorização e frequência de utilização de testes e exames escritos e menor valorização e frequência de utilização de portefólios e reflexões. Os dados sugerem ainda diferenças estatisticamente significativas na utilização de métodos de avaliação em função dos ciclos de estudos, área de conhecimento e formação pedagógica.

**Palavras-Chave:** Ensino superior; Métodos de avaliação; Avaliação das aprendizagens; Professores universitários.

This article is part of a broader research project aimed at identifying the most valued and most used assessment methods by faculty in the Portuguese context. The results were collected through a survey with faculty from five Portuguese public universities (n=185) in all professional categories and cycles of study, of the following scientific areas: Hard Sciences, Engineering and Technology Sciences, Medical and Health Sciences, Social Sciences and Humanities. The exploratory factor analysis (EFA) results indicated a three factors model for both scales: collective methods, portfolios and reflections, and individual methods. The item “written tests and exams” was treated in subsequent analysis independently as observable variable. The results of descriptive statistics identify a greater valorisation and use of written tests and exams and a lower valorisation and use of portfolios and reflections. Data also suggest statistically significant differences in the use of assessment methods according to the cycles of study, the area of knowledge and pedagogical training.

**Keywords:** Higher education; Assessment methods; Assessment of learning; Faculty.

---

\*Contacto: evalopesfernandes@ie.uminho.pt

## 1. Introdução

A avaliação no contexto do Ensino Superior constitui uma temática que tem recebido uma atenção crescente por parte dos decisores políticos, dos investigadores, dos gestores, dos professores e outros *stakeholders* na medida em que influencia a forma como os estudantes organizam o seu tempo e mobilizam os seus esforços (Fernandes, 2015; Myers e Myers, 2015) e as suas perceções sobre o modo como as aprendizagens se desenvolvem (Brown e Knight, 1994). Daí a importância da seleção dos métodos de avaliação mais adequados em função dos objetivos de ensino e aprendizagem (Pereira e Flores, 2016).

Estudos realizados no ensino superior demonstram que os instrumentos de avaliação tradicionais são os mais utilizados (principalmente o exame escrito) (Barreira et al., 2015; Pereira, 2016; Pereira e Flores, 2016), destacando-se o sistema de classificação e a hierarquização (Pereira e Flores, 2013; Perrenoud, 1999). Outras perspetivas de avaliação, como, por exemplo, o *assessment for learning* (McDowell et al., 2011) apontam para um papel de cada vez maior responsabilização dos estudantes no processo de avaliação e de aprendizagem. Este alargamento de papéis aos diversos atores é extensível aos professores, que necessitam de assumir o processo de ensino e aprendizagem de forma mais autónoma, colaborativa e integrada, através de projetos partilhados; da produção coletiva do conhecimento; no sentido de tornar o processo de aprendizagem mais criativo (Flores e Veiga Simão, 2007), compreendendo a avaliação como uma parte integrante da aprendizagem (Zabalza, 2008).

Neste artigo são apresentados dados decorrentes de um projeto mais vasto sobre avaliação no Ensino Superior na perspetiva de docentes universitários, nomeadamente no que diz respeito à valorização e utilização de métodos de avaliação.

A literatura existente distingue métodos tradicionais de métodos alternativos de avaliação (Duncan e Buskirk-Cohen, 2011), destacando-se práticas centradas no professor e centradas no aluno (Burkšaitienė e Teresevičienė, 2008; Fernandes, 2015; Myers e Myers, 2015), e reconhecendo-se, no âmbito destas últimas, o espaço privilegiado da sala de aula na organização de ambientes inovadores e facilitadores da aprendizagem (Black e William, 1998) e no desenvolvimento de formas inovadoras de estruturar o ensino e a avaliação (Fernandes, 2015). A transformação das práticas pedagógicas encerra uma mudança no papel das instituições do ensino superior enquanto contextos que existem para produzir conhecimento (Burkšaitienė e Teresevičienė, 2008). Esta visão assume especial importância no contexto do Processo de Bolonha (Flores e Veiga Simão, 2007), desafiando os professores a promoverem oportunidades de aprendizagem mais questionadoras, inovadoras e criativas (Fernandes e Flores, 2012; Zabalza, 2008).

Os métodos tradicionais, comumente usados no contexto do Ensino Superior (Duncan e Buskirk-Cohen, 2011; Pereira e Flores, 2016), sobretudo no contexto de turmas grandes, a par da sobrecarga de trabalho dos professores (Myers e Myers, 2015), podem enfatizar a reprodução e memorização (Perrenoud, 1999) e baixos níveis de compreensão (Dochy, et al., 2007). Os métodos tradicionais (nomeadamente o teste ou exame escrito) ocorrem num momento pré-determinado, centram-se no produto ou resultados e apresentam usualmente uma forte componente individual (Hadji, 1994), representando um meio de perceber o que se passa na sala de aula (Phelps, 2017).

Por seu turno, os chamados métodos de avaliação centrados no aluno (Myers e Myers, 2015; Webber e Tschepikow, 2013) permitem o desenvolvimento de competências

técnicas e transversais, por exemplo, de resolução de problemas, fomentando um maior envolvimento dos alunos no processo de aprendizagem (Myers e Myers, 2015). Normalmente, estes métodos implicam tarefas mais globais, desenvolvidas ao longo do tempo e que podem incluir, simultaneamente, o processo e o produto e a dimensão individual e coletiva, estimulando a autonomia, a colaboração, responsabilidade, o *feedback* construtivo, a interação com os pares e a construção de conhecimento (Pereira e Flores, 2013; Webber e Tschepikow, 2013), o desenvolvimento de competências e o aprofundamento das aprendizagens (Brew, Riley e Walta, 2009). Estes métodos incluem, por exemplo, os trabalhos práticos experimentais, o trabalho de projeto ou as reflexões (Struyven, Dochy e Steven, 2005; Webber e Tschepikow, 2013). O portefólio destaca-se, pela sua especificidade, ao apresentar um carácter interativo, dinâmico, aberto e global. Trata-se de um método peculiar que encerra diferentes funções e tarefas, e que pode ser individual ou coletivo. Os métodos alternativos são usados ao longo do tempo, potenciando a revisão e análise dos elementos de avaliação produzidos e reconhecendo o papel essencial do *feedback* no processo de aprendizagem do aluno e na sua autorregulação (Carless et al., 2011). Apesar dos seus reconhecidos benefícios, são-lhes, igualmente, apontadas algumas limitações ao nível da prestação de contas e da dificuldade em usar a informação produzida do ponto de vista sumativo (Maclellan, 2004), o que, atendendo aos inúmeros desafios no campo da avaliação, pode justificar o seu uso limitado no contexto do ensino superior.

Estudos recentes apontam para a necessidade de mais investigação neste domínio, nomeadamente no que se refere aos métodos de avaliação mais utilizados e ao seu impacto na aprendizagem dos alunos (Watering et al., 2008), à comparação entre práticas de avaliação em diferentes áreas, instituições e países (Gilles, Detroz e Blais, 2010), à eficácia dos chamados métodos alternativos de avaliação ou centrados nos estudantes (Segers, Gijbels e Thurlings, 2008) e à necessidade de analisar as práticas de avaliação em articulação com mecanismos de *feedback* (Flores et al., 2014).

No sentido de responder a essa necessidade, este estudo pretende conhecer as conceções e práticas de avaliação de docentes do Ensino Superior após a implementação do Processo de Bolonha, mais concretamente através da identificação e análise dos métodos de avaliação mais valorizados e utilizados. Para o efeito, procedeu-se à validação de um instrumento de recolha de dados sobre valorização e utilização de métodos de avaliação adaptado de Pereira (2011, 2016) e Gonçalves (2016), que foi aplicado em cinco universidades públicas portuguesas. O instrumento foi concebido com o objetivo de alargar o espectro do estudo original de Pereira (2016) às perspetivas e práticas de professores universitários no sentido de dar respostas às seguintes questões de investigação:

- 1) Que importância atribuem aos métodos de avaliação?
- 2) Qual a frequência da utilização dos diferentes métodos de avaliação?
- 3) Quais as variáveis associadas à valorização e utilização de diferentes métodos de avaliação?
- 4) Qual a relação entre a valorização e a utilização de diferentes métodos de avaliação?

## 2. Métodos

Este artigo enquadra-se num projeto de investigação mais amplo realizado no âmbito do Doutoramento em Ciências da Educação (Ref. SFRH/BD/103291/2014) que, por sua vez, se integra um projeto de investigação mais alargado, intitulado “Assessment in Higher Education: The potential of alternative methods” (Ref. PTDC/MHC-CED/2703/2014), projetos financiados pela Fundação Nacional para a Ciência e Tecnologia (FCT). Este estudo teve como objetivos:

- 1) Analisar a influência das variáveis demográficas na valorização e utilização de métodos de avaliação;
- 2) Analisar a influência das variáveis profissionais na valorização e utilização de métodos de avaliação;
- 3) Analisar a influência das variáveis relacionadas com as práticas de avaliação e com o Processo de Bolonha na valorização e utilização de métodos de avaliação;
- 4) Testar a existência de uma correlação entre a valorização e a utilização de métodos de avaliação.

### 1.1. Participantes

Os resultados foram recolhidos através de um inquérito por questionário, aplicado entre fevereiro e julho de 2017, em cinco universidades públicas portuguesas. A amostra é composta por 185 professores universitários de todas as categorias profissionais e ciclos de estudos. Tendo em conta os objetivos de investigação optou-se por uma amostra de conveniência (Coutinho, 2014), definida de acordo com as áreas científicas identificadas pela FCT, assegurando que as cinco universidades envolvidas tivessem uma oferta educacional semelhante nas cinco áreas selecionadas.

No quadro 1 apresenta-se uma descrição detalhada das características dos participantes: são principalmente do sexo feminino, têm mais de 45 anos, são professores associados ou auxiliares, com mais de 15 anos de experiência e com formação pedagógica. Lecionam nas seguintes áreas de conhecimento: Ciências Médicas e da Saúde, Ciências Exatas, Ciências da Engenharia e Tecnologia, Ciências Sociais e Humanidades e em diferentes cursos: 83,8% lecionam nos cursos de graduação; 77,3% em cursos de mestrado; 41,5% em cursos de Mestrado Integrado; 55,8% em cursos de doutoramento; e, 1,7% em outros cursos (por exemplo, cursos não conferentes de grau).

### 1.2. Instrumento de recolha de dados

O questionário utilizado neste estudo é uma versão adaptada dos “Métodos de Avaliação no Ensino Superior”, utilizado no estudo de Pereira (2011, 2016) e da secção sobre utilização e importância dos métodos de avaliação do “Inventário de Conceções de Avaliação” (ICA) de Gonçalves (2016), centrando-se em dois aspetos fundamentais: 1) importância atribuída aos diferentes métodos de avaliação; e, 2) métodos os métodos de avaliação mais utilizados pelos docentes universitários. Ambas as escalas são constituídas pelos mesmos 14 itens (consultar quadro 2). Para cada item da primeira escala, os participantes classificaram a importância atribuída ao método utilizando uma escala *likert* de quatro pontos que varia entre 1 (nada importante) e 4 (muito importante). Era ainda oferecida a possibilidade de os participantes selecionarem a opção “Não se aplica”. Estas respostas foram codificadas com o valor mais baixo da escala de resposta (1). Na segunda

escala, os participantes indicaram a frequência da utilização de cada método, recorrendo a uma escala de *likert* de quatro pontos entre 1 (nada utilizados) e 4 (muito utilizados)

Quadro 1. Caracterização dos participantes

<b>CARACTERÍSTICAS DEMOGRÁFICAS</b>		
<b>Universidade</b>	<b>N</b>	<b>%</b>
A	36	19,5
B	34	18,4
C	60	32,4
D	36	19,5
E	19	10,2
<b>Sexo</b>		
Masculino	74	40,0
Feminino	87	47,0
Sem informação	24	13,0
<b>Idade</b>		
Menos de 45 anos de idade	82	44,3
Mais de 45 anos de idade	103	55,7
<b>Área de Conhecimento</b>		
Ciências Médicas e da Saúde	21	11,4
Ciências Exatas	16	8,6
Ciências da Engenharia e Tecnologia	50	27,0
Ciências Sociais	77	41,6
Humanidades	21	11,4
<b>Categoria Profissional</b>		
Professor Catedrático	10	5,4
Professor Associado/Auxiliar com Agregação	19	10,3
Professor Associado/Auxiliar	132	71,3
Outros	24	13,0
<b>Experiência Docente</b>		
Menos de 15 anos	54	29,2
Mais de 15 anos	131	70,8
<b>Formação Pedagógica</b>		
Sim	117	63,2
Não	63	34,1
Sem informação	5	2,7

Fonte: Elaboração própria.

### 1.3. Procedimentos

No contacto com as instituições e ao longo de toda a investigação foram respeitadas as questões éticas de investigação na área das Ciências Sociais e Humanas reconhecidas internacionalmente, tendo o estudo sido aprovado pela Comissão de Ética da Universidade do Minho (Ref. SECSH035/2016 e SECSH036/2016). Foram contactados todos os autores dos instrumentos de recolha de dados, tendo-se obtido autorização para o seu uso e adaptação ao contexto português. Foram celebrados protocolos de investigação com as faculdades e escolas/institutos envolvidos. Todos os participantes assinaram um consentimento informado voluntário após uma breve explicação do projeto e dos objetivos de investigação.

### 1.4. Análise estatística

As análises estatísticas foram realizadas com recurso ao software *IBM SPSS Statistics 24*. Por forma a explorar a estrutura interna do instrumento recorreu-se à análise fatorial exploratória (AFE), utilizando o método de análise de componentes principais (ACP) (Field, 2009). Os pressupostos para a realização desta análise foram aferidos utilizando o

KMO (*Kaiser-Meyer-Olkin*) para análise da adequação da amostra e o teste de esfericidade de Bartlett para testar se as intercorrelações entre os itens não se configuram como uma matriz de identidade (Field, 2009). No KMO foram considerados aceitáveis os valores acima de 0,5 e bons os valores acima de 0,7 (Hutcheson e Sofroniou, 1999). No teste de esfericidade de Bartlett, as estatísticas de teste com níveis de significância  $p < 0,05$  indicam que os dados são apropriados para a realização de análise em componentes principais (Tabachnick e Fidell, 1996). O método de extração foi a ACP com rotação *varimax*. A decisão sobre o número de fatores a reter foi realizada com base no critério de Kaiser, pelo que foram considerados todos os fatores com *eigenvalues* superiores a 1. Para atribuição dos itens a cada fator, consideram-se apenas cargas fatoriais  $> 0,35$ .

A fidelidade das subescalas foi testada com recurso ao método de avaliação da consistência interna dos itens, calculando-se o Alfa de Cronbach. Em geral, consideram-se como valores aceitáveis os valores de alfa superiores a 0,70, contudo, tratando-se de um estudo exploratório, esse valor pode diminuir para um mínimo de 0,60 (Hair, Black, Babin e Anderson, 2009).

Depois de delimitada a estrutura interna das escalas, calcularam-se as pontuações totais de cada dimensão, somando os respetivos itens e dividindo este valor pelo número de itens que a integram. Começou-se por calcular a estatística descritiva para a amostra total, analisando-se ainda a distribuição destas variáveis. Seguidamente, explorou-se a relação entre as pontuações nas diferentes dimensões de valorização e utilização dos métodos de avaliação e as características demográficas e profissionais dos participantes. Relativamente às características demográficas, na variável idade, para efeito deste estudo, foram considerados dois grupos para análise: 1) professores com menos de 45 anos de idade; 2) professores com mais de 45 anos de idade. Esta alteração nas unidades de tempo relativamente ao questionário inicial permitiu obter unidades de análise mais adequadas e grupos com um número de participantes mais equilibrado (Cadime, Silva, Ribeiro e Viana, 2018). Na variável experiência profissional (variáveis profissionais) procedeu-se à agregação do tempo de serviço da seguinte forma: 1) professores com menos de 15 anos de experiência; 2) professores com mais de 15 anos de experiência. Procedeu-se igualmente à junção das categorias profissionais “professor associado com agregação” e “professor auxiliar com agregação”, e “professor associado” e “professor auxiliar”. Esta alteração nas unidades de tempo e categorias profissionais relativamente ao questionário inicial permitiu obter unidades de análise mais adequadas, tendo-se usado a data da implementação do Processo de Bolonha como marco para a definição destas unidades temporais. Para a realização destas análises, recorreu-se à análise de variância multivariada (*MANOVA*). Garantiram-se os pressupostos de independência das observações, normalidade univariada e homogeneidade das matrizes de variância-covariância (Field, 2009). Os valores do eta-quadrado parcial ( $\eta^2$ ) foram calculados como medida de tamanho do efeito, considerando as seguintes linhas orientadoras para a sua interpretação: efeito pequeno,  $\eta^2 > 0,1$ ; efeito médio,  $\eta^2 > 0,3$ ; efeito grande,  $\eta^2 > 0,5$  (Cohen, 1988).

Nos casos em que foram realizados vários testes estatísticos independentes em simultâneo procedeu-se à correção de *Bonferroni* (Field, 2009). Quando os pressupostos para a utilização de testes paramétricos não foram cumpridos, optou-se pela utilização de testes não paramétricos (Testes de *Mann-Whitney* e *Kruskal-Wallis*) (Field, 2009).

Por último, procedeu-se ao teste das correlações entre os métodos de avaliação mais valorizados e os mais utilizados, através do coeficiente de correlação de *Spearman* (Field, 2009). Este coeficiente varia entre -1 e 1, indicando a direção e força da correlação, sendo que uma maior proximidade destes extremos corresponde a uma maior associação entre as variáveis (Field, 2009), tendo-se considerado os valores de 0,1 a 0,3 e -0,1 a -0,3 fracos; os valores de 0,4 a 0,6 e -0,4 a -0,6 moderados; os valores de 0,7 a 0,9 e -0,7 a -0,9 fortes e os valores 1 e -1 perfeitos (Dancey e Reidy, 2007) na interpretação dos valores de *r*.

Em todas as análises foram considerados como patamares para aceitação e/ou rejeição de hipóteses nulas os valores de  $p < 0,05$  (Field, 2009).

### 3. Resultados

#### 3.1. Análise fatorial exploratória e fidelidade das subescalas

Relativamente à escala sobre valorização de métodos de avaliação, o valor de KMO permitiu comprovar a adequação da amostra (KMO=0,828) e o resultado do teste de Bartlett revelou que os dados se adequam à realização desta análise,  $\chi^2(91)=1209,61$ ,  $p < 0,001$ . De igual modo, na escala de métodos de avaliação mais utilizados, o valor de KMO foi também elevado, sugerindo a adequação da amostra (KMO=0,762), e o resultado do teste de Bartlett revelou que as intercorrelações entre os itens são suficientemente elevadas para a realização desta análise,  $\chi^2(91)=836,67$ ,  $p < 0,001$ .

O quadro 2 reporta os resultados da AFE para a valorização e utilização de métodos de avaliação pelos participantes no estudo.

Uma primeira extração com a escala sobre a valorização de métodos de avaliação revelou a presença de três fatores com *eigenvalues* > 1, explicando no total 64,24% da variância. Os itens que se agruparam no mesmo fator sugerem que o fator 1 representa os Métodos Coletivos e Individuais; que o fator 2 representa os Portefólios e Reflexões; e, que o fator 3 representa os Testes e Exames. Os valores do Alfa de Cronbach foram elevados nos fatores 1 ( $\alpha=0,880$ ) e 2 ( $\alpha=0,850$ ), mas pobres para o fator 3 ( $\alpha=0,448$ ), tendo-se optado pela exclusão dos itens que compunham este fator (1-Testes/exames escritos e 2-Testes/exames orais). Contudo, dada a relevância do item 1, Testes/Exames escritos -, ao nível das respostas dos participantes do estudo e também ao nível dos estudos nacionais e internacionais (e.g. Barreira et al., 2015; Flores et al., 2014; Myers e Myers, 2015; Pereira, Flores e Barros, 2017; Pereira e Flores, 2016), optou-se pelo seu tratamento enquanto variável observável.

Após a remoção dos dois itens referidos, procedeu-se a uma nova ACP, utilizando os restantes itens. Os resultados desta análise são apresentados no quadro 2. Obtiveram-se três fatores com *eigenvalues* superiores a 1, explicando 69,17% da variância: F<sub>1</sub>) Métodos Coletivos; F<sub>2</sub>) Portefólios e Reflexões; e F<sub>3</sub>) Métodos Individuais. Todos os itens apresentaram saturações fatoriais elevadas no respetivo fator (ver quadro 2) e valores de Alfa de Cronbach > 0,70. Os itens 9, 11 e 12 saturaram simultaneamente em dois fatores tendo-se optado pela sua permanência no fator em que apresentaram maior carga fatorial (Tabachnik, e Fidell, 1996), dado que o seu conteúdo estava de acordo com o fator.

Na frequência da utilização de métodos de avaliação, uma primeira extração revelou a presença de quatro fatores. Os quatro fatores extraídos na análise foram: 1) Métodos Coletivos; 2) Portefólios e Reflexões; 3) Métodos Individuais; e 4) Testes e Exames. A

estatística de fidelidade, calculada através do Alfa de Cronbach revelou a existência de valores aceitáveis para os fatores 1 ( $\alpha=0,822$ ), 2 ( $\alpha=0,772$ ) e 3 ( $\alpha=0,693$ ). Contudo, os valores do Alfa revelaram-se inaceitáveis para o fator 4 ( $\alpha=0,342$ ), tendo-se optado pela sua exclusão. À semelhança da escala anterior, optou-se igualmente pelo tratamento do item 1 - “Testes/Exames escritos” -, em análises subsequentes, enquanto variável observável devido à sua relevância. Procedeu-se a uma nova extração de fatores, após a remoção dos itens 1 e 2. Os resultados da segunda análise são apresentados no quadro 2, através de uma estrutura de 3 fatores que explica 61,15% da variância total. Todos os itens revelaram saturações muito elevadas nos respetivos fatores. Os itens 5, 8, 11 e 12 apresentaram saturações em mais que um fator, tendo-se optado pela sua inclusão no fator com maior carga fatorial (Tabachnik e Fidell, 1996), dado que o seu conteúdo era congruente com o fator.

Quadro 2. Resultados da AFE para a valorização de métodos de avaliação e a frequência da utilização de métodos de avaliação

	FATORES – VALORIZAÇÃO DE MÉTODOS DE AVALIAÇÃO			FATORES – FREQUÊNCIA DA UTILIZAÇÃO DE MÉTODOS DE AVALIAÇÃO		
	1 MÉTODOS PORTEFÓLIOS COLETIVOS E REFLEXÕES INDIVIDUAIS	2 MÉTODOS PORTEFÓLIOS INDIVIDUAIS	3 MÉTODOS INDIVIDUAIS	1 MÉTODOS PORTEFÓLIOS COLETIVOS E REFLEXÕES INDIVIDUAIS	2 MÉTODOS PORTEFÓLIOS INDIVIDUAIS	3 MÉTODOS INDIVIDUAIS
6 - Trabalhos práticos ou experimentais em grupo	,808			,848		
8 - Projeto realizado em grupo	,713			,622		
10 - Relatórios em grupo	,789			,790		
14 - Apresentações orais em grupo	,730			,733		
3 - Portefólios coletivos		,770			,800	
4 - Portefólios individuais		,840			,811	
11 - Reflexões escritas individuais		,726			,558	
12 - Reflexões escritas em grupo		,666			,635	
5 - Trabalhos práticos ou experimentais individuais			,740			,565
7 - Projeto realizado individualmente			,718			,745
9 - Relatórios individuais			,660			,660
13 - Apresentações orais individuais			,756			,733
Eigenvalues	6,064	1,170	1,067	4,549	1,472	1,317
% Variância	50,53%	9,75%	8,90%	37,91%	12,26%	10,97%
Alfa de Cronbach	,848	,850	,807	,822	,772	,693

Nota. Os itens 1) Testes/exames escritos e 2) Testes/exames orais foram removidos nas duas escalas.

Fonte: Elaboração própria.

### 3.2. Diferenças nos métodos de avaliação mais valorizados e utilizados em função de variáveis demográficas e profissionais

O quadro 3 apresenta os resultados da estatística descritiva relativos à valorização e utilização de métodos de avaliação. Os testes de normalidade revelaram-se estatisticamente significativos, sugerindo a não normalidade das distribuições. Por conseguinte, foi dada prioridade à análise dos valores de assimetria e curtose (ver quadro 3). Os valores de assimetria e curtose são muito próximos de zero para todos os fatores de ambas as escalas, sugerindo a não existência de violações substanciais da normalidade, o

que permitiu avançar com a análise MANOVA. Contudo, verificou-se que os valores de assimetria e curtose da variável observável “testes e exames escritos” eram elevados, o que impossibilitou a inclusão destas pontuações nas análises de variância multivariada, tendo-se optado pela realização de testes não paramétricos para a sua análise.

Os “Testes/Exames Escritos” surgem como os métodos mais valorizados pelos professores, enquanto os “Portefólios e Reflexões” são os menos valorizados (ver quadro 3). Os resultados das estatísticas descritivas revelam também uma tendência positiva na valorização quer dos métodos individuais quer dos métodos coletivos.

Quadro 3. Estatísticas descritivas das escalas métodos de avaliação mais valorizados e utilizados

	N	%	MÉDIA	DESVIO PADRÃO	ASSIMETRIA	CURTOSE	TESTE KS	
							KS	P
<b>Valorização de métodos de avaliação</b>								
Fator 1 - Métodos coletivos			2,60	0,881	-0,333	-0,875	,167	<,001
Fator 2 - Portefólios e Reflexões	163	88,1	2,05	0,901	0,393	-1,067	,139	<,001
Fator 3 - Métodos Individuais			2,67	0,868	-0,399	-0,763	,152	<,001
Testes/Exames Escritos			3,24	0,784	-0,917	0,565	,251	<,001
<b>Frequência da utilização de métodos de avaliação</b>								
Fator 1 - Métodos coletivos			2,58	0,857	-0,217	-0,777	0,108	<,001
Fator 2 - Portefólios e Reflexões	171	92,4	1,82	0,736	0,797	0,156	0,132	<,001
Fator 3 - Métodos Individuais			2,44	0,714	-0,071	-0,444	0,108	<,001
Testes/Exames Escritos			3,37	0,811	-1,106	0,425	0,332	<,001

**Nota.** KS = Kolmogorov-Smirnov.

Fonte: Elaboração própria.

### 3.2.1. Influência das variáveis demográficas na valorização e utilização de métodos de avaliação

Os testes multivariados permitiram verificar que não existe uma influência da idade na valorização (WILK'S  $\Delta=,969$ ,  $F(3, 159)=1,683$ ,  $p=,173$ ;  $\eta^2=,031$ ) e na utilização dos diferentes métodos de avaliação (WILK'S  $\Delta=,960$ ,  $F(3, 167)=2,339$ ,  $p=,075$ ,  $\eta^2=,040$ ), assim como do sexo ao nível da valorização (WILK'S  $\Delta=,965$ ,  $F(3, 136)=1,652$ ,  $p=,180$ ;  $\eta^2=,035$ ) e da utilização dos diferentes métodos de avaliação (WILK'S  $\Delta=,980$ ,  $F(3, 143)=,993$ ,  $p=,398$ ,  $\eta^2=,020$ ). Também não foram encontradas diferenças estatisticamente significativas entre homens e mulheres ( $U=3217$ ,  $p=,994$ ), nem entre grupos etários ( $U=4026$ ,  $p=,635$ ) na valorização de testes e exames escritos. De igual modo, não foram identificadas diferenças estatisticamente significativas entre homens e mulheres, ( $U=2861,50$ ,  $p=,173$ ) e entre grupos etários ( $U=4047$ ,  $p=,588$ ) na frequência da utilização de testes e exames escritos.

### 3.2.2. Influência das variáveis profissionais na valorização e utilização de métodos de avaliação

Os resultados da MANOVA revelaram não existir efeitos estatisticamente significativos da categoria profissional (WILK'S  $\Delta=,963$ ,  $F(9, 382)=,658$ ,  $p=,747$ ,  $\eta^2=,012$ ), dos anos de experiência (WILK'S  $\Delta=,977$ ,  $F(3, 159)=1,220$ ,  $p=,304$ ,  $\eta^2=,023$ ), e da frequência de formação pedagógica (WILK'S  $\Delta=,967$ ,  $F(3, 156)=1,756$ ,  $p=,158$ ,  $\eta^2=,033$ ) na valorização dos diferentes métodos de avaliação. De igual modo, os resultados da análise multivariada indicaram não existir influência da categoria profissional (WILK'S  $\Delta=,926$ ,

$F(9, 402)=1,430$ ,  $p=,173$ ,  $\eta^2=,025$ ), anos de experiência (WILK'S  $\Delta=,968$ ,  $F(3, 167)=1,854$ ,  $p=,139$ ,  $\eta^2=,032$ ), e da frequência de formação pedagógica (WILK'S  $\Delta=,954$ ,  $F(3, 163)=2,632$   $p=,052$ ,  $\eta^2=,046$ ), na utilização dos diferentes métodos de avaliação.

No que concerne à influência dos ciclos de estudos na valorização de métodos de avaliação, os testes multivariados revelaram não existir diferenças estatisticamente significativas ao nível dos cursos de Licenciatura (WILK'S  $\Delta=,967$ ,  $F(3, 159)=1,794$ ,  $p=,150$ ,  $\eta^2=,033$ ) e de Doutoramento, (WILK'S  $\Delta=,971$ ,  $F(3, 155)=1,569$ ,  $p=,199$ ,  $\eta^2=,029$ ). Não obstante, os resultados da MANOVA permitiram verificar a existência de diferenças estatisticamente significativas, embora com tamanho do efeito negligível, entre professores que lecionam nos cursos de Mestrado Integrado e os que não lecionam, WILK'S  $\Delta=,913$ ,  $F(3, 151)=4,786$ ,  $p=,003$ ,  $\eta^2=,087$ . Os testes univariados indicam que os professores que lecionam neste ciclo de estudos, em média, valorizam menos os portefólios e reflexões do que os professores que lecionam apenas noutros ciclos (ver quadro 4). Os testes multivariados sugeriram também diferenças significativas, embora com tamanho do efeito negligível, na valorização dos diferentes métodos de avaliação entre os professores que lecionam em cursos de Mestrado e aqueles que não lecionam, WILK'S  $\Delta=,922$ ,  $F(3, 159)=4,508$ ,  $p=,005$ ,  $\eta^2=,078$ . Os resultados dos testes univariados permitiram concluir que os professores que lecionam neste ciclo de estudos valorizam mais os métodos coletivos, individuais e portefólios e reflexões do que aqueles que não lecionam neste ciclo de estudos (ver quadro 4).

Relativamente à utilização dos diferentes métodos de avaliação de acordo com os diferentes ciclos de estudos, os testes multivariados revelaram não existir diferenças estatisticamente significativas na utilização dos diferentes métodos de avaliação pelos professores que lecionam nos cursos de Licenciatura e aqueles que não lecionam, WILK'S  $\Delta=,959$ ,  $F(3, 167)=2,366$ ,  $p=,073$ ,  $\eta^2=,041$ . Não obstante, os resultados da MANOVA permitiram verificar a existência de diferenças estatisticamente significativas, com tamanho do efeito pequeno, entre professores que lecionam nos cursos de Mestrado Integrado e os que não lecionam, WILK'S  $\Delta=,852$ ,  $F(3, 158)=9,174$ ,  $p<,001$ ,  $\eta^2=,148$ . Os professores que lecionam neste ciclo de estudos utilizam, em média, com maior frequência os métodos coletivos e com menor frequência os portefólios e reflexões (ver quadro 4). Os testes multivariados sugeriram também diferenças significativas na frequência da utilização dos diferentes métodos de avaliação entre os professores que lecionam no Mestrado e aqueles que não lecionam, WILK'S  $\Delta=,898$ ,  $F(3,167)=6,293$ ,  $p<,001$ ,  $\eta^2=,102$ , bem como entre aqueles que lecionam nos cursos de Doutoramento e os que não lecionam, WILK'S  $\Delta=,949$ ,  $F(3, 163)=2,910$ ,  $p=,036$ ,  $\eta^2=,051$ . Em ambos os casos, os resultados dos testes univariados permitiram concluir que os professores que lecionam nestes ciclos de estudos utilizam com maior frequência os métodos de avaliação coletivos, individuais e portefólios e reflexões do que aqueles que não lecionam nesses mesmos ciclos de estudo (ver quadro 4).

Quadro 4. Resultados dos testes univariados para a análise do efeito do Ciclo de Estudos na valorização e frequência da utilização de métodos de avaliação

	LICENCIATURA					MESTRADO INTEGRADO					MESTRADO					DOUTORAMENTO				
	Sim M <sub>e</sub> (DP)	Não M <sub>e</sub> (DP)	F (gl)	P	ηp <sup>2</sup>	Sim M <sub>e</sub> (DP)	Não M <sub>e</sub> (DP)	F (gl)	P	ηp <sup>2</sup>	Sim M <sub>e</sub> (DP)	Não M <sub>e</sub> (DP)	F (gl)	P	ηp <sup>2</sup>	Sim M <sub>e</sub> (DP)	Não M <sub>e</sub> (DP)	F (gl)	P	ηp <sup>2</sup>
<b>Valorização de métodos de avaliação</b>																				
Métodos Coletivos	2,66 (,87)	2,32 (,92)	3,38 (1,16)	,07	,02	2,72 (,77)	2,56 (,96)	1,29 (1,15)	,26	,008	2,72 (,84)	2,18 (,90)	10,63 (1,16)	,001	,06	2,74 (,83)	2,46 (,92)	3,24 (1,16)	,04	,03
Portefólios e reflexões	2,11 (,90)	1,71 (,85)	4,69 (1,16)	,03	,03	1,88 (,84)	2,20 (,94)	4,59 (1,15)	,03	,03	2,13 (,89)	1,76 (,89)	4,63 (1,16)	,03	,03	2,16 (,87)	1,94 (,94)	1,80 (1,16)	,13	,01
Métodos Individuais	2,73 (,84)	2,38 (,98)	3,66 (1,16)	,06	,03	2,69 (,80)	2,72 (,91)	0,04 (1,15)	,84	,00	2,78 (,82)	2,25 (,93)	11,37 (1,16)	,001	,06	2,80 (,75)	2,55 (,98)	2,48 (1,16)	,07	,02
<b>Frequência da utilização de métodos de avaliação</b>																				
Métodos Coletivos	2,60 (,84)	2,49 (,94)	0,39 (1,17)	,53	,002	2,77 (,75)	2,42 (,91)	6,89 (1,16)	,00	,04	2,67 (,85)	2,27 (,82)	6,95 (1,17)	,009	,04	2,73 (,84)	2,40 (,87)	6,23 (1,16)	,01	,04
Portefólios e reflexões	1,88 (,75)	1,54 (,61)	5,18 (1,17)	,02	,03	1,66 (,68)	1,90 (,71)	4,67 (1,16)	,03	,03	1,93 (,76)	1,46 (,52)	13,56 (1,17)	<,001	,07	1,94 (,73)	1,68 (,73)	5,05 (1,16)	,03	,03
Métodos Individuais	2,48 (,69)	2,21 (,77)	3,69 (1,17)	,06	,02	2,38 (,70)	2,49 (,72)	0,99 (1,16)	,32	,006	2,54 (,70)	2,09 (,68)	12,77 (1,17)	<,001	,07	2,56 (,65)	2,31 (,77)	5,28 (1,16)	,02	,03

Nota. M<sub>e</sub> (média); DP (Desvio Padrão); gl (graus de liberdade)

Fonte: Elaboração própria.

Relativamente à área de conhecimento em que os professores lecionam, os testes multivariados revelaram a existência de diferenças estatisticamente significativas, embora com tamanho do efeito negligível, na valorização dos diferentes métodos de avaliação (WILK'S  $\Delta=,744$ ,  $F(12, 413)=4,063$ ,  $p<,001$ ,  $\eta^2=,094$ ) e de diferenças estatisticamente significativas, com tamanho de efeito pequeno na utilização dos diferentes métodos de avaliação WILK'S  $\Delta=,694$ ,  $F(12, 434)=5,360$ ,  $p<,001$ ,  $\eta^2=,115$ .

Os resultados dos testes univariados (quadro 5) e as comparações *pairwise* correspondentes permitiram identificar uma maior valorização de métodos coletivos nos professores que lecionam nas áreas das Ciências da Engenharia e Tecnologia em relação aos professores das Ciências Exatas ( $p=,001$ ) e Ciências Médicas e da Saúde ( $p=,047$ ); e nos professores das Ciências Sociais em relação aos professores das Ciências Exatas ( $p=,008$ ). Permitiram igualmente identificar uma maior valorização de portefólios e reflexões pelos docentes das Ciências Sociais em relação aos docentes das Ciências Exatas ( $p=,007$ ) e das Ciências da Engenharia e Tecnologia ( $p=,049$ ). As restantes comparações não foram estatisticamente significativas.

No que diz respeito à utilização dos métodos de avaliação, os testes univariados identificaram a existência de diferenças estatisticamente significativas entre professores de diferentes áreas de conhecimento, porém com tamanho do efeito negligível, na utilização de portefólios e reflexões e de métodos coletivos (ver quadro 5), favoráveis, de acordo com as comparações *pairwise*, à utilização de métodos coletivos pelos docentes das Ciências da Engenharia e Tecnologia em relação aos docentes das Ciências Exatas ( $p<,001$ ) e Humanidades ( $p=,004$ ); e dos docentes das Ciências Sociais em relação aos docentes das Ciências Exatas ( $p=,001$ ). Relativamente à utilização de portefólios e reflexões, as comparações *pairwise* revelaram uma maior utilização destes métodos pelos docentes das Ciências Sociais em relação aos docentes das Ciências Exatas ( $p=,015$ ), das Ciências Médicas e da Saúde ( $p=,009$ ) e das Ciências da Engenharia e Tecnologia ( $p=,005$ ).

Não foram encontradas diferenças estatisticamente significativas entre a experiência profissional dos participantes ( $U=3330$ ,  $p=,550$ ); os ciclos de estudo, Licenciatura ( $U=2090,50$ ,  $p=,369$ ), Mestrado Integrado ( $U=3647$ ,  $p=,799$ ), Mestrado ( $U=2689$ ,  $p=,378$ ) e Doutoramento ( $U=3570$ ,  $p=,184$ ); e a categoria profissional ( $\chi^2=0,990(2)$ ,  $p=609$ ), na valorização de testes e exames escritos. Contudo, foram encontradas diferenças estatisticamente significativas ao nível da área científica e da frequência de formação pedagógica (ver quadro 6). A valorização dos testes e exames escritos é significativamente maior nos professores que não possuem formação pedagógica. Relativamente às diferenças identificadas ao nível da área científica, foram aplicados testes *Mann-Whitney* subsequentes para investigar as diferenças entre grupos (*pairwise*). Foi aplicada uma correção de *Bonferroni* e adotado o nível de significância de  $p<,005$  (valor resultante da divisão do valor de significância,05 pelo número de testes realizados (10)). Os professores das Ciências Exatas reportaram uma maior valorização de testes e exames escritos do que os professores das Ciências Sociais ( $U=307,50$ ,  $p=,002$ ) e os professores das Humanidades reportaram uma maior valorização em relação aos professores das Ciências Sociais ( $U=391,50$ ,  $p<,001$ ). As restantes comparações entre grupos não foram significativas.

Quadro 5. Resultados dos testes univariados para a análise do efeito da área de conhecimento na valorização e frequência da utilização de métodos de avaliação

	ÁREA DE CONHECIMENTO					F (gl)	P	$\eta^2$
	CE M <sub>e</sub> (DP)	CET M <sub>e</sub> (DP)	CMS M <sub>e</sub> (DP)	CS M <sub>e</sub> (DP)	H M <sub>e</sub> (DP)			
<b>Valorização de métodos de avaliação</b>								
Métodos	1,89	2,90	2,26	2,72	2,28	5,919	<,001	,130
Coletivos	(,897)	(,644)	(,926)	(,887)	(,895)	(4,158)		
Portefólios e reflexões	1,46	1,87	1,83	2,35	2,11	4,361	,002	,099
	(,611)	(,828)	(,946)	(,948)	(,698)	(4,158)		
Métodos Individuais	2,23	2,76	2,51	2,74	2,69	1,280	,280	,031
	(,968)	(,824)	(,985)	(,864)	(,735)	(4,158)		
<b>Frequência da utilização de métodos de avaliação</b>								
Métodos	1,79	2,90	2,41	2,69	2,12	7,582	<,001	,154
Coletivos	(,587)	(,729)	(,834)	(,862)	(,805)	(4,166)		
Portefólios e reflexões	1,46	1,66	1,51	2,12	1,68	6,007	<,001	,126
	(,489)	(,689)	(,580)	(,817)	(,352)	(4,166)		
Métodos Individuais	2,20	2,35	2,24	2,56	2,57	1,689	,155	,039
	(,735)	(,676)	(,729)	(,722)	(,701)	(4,166)		

Nota. M<sub>e</sub> (média); DP (Desvio Padrão); gl (graus de liberdade), CE (Ciências Exatas), CET (Ciências da Engenharia e da Tecnologia); CMS (Ciências Médicas e da Saúde), CS (Ciências Sociais), H (Humanidades).

Fonte: Elaboração própria.

Ao nível da frequência da utilização de testes e exames escritos não foram igualmente encontradas diferenças estatisticamente significativas entre a experiência profissional dos participantes ( $U=3331$ ,  $p=,488$ ); os ciclos de estudo de Licenciatura ( $U=2215$ ,  $p=,648$ ), Mestrado Integrado ( $U=3501$ ,  $p=,385$ ), Mestrado ( $U=2810$ ,  $p=,481$ ) e Doutorado ( $U=3935$ ,  $p=,737$ ); e a categoria profissional ( $\chi^2= 2,525(2)$ ,  $p=,283$ ). Porém, foram encontradas diferenças estatisticamente significativas ao nível da área científica e da frequência de ações de formação na utilização de testes e exames escritos (ver quadro 6). A valorização dos testes e exames escritos é significativamente maior nos professores que não possuem formação pedagógica. Relativamente às diferenças identificadas ao nível da área de conhecimento, foram aplicados testes *Mann-Whitney* subsequentes para investigar as diferenças entre grupos (*pairwise*). Foi aplicada uma correção de *Bonferroni* e adotado o nível de significância de  $p<,005$  (valor resultante da divisão do valor de significância  $p<,05$  pelo número de testes realizados (10)). Os professores das Ciências Exatas reportaram uma maior frequência da utilização de testes e exames escritos em relação aos professores das Ciências da Engenharia e Tecnologia ( $U=280,50$ ,  $p=,023$ ), aos professores das Ciências Médicas e da Saúde ( $U=97,50$ ,  $p=,029$ ) e aos professores das Ciências Sociais ( $U=282,50$ ,  $p<,001$ ); e os professores das Ciências da Engenharia e Tecnologia reportaram uma maior frequência da utilização de testes e exames escritos em relação aos professores das Ciências Sociais ( $U=1298$ ,  $p=,001$ ). As restantes comparações entre grupos não foram significativas.

Quadro 6. Resultados dos testes não paramétricos para a análise do efeito da frequência de ações de formação pedagógica, área de conhecimento e indicação de práticas de avaliação na valorização e utilização de testes e exames escritos

	FREQUÊNCIA DE AÇÕES DE FORMAÇÃO PEDAGÓGICA				ÁREA DE CONHECIMENTO							INDICAÇÃO DE ALTERAÇÃO DE PRÁTICAS DE AVALIAÇÃO AO LONGO DA CARREIRA				
	Sim PM	Não PM	U	P	CE PM	CET PM	CMS PM	CS PM	H PM	$\chi^2$ (gl)	P	Sim	Não	Talvez	$\chi^2$ (gl)	P
<b>Valorização de testes e exames escritos</b>	83,98	102,60	2923	,012	120,60	96,28	98,95	74,42	123,29	24,64 (4)	<,001	88,58	119,44	97,44	6,68 (2)	,035
<b>Frequência da utilização de testes e exames escritos</b>	82,72	104,95	2775	,002	129,63	105,64	92,26	76,19	97,36	22,35 (4)	<,001	96,34	90,82	83,33	2,53 (2)	,283

Nota. PM (ponto médio); U (U de Mann-Whitney);  $\chi^2$  (Qui-quadrado); gl (graus de liberdade), CE (Ciências Exatas), CET (Ciências da Engenharia e da Tecnologia); CMS (Ciências Médicas e da Saúde), CS (Ciências Sociais), H (Humanidades)

Fonte: Elaboração própria.

### 3.2.3. Influência das variáveis relacionadas com as práticas de avaliação e com o Processo de Bolonha na valorização e utilização de métodos de avaliação

Um aspeto importante no contexto da investigação era compreender se os professores que afirmam ter alterado a forma como avaliam os seus alunos valorizam e utilizam métodos de avaliação diferentes dos professores que responderam negativamente. Os resultados da MANOVA revelaram não existir uma influência da mudança das práticas de avaliação na valorização (WILK'S  $\Delta=,931$ ,  $F(6, 314)= 1,905$ ,  $p=,80$ ,  $\eta^2=,035$ ) e na frequência da utilização (WILK'S  $\Delta=,951$ ,  $F(6, 330)=1,393$ ,  $p=,217$ ,  $\eta^2=,025$ ) dos diferentes métodos de avaliação. De igual modo, os resultados dos testes multivariados não permitiram identificar uma influência do reconhecimento do papel do Processo de Bolonha na mudança das práticas de avaliação na valorização (WILK'S  $\Delta=,985$ ,  $F(6, 310)=,402$ ,  $p=,878$ ,  $\eta^2=,008$ ) e na frequência da utilização (WILK'S  $\Delta=,949$ ,  $F(6, 326)=1,430$ ,  $p=,202$ ,  $\eta^2=,026$ ) dos diferentes métodos de avaliação.

Não foram encontradas diferenças estatisticamente significativas entre o reconhecimento do papel do Processo Bolonha na mudança das práticas de avaliação no Ensino Superior ( $\chi^2=0,101(2)$ ,  $p=,951$ ) na valorização de testes e exames escritos. Contudo, foram encontradas diferenças estatisticamente significativas ao nível da mudança de práticas de avaliação na valorização de testes e exames escritos (ver quadro 6). Foram aplicados testes *Mann-Whitney* para investigar as diferenças entre grupos. Foi aplicada uma correção de *Bonferroni* e adotado o nível de significância de  $p<,016$  (valor resultante da divisão do valor de significância  $p<,05$  pelo número de testes realizados (3)). Os professores que não alteraram as suas práticas de avaliação no Ensino Superior reportaram uma maior valorização de testes e exames escritos em relação aos professores que responderam afirmativamente ( $U=935,50$ ,  $p=,01$ ). As restantes comparações entre grupos não foram significativas.

Relativamente à frequência da utilização de testes e exames escritos não foram encontradas diferenças estatisticamente significativas ao nível do da influência do Processo Bolonha na mudança das práticas de avaliação no Ensino Superior ( $\chi^2=0,104(2)$ ,  $p=,949$ ), assim como da mudança de práticas de avaliação (ver quadro 6).

### 3.3. Correlação entre os métodos de avaliação mais valorizados e os métodos de avaliação mais utilizados

O quadro 7 apresenta a matriz de correlações entre a valorização e a frequência da utilização de métodos de avaliação. Foi identificada uma correlação significativa forte entre a valorização e frequência da utilização dos diferentes métodos de avaliação: uma maior valorização de métodos coletivos, portefólios e reflexões, e métodos individuais está associada a uma maior frequência da utilização dos mesmos. Foi também encontrada uma relação positiva e moderada entre a valorização e a frequência da utilização de testes e exames escritos.

Não obstante, a análise dos resultados do coeficiente de *Spearman* permitem identificar que uma maior valorização de portefólios e reflexões está associada a menor valorização e também menor utilização de testes e exames; uma maior utilização de portefólios e reflexões está associada a menor utilização de testes e exames; uma maior utilização de métodos coletivos está associada a menor utilização de testes e exames; e uma maior utilização de testes e exames escritos está associada a menor utilização e valorização de portefólios e reflexões; e a uma menor utilização de métodos coletivos.

Quadro 7. Matriz de correlação entre a valorização de métodos de avaliação e a frequência da utilização de métodos de avaliação (Correlação de Spearman)

	VALORIZAÇÃO DE MÉTODOS DE AVALIAÇÃO				FREQUÊNCIA DA UTILIZAÇÃO DE MÉTODOS DE AVALIAÇÃO			
	Métodos Coletivos	Portefólios e Reflexões	Métodos Individuais	Testes/exames escritos	Métodos Coletivos	Portefólios e Reflexões	Métodos Individuais	Testes/exames escritos
<b>Valorização de métodos de avaliação</b>								
Métodos Coletivos	1	,563***	,623***	-,115	,787***	,400***	,421***	-,138
Portefólios e Reflexões		1	,571***	-,172*	,482***	,822***	,489***	-,317***
Métodos Individuais			1	,054	,408***	,369***	,735***	,012
Testes/exames escritos				1	-,246**	-,295***	-,079	,584***
<b>Frequência da utilização de métodos de avaliação</b>								
Métodos Coletivos					1	,457***	,407***	-,171*
Portefólios e Reflexões						1	,409***	-,402***
Métodos Individuais							1	-,011
Testes/exames escritos								1

Nota. \* p < ,05; \*\* p < ,01; \*\*\* p < ,001.

Fonte: Elaboração própria.

#### 4. Discussão e conclusões

Este artigo visou explorar as seguintes questões de investigação: 1) Que importância atribuem aos métodos de avaliação?; 2) Qual a frequência da utilização dos diferentes métodos de avaliação?; 3) Quais as variáveis associadas à valorização e utilização de diferentes métodos de avaliação?; e 4) Qual a relação entre a valorização e a utilização de diferentes métodos de avaliação?

Para explorar estas questões, utilizou-se uma medida de valorização e outra de frequência de métodos de avaliação. Ambas as medidas demonstraram propriedades psicométricas adequadas, sendo a estrutura interna de ambas congruente com uma estrutura de três fatores: métodos coletivos, métodos individuais e portefólios e reflexões. Os itens de cada fator revelaram uma adequada consistência interna, suportando a fidelidade das pontuações obtidas. A valorização e frequência de utilização de testes/exames escritos foi analisada separadamente dada a ausência de saturação nos fatores anteriormente identificados.

Relativamente à influência das variáveis demográficas na valorização e frequência da utilização de métodos de avaliação, verificou-se não existir uma influência das variáveis demográficas (idade e sexo) ao nível da valorização e frequência da utilização dos métodos de avaliação em análise.

Ao nível das variáveis profissionais, a MANOVA permitiu identificar diferenças estatisticamente significativas em função dos ciclos de estudos na valorização e utilização

de métodos coletivos, individuais, e de portefólios e reflexões; e da área de conhecimento na valorização e utilização de métodos coletivos e portefólios e reflexões. Estudos anteriores realizados com alunos do Ensino Superior português (Pereira, 2016) identificaram igualmente uma influência do curso na utilização dos diferentes métodos de avaliação. Os resultados dos testes não paramétricos permitiram identificar uma influência da formação pedagógica e da área científica ao nível da valorização e frequência da utilização de testes e exames escritos. As respostas dos participantes permitiram identificar, por um lado, uma maior valorização de testes e exames escritos pelos professores que não possuem formação pedagógica e, por outro lado, uma maior valorização dos testes e exames escritos pelos professores das Ciências Exatas e das Humanidades em relação aos professores das Ciências Sociais. Identificou-se também uma maior frequência de utilização dos testes e exames escritos pelos professores das Ciências Exatas por comparação aos professores das Ciências da Engenharia e da Tecnologia, das Ciências Médicas e da saúde e das Ciências Sociais; e dos professores das Ciências da Engenharia e da Tecnologia em comparação com os professores das Ciências Sociais.

Em relação à influência das variáveis relacionadas com as práticas de avaliação e com o Processo de Bolonha na valorização e frequência da utilização de métodos de avaliação, não se verificou uma influência destas variáveis na valorização e utilização de métodos individuais, coletivos e de portefólios e reflexões. Porém, verificou-se a influência da mudança de práticas de avaliação ao nível da valorização de testes e exames escritos, sugerindo uma maior valorização de testes e exames pelos professores que afirmam não ter alterado as suas práticas de avaliação ao longo da sua carreira.

Na análise das correlações entre a valorização e a frequência da utilização de métodos de avaliação destaca-se a correlação positiva forte entre a valorização e utilização dos métodos de avaliação. Contudo, essa correlação é moderada ao nível da valorização e utilização de testes e exames escritos.

Estes resultados apontam para a valorização e frequência da utilização de um leque diversificado de métodos de avaliação pelos participantes neste estudo. Porém, os resultados das estatísticas descritivas apontam para uma maior valorização e frequência de utilização de testes e exames escritos e para uma menor valorização e frequência de utilização de portefólios e reflexões. Estes resultados corroboram outros estudos realizados no contexto português (Barreira et al., 2015; Pereira, 2016) que destacam o carácter sumativo da avaliação e a prevalência do uso de testes e exames escritos, articulada com a utilização de outros métodos de avaliação.

Estudos anteriores no contexto português apontam para a existência de contradições entre as práticas de avaliação utilizadas e as conceções de avaliação dos docentes (Gonçalves, 2016; Pereira e Flores, 2016) justificadas pelo volume de trabalho, escassez de recursos humanos e físicos e a imposição institucional na utilização de avaliação sumativa, que perpetuam o uso de determinados métodos de práticas (Pereira e Flores, 2016). Os resultados deste estudo indicam alguma coerência entre a valorização dos diferentes métodos de avaliação e a frequência da sua utilização, todavia, com menor expressividade na valorização e frequência da utilização de testes e exames escritos, o que poderá ser um resultado da imposição e prevalência da avaliação sumativa (Pereira e Flores, 2016).

Estes resultados realçam a complexidade dos cenários formativos universitários (Zabalza, 2004), da avaliação (Brown e Knight, 1994) e da profissão docente no sentido de encarar

a docência como um processo de criação e desenvolvimento de conhecimento através do estudo e exploração das suas diferentes dimensões, nomeadamente, o desenvolvimento profissional docente com particular ênfase em práticas de avaliação inovadoras (Fernandes, 2015).

Os resultados deste estudo contribuem para compreender o processo de avaliação na ótica de professores universitários portugueses, sugerindo-se a aplicação futura do instrumento a um número mais expressivo de docentes, de diferentes contextos universitários, para melhor conhecer as suas perceções acerca dos métodos e práticas de avaliação.

## Agradecimentos

Este estudo é financiado por Fundos Nacionais através da FCT (Fundação para a Ciência e a Tecnologia) e cofinanciado pelo Fundo Europeu de Desenvolvimento Regional (FEDER) através do COMPETE 2020 – Programa Operacional Competitividade e Internacionalização (POCI) com a referência POCI-01-0145-FEDER-007562 no âmbito do projeto “Assessment in Higher Education: the potential of alternative methods”, com a referência PTDC/MHCCED/2703/2014, e do projeto de doutoramento em Ciências da Educação - Especialidade em Desenvolvimento Curricular intitulado “Conceções e práticas de avaliação no Ensino Superior após a implementação do Processo de Bolonha: um estudo com professores universitários”, com a referência SFRH/BD/103291/2014.

O projeto é desenvolvido no Centro de Investigação em Estudos da Criança, no âmbito do Projeto Estratégico UID/ CED/00317/2013, por Fundos Nacionais através da FCT (Fundação para a Ciência e Tecnologia) e co-financiado pelos Fundos Europeus de Desenvolvimento Regional (FEDER), através do Programa Operacional de Competitividade e Internacionalização (POCI) com a referência POCI-01-0145-FEDER-007562.



## Referências

- Barreira, C., Bidarra, M. G., Vaz-Rebelo, P. Monteiro, F. e Alferes, V. (2015). Perceções dos professores e estudantes de quatro universidades portuguesas acerca do ensino e avaliação das aprendizagens. En D. Fernandes, A. Borralho, C. Barreira, A. Monteiro, D. Catani, E. Cunha, e P. Alves (Orgs.), *Avaliação, ensino e aprendizagem em Portugal e no Brasil: Realidades e perspectivas* (pp. 309-326). EDUCA.
- Black, P. e Wiliam, D. (1998), Assessment and classroom learning, *Assessment in Education*, 5(1), 7-75. <https://doi.org/10.1080/0969595980050102>
- Brew, C., Riley, P. e Walta, C. (2009). Participative assessment practices: A comparison of pre-service primary teachers and teaching staff views. *Assessment and Evaluation in Higher Education*, 34(6), 641-657. <https://doi.org/10.1080/02602930802468567>
- Brown, S. e Knight, P. (1994). *Assessing learners in higher education*. Kogan Page.

- Burkšaitienė, N. e Teresevičienė, M. (2008). Integrating alternative learning and assessment in a course of English for law students. *Assessment e Evaluation in Higher Education*, 33(2), 155-166. <https://doi.org/10.1080/02602930601125699>
- Cadime, I., Silva, C., Ribeiro, I. e Viana, F. L. (2018). Early lexical development: Do day care attendance and maternal education matter? *First Language*, 38(5), 503-519. <https://doi.org/10.1177/0142723718778916>
- Carless, D., Salter, M., Yang, M. e Lam, J. (2011). Developing Sustainable Feedback Practices. *Studies in Higher Education*, 36(4), 395-407. <https://doi.org/10.1080/03075071003642449>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Earlbaum Associates.
- Coutinho, C. P. (2014). *Metodologias de investigação em ciências sociais e humanas: Teoria e prática*. Edições Almedina, S.A.
- Dancey, C. P., Reidy, J. (2007). *Statistics without maths for psychology*. Pearson Education.
- Dochy, F., Segers, M., Gijbels, D. e Struyven, K. (2007). Assessment Engineering. Breaking down barriers between teaching, learning and assessment. En D. Boud e N. Falchikov (Eds.), *Rethinking assessment in higher education: learning for the longer term* (pp. 83-100). Routledge.
- Duncan, T. e Buskirk-Cohen, A. (2011). Exploring Learner-Centered Assessment: A Cross-Disciplinary Approach. *International Journal of Teaching and Learning in Higher Education*, 23(2), 246-259.
- Fernandes, D. (2015). Práticas de ensino e de avaliação de docentes de quatro universidades portuguesas. En D. Fernandes, A. Borralho, C. Barreira, A. Monteiro, D. Catani, E. Cunha, e P. Alves (Orgs.), *Avaliação, ensino e aprendizagem em Portugal e no Brasil: Realidades e perspectivas I* (pp. 97-135). EDUCA.
- Fernandes, S. e Flores, M. A. (2012). A docência no contexto da avaliação do desempenho no ensino superior: Reflexões no âmbito de um estudo em curso. *Revista Iberoamericana de Evaluación Educativa*, 5(2), 82-98.
- Field, A. (2009). *Discovering statistics using SPSS*. Sage Publications Ltd.
- Flores, M. A. e Veiga Simão, A. M. (2007, junho). Competências desenvolvidas no contexto do Ensino Superior: a perspectiva dos diplomados. In *V Jornadas de Redes de Investigación en Docencia Universitaria*. Alicante, Espanha.
- Flores, M. A., Veiga Simão A. M., Barros A. e Pereira, D. (2014). Perceptions of effectiveness, fairness and feedback of assessment methods: a study in higher education. *Studies in Higher Education*, 40(9), 1523-1534. <https://doi.org/10.1080/03075079.2014.881348>
- Gilles, J. L., Detroz, P. e Blais, J. G. (2010). An international online survey of the practices and perceptions of higher education professors with respect to the assessment of learning in the classroom. *Assessment e Evaluation in Higher Education* 36(6), 719-733. <https://doi.org/10.1080/02602938.2010.484880>
- Gonçalves, R. (2016). *Conceções de avaliação em contexto de ensino clínico de enfermagem: Um estudo na escola superior de enfermagem de Coimbra*. Tese de Doutoramento em Didática e Formação. Universidade de Aveiro, Portugal.
- Hadji, C. (1994). *A avaliação, regras do jogo*. Porto: Porto Editora.
- Hair, J. F., Black, W. C., Babin, B. J. e Anderson, R. E. (2009). *Multivariate data analysis*. Prentice Hall.
- Hutcheson, G. D. e Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics*. Sage Publications.

- Maclellan, E. (2004). How convincing is alternative assessment for use in higher education?. *Assessment e Evaluation in Higher Education*, 29(3), 311-321. <https://doi.org/10.1080/0260293042000188267>
- McDowell, L., Wakelin, D., Montgomery, C. e King, S. (2011). Does assessment for learning make a difference? The development of a questionnaire to explore the student response. *Assessment e Evaluation in Higher Education*, 36(7), 749-765. <https://doi.org/10.1080/02602938.2010.488792>
- Myers, C. B. e Myers, S. M. (2015). The use of learner-centered assessment practices in the United States: the influence of individual and institutional contexts. *Studies in Higher Education*, 40(10), 1904-1918. <https://doi.org/10.1080/03075079.2014.914164>
- Pereira, D. R. (2011). *A avaliação das aprendizagens no ensino superior na perspetiva dos estudantes. Um estudo exploratório*. Dissertação de Mestrado em Ciências da Educação. Universidade do Minho, Portugal.
- Pereira, D. R. (2016). *Assessment in higher education and quality of learning: Perceptions, practices and implications*. Tese de Doutoramento em Ciências da Educação. Universidade do Minho, Portugal.
- Pereira, D. R. e Flores, M. A. (2013). Avaliação e feedback no ensino superior: um estudo na Universidade do Minho. *Revista Iberoamericana de Educación Superior*, 4(10), 40-54.
- Pereira, D. R. e Flores, M. A. (2016). Conceptions and practices of assessment in Higher Education: A study of Portuguese university teachers. *Revista Iberoamericana de Evaluación Educativa*, 9(1), 9-29. <https://doi.org/10.15366/riee2016.9.1.001>
- Pereira, D., Flores, M. A. e Barros, A. (2017). Perceptions of Portuguese undergraduate students about assessment: a study in five public universities. *Educational Studies*, 43(4), 442-463. <https://doi.org/10.1080/03055698.2017.1293505>
- Perrenoud, P. (1999). *Avaliação: da excelência à regulação das aprendizagens: entre duas lógicas*. Porto Alegre: Artmed.
- Phelps, R. P. (2017). The “teaching to the test” family of fallacies. *Revista Iberoamericana de Evaluación Educativa*, 10(1), 33-49. <https://doi.org/10.15366/riee2017.10.1.002>
- Segers, M., Gijbels, D. e Thurlings, M. (2008). The relationship between students' perceptions of portfolio assessment practice and their approaches to learning. *Educational Studies*, 34(1), 35-44. <https://doi.org/10.1080/03055690701785269>
- Struyven, k., Dochy, F. e Steven, J. (2005). Students' perceptions about evaluation and assessment in higher education: a review. *Assessment e Evaluation in Higher Education*, 30(4), 325-341. <https://doi.org/10.1080/02602930500099102>
- Tabachnik, B. G. e Fidell, L. S. (1996). *Using multivariate statistics*. Harper e Row.
- Watering, G., Gijbels, D., Dochy, F. e Rijt, J. (2008). Students' assessment preferences, perceptions of assessment and their relationships to study results. *Higher Education*, 56, 645-58. <https://doi.org/10.1007/s10734-008-9116-6>
- Webber, K. L. e Tschepikow, K. (2013). The role of learner-centred assessment in postsecondary organisational change. *Assessment in Education: Principles, Policy e Practice*, 20(2), 187-204. <https://doi.org/10.1080/0969594x.2012.717064>
- Zabalza, M. A. (2008). *Competencias docentes del profesorado universitario. Calidad y desarrollo profesional*. Narcea.
- Zabalza, M. A. (2004). *La enseñanza universitaria. El escenario y sus protagonistas*. Narcea.

## Breve Cv de las autoras

### **Eva Lopes Fernandes**

Eva Lopes Fernandes é aluna do doutoramento em Ciências da Educação, especialização em Desenvolvimento Curricular na Universidade do Minho, com o tema: "Conceptions and Practices of Assessment in Higher Education: A study of University Teachers" (SFRH/BD/103291/2014). É Licenciada em Educação e Mestre em Ciências da Educação pela Universidade do Minho. Tem várias publicações sobre temas relacionados com o trabalho dos professores, profissionalismo docente, liderança docente e as vozes dos alunos. Atualmente integra a equipa de investigação do projeto "Impact - Investigando os Efeitos das Lideranças Escolares nos Resultados dos Alunos", tendo integrado várias equipas de investigação de projetos nacionais e internacionais (e.g. "Assessment in Higher Education: the potential of alternative methods"; "Teachers Exercising Leadership" and "Teachers professional trajectories"). ORCID ID: <https://orcid.org/0000-0002-3838-9846>. Email: [evalopesfernandes@ie.uminho.pt](mailto:evalopesfernandes@ie.uminho.pt)

### **Maria Assunção Flores**

Professora na Universidade do Minho. Doutorou-se em Educação na Universidade de Nottingham, Reino Unido, tendo sido *visiting scholar* na Universidade de Cambridge e na Universidade de Glasgow. É membro de várias associações científicas internacionais e pertence ao corpo editorial de várias revistas internacionais. É diretora executiva da revista *Teachers and Teaching Theory and Practice* e co-diretora da *European Journal of Teacher Education*. Foi presidente da *International Study Association on Teachers and Teaching*, tendo ainda presidido à direção do *International Council on Education for Teaching*. É coordenadora e membro fundador da Rede Internacional de Investigação-Ação Colaborativa (Estreiadialogos). É ainda membro do Conselho Geral do IAVE, I.P. As suas áreas de investigação incluem formação e desenvolvimento profissional de professores, profissionalismo docente e identidade profissional, currículo, avaliação e ensino superior. Tem mais de 200 publicações nestes domínios, incluindo livros, capítulos de livros e artigos em revistas nacionais e internacionais. Coordenou vários projetos de investigação, sendo os mais recentes "Os professores e o exercício da Liderança" e "Investigando os efeitos das lideranças nos resultados escolares dos alunos", financiados pela Fundação para a Ciência e a Tecnologia. <https://orcid.org/0000-0002-4698-7483>. Email: [aflores@ie.uminho.pt](mailto:aflores@ie.uminho.pt)

### **Irene Cadime**

Licenciada e Doutorada em Psicologia, na especialidade de Psicologia da Educação, pela Universidade do Minho. Mestre em Intervenção Psicológica, Educação e Desenvolvimento Humano pela Universidade do Porto. Atualmente é investigadora e membro integrado do Centro de Investigação em Estudos da Criança do Instituto de Educação da Universidade do Minho. Desenvolve trabalhos de investigação nos domínios da psicometria, do desenvolvimento de competências linguísticas e do sucesso académico. ORCID ID: <https://orcid.org/0000-0001-8285-4824>. Email: [ireneCadime@ie.uminho.pt](mailto:ireneCadime@ie.uminho.pt)

### **Clara Pereira Coutinho**

Licenciada em Economia, Mestre em Educação na área de especialização de Tecnologia Educativa e Doutora em Educação na área de especialização de Tecnologia Educativa, grau que obteve no ano de 2003 na Universidade do Minho, Braga, Portugal. Atualmente é Professora Auxiliar Aposentada do Departamento de Estudos Curriculares e Tecnologia Educativa do Instituto de Educação da Universidade do Minho sendo responsável pela coordenação da linha de investigação “Recursos Educativos Digitais” no Centro de Investigação em Estudos da Criança. Tem desenvolvido atividades de pesquisa no âmbito da Formação de Professores em Tecnologias de Informação e Comunicação e ainda no domínio das Metodologias de Investigação em Educação. Mais recentemente, desenvolve investigação ao nível do *mobile learning* e da utilização de aplicativos da Web 2.0 como ferramentas de apoio ao ensino e à aprendizagem, tendo publicados dezenas de artigos em revistas internacionais de referência bem como em atas de reuniões científicas nacionais e internacionais. Participa em diversos projetos financiados por agências nacionais e internacionais tendo recebido diversos prémios, como é o caso do projeto *t-words*, um manipulativo digital para crianças do pré-escolar que recebeu o World Technology Award em 2013. Publicou dois livros um dos quais no domínio das Metodologias de Investigação em Ciências Sociais e Humanas que teve a sua 2ª edição publicada em 2013. Informação adicional pode ser encontrada na sua página pessoal disponível em <http://www.degois.pt/visualizador/curriculum.jsp?key=1426606078182665>. ORCID ID: <https://orcid.org/0000-0002-2309-4084>. Email: [ccoutinho@ie.uminho.pt](mailto:ccoutinho@ie.uminho.pt)