Revista EaD &
tecnologias digitais na educação

Das provas de nível presenciais em língua estrangeira aos testes de nível online: o caso da Universidade Aberta

Margarida Martins

margarida.martins@uab.pt

Ana Burgues Setién

ana.burgues@uab.pt

Isabelle Simões Marques

isabelle.marques@uab.pt

Katja Göttsche

katja.clara@uab.pt

Universidade Aberta, Portugal

Resumo: *No presente artigo propomos uma análise da implementação de provas de diagnostico com resposta automática online na Universidade Aberta (Portugal) como resposta à pandemia COVID-19 de 2020, devido a qual não foi possível os estudantes se deslocarem à Universidade para realizarem as provas presencialmente. São discutidas as vantagens dos testes online no que diz respeito a planeamento, correção e aferição de nível. Também analisamos os critérios de harmonização que foram aplicados para as quatro línguas lecionadas na Universidade Aberta (Alemão, Espanhol, Francês e Inglês) assim como as pedagogias relacionadas com este tipo de avaliação. Por fim elaborámos uma reflexão sobre esta experiência, demonstrando as especificidades deste tipo de avaliação no ensino superior.*

Palavras-chave: *prova de nível; ensino online; ensino a distância; língua estrangeira*

Abstract: *In this article, we propose an analysis of the implementation of automatic diagnostic online placement tests at Universidade Aberta (Portugal) as a result of the 2020 pandemic, which prevented undergraduate paper-based exams from being administered at the University. We will discuss the*

advantages online testing has on planning, marking and student placement. We also analyse the harmonising criteria that were put in place for the four existing foreign languages at the Universidade Aberta (German, English, French and Spanish) as well as the pedagogical strategies related to this type of assessment. Finally, we draw up an evaluation of this experience, demonstrating the specificities of this type of assessment in higher distance education.

Keywords: *placement test; online education; distance learning; foreign languages*

"All tests contain the condition of intent: they are planned, arranged, given or conducted by someone with some purpose in mind." (Hanson, 1993: 16)

1. Contexto de ensino-aprendizagem: a Universidade Aberta

A Universidade Aberta é a única instituição pública de ensino superior em Portugal inteiramente online. Fundada em 1988, o seu público-alvo é composto por um público adulto residente em Portugal ou em diversos países do mundo, e que enfrenta muitas vezes a sua primeira, segunda ou terceira experiência de ensino superior em todos os ciclos de estudo (1º, 2º e 3º ciclos). Em 2008, o ensino passou a ser exclusivamente online, com base nos princípios metodológicos definidos pelo Modelo Pedagógico Virtual, publicado em 2007 e atualizado em 20186. O Modelo Pedagógico Virtual da Universidade Aberta está baseado nos princípios da aprendizagem centrada no estudante e nos primados da flexibilidade, da interação e da inclusão digital. Este modelo pedagógico, em regime inteiramente online, recorre às tecnologias mais recentes e à comunicação em rede, promovendo uma aprendizagem colaborativa e a criação de uma comunidade de aprendizagem virtual.

A Universidade Aberta acolhe mais de 5000 estudantes com perfis muito diversos, embora a grande maioria já esteja inserida no mercado de trabalho e recorra ao ensino superior a distância para melhorar a sua posição profissional ou como forma de aprofundar e completar os seus estudos e conhecimentos. A flexibilidade da gestão do tempo investido na realização de atividades acadêmicas que oferecem esta modalidade de ensino, torna-se muito atraente e conveniente para este público alvo. No entanto, o estudante que enfrenta pela primeira vez esta experiência de aprendizagem a distância precisa também de adquirir e desenvolver competências e estratégias próprias, não só cognitivas como também metacognitivas.

2. O contexto da pandemia mundial: novos horizontes nas propostas de avaliação dos estudantes

[6] PEREIRA Alda, MENDES António Quintas, MORGADO Lina, AMANTE Lúcia, BIDARRA José (2007). *Modelo pedagógico virtual da Universidade Aberta: para uma universidade do futuro*. Lisboa: Universidade Aberta.
MENDES António Quintas, BASTOS Glória, AMANTE Lúcia, AIRES Luísa Lebres, CARDOSO Teresa (2018). *Modelo pedagógico virtual: cenários de desenvolvimento*. Lisboa: Universidade Aberta.

ISSN 2318-4051

A partir do 2º semestre de 2020, devido à pandemia de COVID 19, que teve repercussões em todos os setores a nível global, a avaliação final das unidades curriculares de 1º ciclo, até à data realizada presencialmente, passou a ser digital. Da mesma forma, as provas de nível em língua estrangeira, que têm como objetivo avaliar as competências dos candidatos em língua estrangeira, e efetuar a sua colocação no nível correspondente num curso de 1º ciclo (Licenciatura em Línguas Aplicadas, Estudos Europeus, Humanidades, História) de acordo com o Quadro Europeu Comum de Referência para as Línguas (Conselho da Europa, 2001), anteriormente presenciais, passaram a ser realizadas online através da ferramenta "Teste" da plataforma Moodle, plataforma usada na Universidade Aberta.

A avaliação constitui uma componente importante no contexto educacional, pois demonstra através de resultados (qualificativos e/ou quantitativos) a aquisição de conhecimentos e competências ao longo de um determinado tempo de aprendizagem. Por esta razão, a testagem de proficiência em língua estrangeira deve ser um assunto sujeito a uma reflexão aprofundada e à análise de resultados.

Com o impacto da pandemia e as repercussões no funcionamento da vida académica (para não mencionar o impacto na vida familiar e pessoal que também se refletiu na produção e no desempenho académico dos estudantes), a Universidade Aberta foi obrigada, assim como todas as instituições de ensino, a reinventar-se face às mudanças impostas pelo isolamento e distanciamento social. No segundo semestre do ano 2020 foi implementada a avaliação final digital para os estudantes do 1º ciclo de estudos. Esta solução veio colmatar a impossibilidade da presença física num espaço destinado à realização da prova presencial escrita, substituindo-a por um formato inteiramente digital com submissão via plataforma Moodle. A adaptação ao contexto digital aconteceu num espaço de tempo muito curto, sem formação ou partilha de experiências prévias da parte dos docentes e estudantes. A mudança das provas em papel (presenciais) para digital (a distância) orientou uma reflexão, por parte das docentes, essencial sobre as provas, a saber: a sua estrutura, os seus objetivos e conteúdos, o seu modo de realização e resultados.

O presente artigo, elaborado após a primeira experiência da realização das provas de nível em formato digital utilizando a ferramenta "Teste"[7] da plataforma Moodle, tem como objetivo geral avaliar o impacto que esta alteração teve para avaliados, avaliadores e instituição, como também refletir sobre aspetos específicos da construção das provas que surgiram antes, durante e após esta primeira realização das provas de nível neste novo formato.

Nesse sentido, este artigo tem os seguintes objetivos específicos:

- analisar o contexto e processo de mudança do presencial para o digital;
- demonstrar como foram definidos critérios gerais e específicos para a construção das provas nas quatro línguas em oferta na Universidade Aberta;
- apresentar uma síntese da organização de cada prova (itens, cotação, estrutura);
- avaliar os resultados e refletir sobre propostas de melhoria;

[7] https://docs.moodle.org/310/en/Quiz_activity

- realçar a importância do trabalho reflexivo e colaborativo no contexto educacional.

A metodologia utilizada na intervenção, análise do processo de avaliação e implementação das provas é a da Investigação em Ação (IA, Lewin, 1944). Esta metodologia tem vindo a ser aplicada em duas áreas: a sociologia, desenvolvida precisamente por Lewin (entre outros Tax, 1958; Borda, 1970); e a educação, inspirada nas ideias e práticas de Freire (1974), Taba (1957) e mais recentemente desenvolvida por Elliot (1990), Hernández Sampieri (1997, 2018), Fernández Collado (1997,) ou Baptista Lucio (1997). Quando aplicada no âmbito da educação o seu foco de estudo está relacionado com a complexidade das atividades em sala de aula (materiais, dinâmicas, conteúdos, avaliação), segundo o ponto de vista dos intervenientes. Assim, foram selecionadas uma série de ações pedagógicas que acabaram por ser também etapas da investigação, a saber: elaborar, experimentar, avaliar e redefinir através de um processo de autocrítica e reflexão cooperativa os modos de intervenção, os processos de ensino-aprendizagem, o desenvolvimento do currículo e o desenvolvimento profissional dos docentes e as instituições educativas (Miguélez, 2000: 30).

Esta metodologia abarca simultaneamente a investigação científica bem como a resolução de problemas em sala de aula, e desenvolve competências fundamentais nos investigadores. A IA funciona de tal maneira que permite a constante intervenção nos processos educativos através de um ciclo de: identificação do problema, criação do plano de intervenção, execução, avaliação da superação do problema, auto-avaliação, repetição de ciclo de forma a assegurar a correta intervenção em sala de aula. O objetivo é sempre tornar os docentes investigadores da sua própria práxis (Miguélez, 2000: 31). Consideramos, após a experiência, que a IA facilitou a análise e posterior otimização (ainda em curso dadas as características da metodologia) dos processos de avaliação, mais concretamente, da avaliação diagnóstica do nível de domínio de uma língua estrangeira.

3. Porquê avaliar o nível de língua?

De modo a responder a esta pergunta é relevante abordar a noção de critical language testing (Shohami, 1998) no que diz respeito não só à necessidade de avaliar o nível de língua dos estudantes que iniciam a sua licenciatura, como também o impacto que as alterações implementadas na Universidade Aberta tiveram para docentes e estudantes.

> Critical language testing (CLT) refers to the examination of the uses and consequences of tests in education and society (Shohamy 2001a, b; Spolsky 1995). The topic gained attention by various scholars and particularly Messick (1981, 1989), who argued for expanding the definition of construct validity as a criterion for evaluating the quality of tests, to include components related to tests use, such as values, impact, and consequences. (Shohamy 2017)

Os resultados das provas de nível são, não só determinantes enquanto indicadores do nível de língua do estudante avaliado, como se refletem na classificação final do curso de 1º ciclo e têm impacto no futuro académico dos estudantes. Com base no re-

sultado da prova, os estudantes são colocados no nível correspondente pelo que, de acordo com a precedência dos níveis de língua em vigor na Universidade Aberta, prosseguem para o nível seguinte e assim sucessivamente durante os três anos da licenciatura. A prova de nível de língua é opcional e os estudantes que não optam por realizá-la ficam inscritos por defeito no nível I de língua ou no nível III se tiverem conhecimentos e/ou qualificações que comprovem a sua proficiência de língua. Tornar a prova de nível obrigatória é uma questão que poderá ser posta à consideração, se isso significar uma distribuição mais justa dos estudantes por níveis de língua e tornar as turmas mais homogêneas.

É fundamental que o avaliado que decida inscrever-se nas provas de nível de língua estrangeira compreenda o objetivo da prova, como se realiza, as suas consequências, ou seja, o que representam os resultados (Hanson 1993) para o seu percurso académico. Aliás a prova de nível pode ter um impacto direto na nota final do curso, pois a classificação obtida fica validada para os níveis concedidos. Para além disso, a sua classificação define o nível em que o avaliado deve ser colocado de acordo com a proficiência demonstrada, como já foi referido. Por outro lado, o docente deve elaborar a prova de acordo com critérios bem definidos que sirvam para adequadamente avaliar o estudante que a realiza. Segundo Shohamy, "the power of tests also lies in the ownership of the information. The tester, not the test taker, owns the testing information, the "the scientific knowledge"" (1998: 338). A investigação, a reflexão e a implementação de estratégias são essenciais para boas práticas de avaliação de maneira a que o avaliado, o avaliador e a instituição estejam preparados e seguros com este novo tipo de prova em formato digital. A avaliação deve adaptar-se às circunstâncias em que se realiza, tendo em conta o público, o tipo de instituição e os seus objetivos (Huver & Springer, 2011). Podemos definir este tipo de teste, retomando a definição dada por Tagliante (2005):

> Ces tests servent à orienter l'élève vers un groupe de niveau homogène et à l'informer de ses capacités langagières, afin qu'il puisse prendre, en connaissance de cause, les décisions qui concernent son apprentissage. Le test de niveau permet de faire une sorte de bilan des atouts et des points faibles de l'élève et constitue une source d'information essentielle pour l'enseignant. (Tagliante, 2005: 14)

Percebemos a importância da qualidade do teste para posterior bom funcionamento das aulas durante o semestre. Para tal, o teste deve cobrir uma grande quantidade de conteúdos linguísticos, avaliar o maior número possível de competências e os itens devem ser variados (Veltcheff & Hilton, 2003).

4. As provas de nível de língua estrangeira (de presencial a digital)

As provas de nível de língua estrangeira, do mesmo modo que as provas de primeira e segunda épocas, tiveram de ser adaptadas à realidade do confinamento geral. No caso dos estudantes que já se encontravam a frequentar o 1º ciclo na Universidade Aberta, a solução para a avaliação final foi operacionalizada dentro das próprias unidades curriculares. Os docentes criavam e colocavam as provas no dispositivo inserido num tópico de avaliação, parametrizado pelos serviços de apoio tecnológico ao ensino.

No entanto, para novos candidatos à Universidade Aberta, num curso de 1º ciclo, sem nunca terem tido acesso à plataforma Moodle, as provas de nível não poderiam ser

realizadas desta forma. Foi necessário encontrar outra solução. Como já foi anteriormente referido, os estudantes têm a possibilidade de realizar uma prova de nível em língua estrangeira cujo resultado define o nível de língua e respetiva unidade curricular em que se deve inscrever. Estas provas diagnósticas têm como objetivo 'medir' com precisão (Brown, 2001) competências e conhecimentos de maneira a que os estudantes sejam adequadamente colocados num nível de língua.8 Até 2020 as provas de nível em língua estrangeira, que tinham uma duração de 2h30 (1h30 + 60 minutos de tolerância), com a cotação de 20 valores, eram realizadas pouco antes do início do ano letivo, na última semana do mês de setembro, presencialmente, nos vários locais de realização de exame, quer em Portugal, nas instalações da Universidade Aberta em Lisboa, Coimbra e Porto, quer nos diversos Centros Locais de Aprendizagem (CLAs) ou no estrangeiro, geralmente em consulados portugueses.

As provas (enunciados e folhas de teste) eram distribuídas em papel aos candidatos. Após a realização das provas, essas eram encaminhadas por via postal ou digital até à sede da Universidade Aberta em Lisboa, encaminhadas para o docente responsável, corrigidas e devolvidas aos serviços administrativos com a devida pauta assinada em papel. Esse processo administrativo, longo e demorado, retardava a colocação dos estudantes nas respectivas UCs do primeiro semestre. As provas escritas tinham como objetivo avaliar de um modo geral as competências de compreensão e produção escritas assim como a gramática através de uma seleção de itens que serviam como 'amostra' de conteúdos programáticos das UCs. Segundo Hansen (1993: 18), "this indicates that a test is a special sort of investigation in which the information that is collected is not itself the information that one seeks but instead a representation of it." E todos os anos se repetia o processo com a criação de uma prova, nos mesmos moldes, mas com perguntas diferentes. Como este procedimento se tornou numa rotina individual de início de semestre, nunca houve verdadeiramente uma reflexão mais aprofundada e conjunta sobre o formato, conteúdos, critérios e resultados, transversal às quatro línguas em oferta. Esta experiência obrigou-nos a interromper um processo automatizado e pôr em prática um trabalho colaborativo e reflexivo importante. A citação que se segue de Richards & Lockhart descreve bem a 'nossa' experiência com esta mudança de formato do presencial para o digital, como resultado da pandemia,

> events occur which the teacher can use to develop a deeper understanding of teaching. Teachers sometimes fail to exploit these events, letting momentum of all the other events of the day take precedence. And yet these experiences can serve as the basis for critical reflection, if teachers can find ways to capture the thoughts of and reactions to these events, as well as ways to gather fuller information on the events themselves. From this basis teachers can develop strategies for intervention or change, depending on their needs. (Richards & Lockhart 1996: 6)

Face à pandemia e ao confinamento geral, as provas de nível em língua estrangeira passaram a ser realizadas a distância e as datas foram antecipadas para o início do mês de setembro de 2020, um mês antes do arranque do ano letivo. Para além disso, optou-se igualmente pela criação da prova na ferramenta "Teste", disponível na plata-

[8] Ver tabelas 1 e 2.

forma Moodle, dando acesso aos estudantes a um espaço na plataforma de aprendizagem criada especialmente para o efeito. Essa ferramenta utilizada na criação e realização de testes formativos, e agora pela primeira vez numa prova formal, trouxe inúmeras vantagens para a instituição, para as docentes e para os estudantes, pois desmaterializou por completo o processo de diagnóstico e acelerou-o. Enumeramos algumas das vantagens que apuramos:

- deixou de ser necessário a deslocação do estudante até o local de realização do exame;
- deixou de ser necessário contar com salas, examinadores e pessoal docente para a realização das provas;
- deixou de haver provas físicas;[9]
- a correção das provas passou a ser automática;[10]
- o tempo de disponibilização da nota foi reduzido, permitindo aos estudantes conhecer o nível em que são colocados quase de forma instantânea;[11]
- a partir dos resultados, as docentes introduziram as notas numa pauta eletrónica, posteriormente enviada por e-mail aos serviços académicos;
- os estudantes ficaram inscritos nas turmas correspondentes antes do início do semestre.

Após análise dos resultados da mudança de formato de prova, fomentada pela pandemia, foi possível chegar à conclusão que, como indica o antigo provérbio inglês, "necessity is the mother of invention"12; assim como aconteceu em muitas outras áreas da sociedade, uma situação de emergência resultou num modelo com resultados comprovadamente positivos para os estudantes, para as docentes e para a instituição.

Iremos agora analisar o processo de criação da prova em termos da seleção de conteúdos e respetivos critérios, para uma avaliação adequada das competências e conhecimentos necessários para frequentar as disciplinas de língua estrangeiras na Universidade Aberta.

5. Definição de critérios gerais e específicos para a construção das provas

Na Universidade Aberta, assim como em todo o paradigma do ensino a distância, online e em rede, desenvolve-se uma relação entre docente, estudante e tecnologia (que engloba recursos e ferramentas digitais),

[9] Uma opção que também tem implicações ao nível de sustentabilidade ambiental pois não são necessários recursos físicos, nem a utilização de espaços físicos, nem a deslocação e presença de colaboradores de apoio à prova e são evitados consumos excessivos de energia (García & Navarro, 2008).

[10] A primeira edição das provas de nível online em 2020 não teve produção escrita com correção manual.

[11] A ferramenta permite a disponibilização imediata de feedback; no entanto, uma vez que se tratava de uma primeira experiência pedagógica, as docentes resolveram rever manualmente as respostas antes de disponibilizar os resultados (calculados, sim, de forma automática).

[12] O provérbio é uma adaptação ao longo dos tempos de uma célebre frase de *A República* de Platão, originalmente traduzida como "our need will be the real creator".
https://www.thedrum.com/profile/mediarithmics/news/the-ongoing-impact-of-the-covid-19-pandemic-means-a-data-marketing-infrastructure-is-more-important-than-ever

> Nesta "nova" realidade da Educação Digital, as salas de aula necessitam de estar conectadas a instrumentos e redes de conhecimento em permanente atualização, sendo que a hierarquia clássica da relação entre professor e estudante está a transformar-se num ecossistema de conhecimento cada vez mais complexo. (Moreira et al., 2020: 5)

O ecossistema do ensino online entre docentes, estudantes e plataforma Moodle permite a criação de atividades e provas, com conhecimento prévio dos conteúdos e metodologias da disciplina (por parte dos estudantes) e conhecimento dos estudantes (por parte dos docentes). O mesmo não acontece nas provas de nível onde existe um certo anonimato de ambos os lados, uma vez que a maioria dos estudantes ainda não integrou a Universidade Aberta. Enquanto que as provas de avaliação final nas unidades curriculares são elaboradas de acordo com critérios e conteúdos específicos do nível de língua trabalhado em determinada unidade curricular, as provas de nível são desenvolvidas com base num conjunto de itens representativos dos conhecimentos e competências necessários para frequência de um determinado nível de língua (para qual os estudantes estão a ser avaliados).

Do ponto de vista do desenho da prova de nível de língua estrangeira, o foco responde aos critérios de equidade, validade, praticabilidade e fiabilidade13 enquadrados no QECR e adaptados aos níveis de língua em oferta na universidade, procurando um impacto positivo nos agentes envolvidos no processo (avaliadores e avaliados). Deste modo, procurámos incorporar estes critérios gerais na prova, assegurando:

- que a prova fosse coerente e precisa enquanto instrumento de avaliação (fiabilidade);
- que efetivamente medisse de forma rigorosa e consistente o que é pretendido medir (validade);
- que fosse viável e acessível no contexto da sua realização (praticabilidade);
- que desse aos avaliados toda a informação relacionada com as provas (equidade).

A decisão sobre as componentes selecionadas para a avaliação, o grau de dificuldade e complexidade, a duração, os critérios e a cotação de cada seção da prova, assim como o reconhecimento da coerência dos conteúdos por parte do estudante garantem a validade e fiabilidade da prova na sua função de avaliar o nível linguístico. Segundo Brown (2001: 391),

---

13 Estes critérios, utilizados para avaliar uma prova, foram consultados em espanhol, francês e em inglês: *validad, fiabilidad, viabilidad / validité, fiabilité, praticabilité / validity, reliability, practicality*. A tradução para o português foi das autoras.
c.f. BERNÁRDEZ Paula, ÁLVAREZ Reyes (2016), Formación del profesorado de ELE y evaluación: construcción de una guía para autentificar pruebas de español lengua extranjera. *Porta Linguarum: revista internacional de didáctica de las lenguas extranjeras*, 25, 165.
c.f. BROWN Hamish Douglas (2001), *Teaching by Principles: An Interactive Approach to Language Pedagogy 2nd Edition*. Longman, 386-389.
c.f. LUSSIER Denise, TURNER Carolyn (1993), *Le point sur l'évaluation en didactique des langues*. Anjou, Québec : Centre Éducatif et Culturel.

ISSN 2318-4051

> A placement test typically includes a sampling of material to be covered in the curriculum (that is, it has content validity), and it thereby provides an indication of the point at which the student will find a level or class to be neither too easy nor too difficult, but appropriately challenging. (Brown, 2001: 391)

Com base nestes critérios de avaliação da própria prova ("testing the test", Brown, 2001: 385), foi possível não só elaborar um modelo de medição do grau de eficiência do estudante em diversas competências comunicativas da língua estrangeira, como também encontrar um modelo que fosse transversal às 4 línguas em oferta na Universidade Aberta, apesar de algumas diferenças estruturais, que iremos discutir mais adiante. Os resultados das provas permitiram aferir de forma mais sistemática e objetiva o nível de conhecimentos e as capacidades do avaliado, em qualquer uma das línguas estrangeiras oferecidas no 1º ciclo na Universidade Aberta. Os resultados demonstraram que é possível existir uma categorização estandardizada que permite a interpretação dos resultados como representativos de uma classificação que corresponde a um dado nível de língua, segundo o QECR.

No entanto, na reflexão posterior à realização das provas, temos de realçar uma certa margem de erro pois a fiabilidade da prova depende de vários fatores: da prova em si (ambiguidades, questões técnicas, clareza, duração), situacionais (interferências, interrupções, ruído) e individuais (motivação, ansiedade, fatores psicológicos) (Cohen, 2001: 525). Sendo uma prova online, nem sempre é possível assegurar a realização dos testes por parte de todos os participantes sem percalços, sejam estes de ordem técnica, de natureza pessoal ou familiar, ou devido a falhas na redação e/ou interpretação das perguntas e instruções. A questão da equidade deve ser, portanto, tomada em conta na realização do exame para que todos os candidatos tenham as mesmas hipóteses de sucesso (Noël-Jothy & Sampsonis, 2006). Evidentemente, estas possíveis limitações foram tidas em conta no processo de tomada de decisões relativas aos níveis de aferição determinados pela instituição e as docentes responsáveis. Tendo em conta que a classificação obtida nas provas de nível confere habilitações de acesso às unidades curriculares de Língua I, II, III e IV, foi determinado um regime de avaliação e classificação que confere uma certa flexibilidade para garantir a validade e confiabilidade da prova. Assim, como foi referido na introdução, as provas de nível são classificadas numa escala numérica de 0 a 20 valores, e os resultados das mesmas correspondem aos níveis indicados na seguinte tabela:

Classificação	Unidade Curricular
Até 11 valores	Língua I
De 12 a 14 valores	Língua II
De 15 a 17 valores	Língua III
De 17 a 20 valores	Língua IV

Tabela 1: Nível de língua correspondente ao resultado da prova de língua estrangeira

Os níveis das quatro línguas estrangeiras em oferta na Universidade Aberta diferem um pouco entre elas, assim, apresentamos a seguinte tabela com a informação do nível segundo o QECR:

Unidade Curricular	Nível correspondente
Língua I	A1.1: alemão A1.2: francês, espanhol B1.1: inglês
Língua II	A1.2: alemão A2.1: francês, espanhol B1.2: inglês
Língua III	A2.1: alemão A2.2: francês B1: espanhol B2.1: inglês
Língua IV	A2: alemão B1.1: francês B2: espanhol B2.2: inglês

Tabela 2: Nível de língua para cada língua estrangeira segundo o QECR

Uma vez estabelecidos os objetivos e critérios gerais das provas de nível (Brown, 2001), procedemos às definições específicas para a criação das provas. Nesta fase, embora estivesse sempre presente a colaboração e reflexão conjunta, cada docente desenhou a sua própria prova para a língua que leciona.

> Test constructors first make an inventory of the objectives that they want to test. This involves distinguishing broad objectives from more specific ones and important objectives from trivial ones. Test items and procedures are then developed to assess these objectives either separately or along with other objectives. (Cohen, 2001: 523)

Desta forma, e após a reflexão conjunta, cada docente optou por criar a sua prova com base em escolhas sobre os seguintes aspetos da estrutura das provas:

- seleção de itens de avaliação;
- distribuição dos itens por secções ou sequência de perguntas sem secção;
- número de perguntas na prova ou em cada secção;
- ordem das perguntas e/ou secções de acordo com a competência a ser avaliada;
- cotações para cada pergunta, grupo de perguntas e/ou secção;
- inclusão ou não de compreensão e produção escrita e oral;
- colocação das perguntas com ordem ascendente de grau de dificuldade.

Para a construção das provas foram selecionados itens para a avaliação das competências, e conhecimentos linguísticos adequados aos conteúdos programáticos das disciplinas. Sendo estas provas de correção automática, com o objetivo de avaliar a proficiência para colocação do estudante num nível de língua, optámos por dois tipos de

ISSN 2318-4051

prova, dependendo da língua14, com 30-60 perguntas de escolha múltipla, resposta única, resposta curta e cloze entre outras disponíveis na ferramenta "Teste": 1) prova dividida por secções; 2) prova sem secções com uma sequência de perguntas.

A seleção do tipo de pergunta depende do objetivo e conhecimentos a serem avaliados, por isso a dificuldade cognitiva das mesmas deve ser levada em consideração. No caso das provas de nível de língua estrangeira15, esta responde à taxonomia de Bloom (1956) que assume que a aprendizagem nos níveis superiores depende da aquisição de conhecimentos e habilidades de certos níveis inferiores. Ou seja, o desenho da prova é condicionado pelo nível de dificuldade das perguntas que, à medida que o teste avança, incrementam o seu grau de desafio cognitivo e a utilização dos recursos comunicativos na língua alvo.

A aplicação dos pressupostos desta taxonomia nos testes permite a hierarquização dos processos cognitivos em diferentes níveis, facilitando o processo de avaliação. Assim, o tipo de pergunta selecionada corresponde a uma série de critérios cognitivos, que são os seguintes: conhecer, compreender, aplicar, analisar, criar, avaliar.

Segue uma breve descrição de cada prova:

Alemão

A prova de nível de Alemão era composta por 60 perguntas de diferentes tipologias com um grau de dificuldade ascendente e com uma ordem sequencial, ou seja, as perguntas não eram ordenadas aleatoriamente. No entanto, as respostas às perguntas do tipo "escolha múltipla" e do tipo "relacionar" foram misturadas.

Para a elaboração do teste foi calculado como ponto de partida um tempo médio de 2 minutos para a resolução de cada pergunta, tendo em conta o tempo total da prova de 120 minutos. Houve naturalmente perguntas com um grau de dificuldade menor que iriam requerer menos tempo para a resolução e outras, por exemplo de compreensão de texto, onde seria necessário mais tempo para a sua resolução.

Ainda que o teste formalmente não estivesse dividido em secções, foi dividido em três partes, que correspondiam a um determinado nível de dificuldade definido pelo QECR. Cada uma destas secções era constituída por 20 perguntas de ordem de grau de dificuldade crescente. Dentro das secções, as perguntas não eram divididas por competências a avaliar, mas sim ordenadas por grau de dificuldade. Com base em perguntas de tipologia variada foram avaliadas diferentes competências linguísticas, ou seja, competências gramaticais, domínio do léxico, domínio de fórmulas discursivas, compreensão escrita, etc.

Para avaliar a compreensão escrita foi contemplado um leque abrangente de textos, desde o nível A1 até ao nível B1, leque esse que incluía, por exemplo, mensagens eletrónicas, postais, anúncios de apartamentos, ofertas de emprego ou um excerto de um contrato de trabalho. Em função do seu grau de dificuldade, estes textos faziam parte das respetivas "secções" do teste de nível.

Por fim, esta prova não contemplava a produção escrita, nomeadamente perguntas de desenvolvimento, ao contrário do que era habitual nos testes de nível de alemão realizados presencialmente.

Espanhol

[14] A decisão sobre a estrutura da prova não tem relação direta com a língua, mas sim com a docente.

[15] Com exceção da prova de inglês.

A prova de Espanhol contava com 42 perguntas que pretendiam avaliar a competência gramatical (estruturas e léxico), a compreensão escrita e a interpretação textual. Esta estava estruturada por secções que correspondem precisamente às diversas competências em avaliação. Mais concretamente, a prova foi desenhada da seguinte maneira:

- Secção 1, competência gramatical, 27 questões de escolha múltipla;
- Secção 2, compreensão escrita e interpretação de texto, 8 questões "verdadeiro/falso" sobre um texto disponibilizado (excertos de um romance de Miguel Delibes, *El hereje*) ;
- Seção 3, competência gramatical e lexical, 7 questões de resposta aberta com erros gramaticais e léxicos que o estudante devia corrigir.

A prova contemplou um grau de dificuldade ascendente, especialmente no que diz respeito ao domínio gramatical e lexical (as semelhanças linguísticas entre Português e Espanhol- línguas de parentesco- permitem que um estudante com um baixo nível de domínio da língua consiga compreender quase na sua totalidade um texto escrito de nível intermédio-alto). Assim, as perguntas de resposta múltipla foram ordenadas (não permitindo assim, a navegação livre pelo teste) em base nos conteúdos lecionados nos níveis A2-B2 numa ordem ascendente.

Francês

No que diz respeito à prova de francês, essa era composta por duas partes distintas: uma primeira parte dedicada à gramática e a segunda parte dedicada à compreensão escrita e vocabulário. A parte da gramática continha 30 perguntas sobre diferentes aspetos gramaticais e as questões tinham uma tipologia variada: "escolha múltipla", "resposta curta", "verdadeiro/falso", "correspondência", "arrastar e largar", e "selecionar palavras em falta". Cada pergunta tinha a cotação de 2 valores, num total de 60 valores. Nessa mesma parte também foi colocada uma questão que tinha como suporte um documento autêntico, a saber um cartaz publicitário e era cotada com 5 valores.

A segunda parte, dedicada à compreensão escrita a ao vocabulário, com uma cotação global de 35 valores, apresentava dois textos escritos com perguntas "verdadeiro/falso" assim como um terceiro texto no qual deviam colocar as palavras em falta.

As perguntas, que incidiam sobre aspectos verbais, preposicionais, sintáticos, lexicais e pragmáticos, não foram baralhadas e seguiram um grau de dificuldade ascendente, partindo do nível A1.2 para, no final do teste, ter perguntas de nível B1.1, nível máximo no qual os candidatos podem ser colocados, a saber Francês IV.

Inglês

A prova de nível de inglês com um total de 20 valores estava organizada por secções ordenadas da seguinte forma: gramática (8 valores), compreensão de texto (8 valores) e técnicas de escrita (4 valores). Os avaliados eram testados sobre os seus conhecimentos de tempos verbais, preposições, sintaxe, léxico, compreensão de texto e conectores, com perguntas de "escolha múltipla", exercício "cloze", "verdadeiro ou falso" e "correspondência".

A primeira secção de gramática era composta por 4 partes: 1) pergunta cloze com 10 questões onde os avaliados tinham de escrever o verbo em entre parênteses no tempo verbal correto; 2) 10 perguntas de escolha múltipla onde era necessário escolher a forma correta do verbo para completar uma frase; 3) 10 questões de correspondência onde era necessário selecionar a preposição correta para completar uma frase; 4) 10

questões de correspondência onde era necessário completar uma oração para completar as frases.

Na segunda secção de compreensão de texto os avaliados tinham de ler um excerto do texto The Old Man and the Sea de Ernest Hemingway e responder a 10 perguntas de "verdadeiro ou falso" e duas perguntas cloze cada uma com 10 questões, onde era necessário completar as frases com palavras retiradas do texto e identificadas, e outra onde era necessário completar as frases com palavras que tinham que ser encontradas no texto.

A terceira e última secção da prova, dado que optámos por não incluir perguntas de desenvolvimento, serviu para avaliar técnicas de escrita. Os avaliados tinham que completar um texto com conectores discursivos (however, although, also, but, moreover, nevertheless, etc).

A prova não foi organizada por ordem ascendente de grau de dificuldade mas continha um conjunto de itens que refletia o tipo de matéria e conteúdos linguísticos trabalhados nos 4 níveis de língua Inglesa no qual os estudantes podem ser colocados com base nos resultados da prova de nível, B1.1 - B2.2.

A opção de não incluir perguntas de desenvolvimento com correção manual foi tomada em conjunto no sentido de facilitar e acelerar a publicação de resultados, para além de tornar o resultado mais objetivo:

> The objectivity of an item refers to the way it is scored. A multiple-choice item, for example, is objective in that there is usually only one right answer. A free composition may be more subjective in nature if the scorer is not looking for any one right answer, but rather for evidence of a series of features, including, say, creativity, style, cohesion and coherence, grammar and mechanics (Cohen 2001: 517).

Por questões técnicas, as provas não incluíam avaliação de compreensão e produção oral, mas esta é uma questão sobre a qual estamos a refletir para provas futuras.

6. Validação das provas de nível

A revisão e verificação das provas foi um passo fundamental na criação dos testes. Como a validação dos testes prendia-se com aspetos pedagógicos e técnicos, esta foi realizada pelas docentes com o apoio técnico e verificação dos serviços de apoio tecnológico ao ensino. Uma vez que estas provas foram as primeiras a serem realizadas online, os resultados e a validação ficaram sujeitos à análise posterior à disponibilização dos resultados, que foram automáticos.

Ao contrário das provas ou outros trabalhos de avaliação ao longo do semestre, estas provas de nível não têm feedback individualizado - um aspeto importante da avaliação de língua como afirma Cohen (2001: 515): "a more constructive view of language testing exists when (...) (g) tests are returned promptly and (h) the results are discussed (Shohamy, 1985)". Nas provas de nível não houve lugar para esse feedback com discussão de resultados e sugestões de melhoria, mas somente a indicação da resposta certa ou errada. A validação partiu da reflexão e aplicação de critérios que garantiam que as provas fossem criadas com responsabilidade e rigor e que de facto mediam aquilo que propunham medir, pois, "the real aim accordingly was to create tests which are more fair, considerate, constructive, and ethical in terms of their power" (Shohamy, 2017). A

fiabilidade das provas só poderia ser avaliada pelo desempenho do estudante ao longo do semestre no nível em que ficou colocado.

7. Reflexão sobre os resultados

Após a realização das provas, e de acordo com o plano de intervenção, as docentes envolveram-se num processo de auto-reflexão individual e conjunta com o intuito de otimizar a avaliação no âmbito das provas de nível e, eventualmente, nas unidades curriculares de avaliação formal. Sendo conscientes de que a IA fomenta a consolidação de estratégias, técnicas e procedimentos para tornar esse processo rigoroso, sistemático e crítico (Miguélez, 2000), plasmámos num relatório conjunto os problemas ocorridos e sugestões de melhoria do processo de avaliação (desenho, criação, monitorização, correção e acompanhamento do estudante). Deste modo, conseguimos a sistematização de um método de diagnóstico do problema, análise e intervenção que permitisse a correta aferição dos estudantes no seu nível de língua, bem como o seu seguimento nas UCs selecionadas.

Assim, com a aplicação desta metodologia e através do relatório mencionado, foi possível apurar conclusões sobre as vantagens e desvantagens deste formato de prova, a sua estrutura, o tipo de perguntas e resultados obtidos, bem como planear potenciais alterações em provas futuras. Foi-nos possível compreender aspetos que induziram em erro os estudantes por não estarem bem definidos, claros ou configurados. Esta reflexão ajudou-nos a perceber, entre outras, a importância da clareza das instruções e da estrutura da prova: a prova deve definir de forma clara cada secção, com instruções simples e sem ambiguidades (Cohen, 2011). Os estudantes que frequentem unidades curriculares estão preparados para as provas no final do semestre e estão familiarizados, não só com os conteúdos, como também com o formato das perguntas e do tipo de resposta esperado. Os avaliados nas provas de nível não tinham conhecimento prévio nem dos conteúdos, nem dos tipos de perguntas. Consideramos a hipótese de colocar uns exemplos de perguntas, ou uma prova modelo no portal juntamente com o formulário de inscrição, ou, antes de cada pergunta, colocar, para além das instruções, um exemplo de pergunta e resposta. Segundo Cohen,

> Examples may help, but on the other hand may hinder if they do not give the whole picture and become a substitute for reading the instructions. (...) They should also know the value of each item and section of the test. Finally, the time allowed for each subtest and/or for the total test should be announced. (Cohen, 2011: 524)

De referir que as instruções, a cotação, a duração da prova assim e de cada secção estivam presentes por escrito no início da prova e de cada secção (no caso das provas com secção). As provas tinham ainda um cronómetro que começava a contagem decrescente a partir do momento em que o avaliado iniciava a sua prova.

De modo a avaliar a validade da prova, a testagem por pares é um procedimento útil e através do qual podem ser feitos ajustes nos itens e nas instruções antes da realização da prova. Também há de se ter em conta que, especialmente no caso de perguntas com mais de uma opção (como é o caso das questões “verdadeiro/falso” ou “escolha múltipla”), os estudantes podem escolher uma resposta aleatória, o que invalida os ob-

jetivos de uma prova de nível com este formato (Mothe, 1975), sem perguntas de desenvolvimento ou com respostas mais abertas (produção escrita e oral).

8. Considerações finais

É inegável que a avaliação é parte integrante do processo de aprendizagem de uma língua estrangeira. No entanto, existem diferentes formas de avaliar e o debate sobre o melhor tipo de avaliação é ainda inconclusivo. A avaliação não é necessariamente sinónimo de "teste", pois é um conceito muito mais abrangente. Existem vários elementos que podem contribuir para uma avaliação rigorosa e representativa da aprendizagem, dos conhecimentos, da proficiência e do desempenho do estudante. Como afirma Brown (2001) um bom professor está sempre envolvido direta ou indiretamente na avaliação dos seus estudantes.

Em língua estrangeira há ainda outro fator a considerar: uma aprendizagem completa implica a aquisição de 4 competências linguísticas: compreensão e produção oral, compreensão e produção escrita. Torna-se ainda mais complexo o processo de avaliação pois é necessário decidir se é possível avaliar todas estas competências numa única prova (a chamada avaliação sumativa) ou se ao longo de um determinado tempo (período ou semestre) com vários momentos de avaliação (avaliação formativa). O processo de avaliação requer um método (Brown 2001) com estratégias definidas e com um envolvimento ativo por parte do docente e do estudante. No caso específico das provas de nível (avaliação diagnóstica) e como já foi referido, existiram outros fatores a considerar na definição de itens de avaliação (objetividade na aferição de nível, rápida publicação de resultados, impossibilidade de realização presencial devido à pandemia).

No caso concreto da avaliação de uma língua estrangeira através da prova de nível, consideramos que a mudança para o formato digital via ferramenta "teste" da plataforma Moodle constituiu uma excelente oportunidade para dar continuidade a este formato nas disciplinas de línguas estrangeiras lecionadas nos cursos de 1º ciclo16. Através do uso de ferramentas digitais podemos considerar a hipótese de incluir, na avaliação final, outras competências comunicativas que tradicionalmente estão restringidas à avaliação semestral, i.e. a compreensão e produção oral. Deste modo, pretendemos promover uma continuidade e homogeneidade das competências avaliadas (uma vez que a Universidade Aberta é uma instituição de ensino online) no processo de aprendizagem que, de facto, se reflete também na avaliação.

Do ponto de vista do estudante, o facto da prova ser a distância vai ao encontro da modalidade de ensino escolhida inicialmente, num ambiente virtual onde se sente seguro. Estes fatores, aliados ao facto de saber antecipadamente que irá obter resultados imediatos, influem positivamente na realização da prova e posterior classificação.

Do ponto de vista dos docentes, consideramos que trabalhar com ferramentas digitais no contexto da avaliação facilita e acelera o processo de correção, chegando mesmo a aperfeiçoá-lo. As questões puramente gramaticais são auto-corrigidas permitindo assim incrementar o tempo que o docente investe na correção e fornecimento de feedback mais significativo e pormenorizado em questões de índole pragmática, de uso da língua, de coerência textual etc.

16 Esta experiência, pioneira, deu origem à criação de um projeto piloto (2021) para adaptação da avaliação presencial ao formato digital, com recurso à ferramenta "teste" noutras áreas disciplinares no seio da universidade Aberta.

Sabemos agora, após a experiência de aprendizagem que implicou esta adaptação, que estamos prontos, enquanto comunidade académica, para a mudança de comportamentos, modelos e dinâmicas, em parte graças ao desenvolvimento e à sistematização das competências digitais que temos vindo a inculcar aos nossos estudantes e a nós próprios. É o nosso objetivo continuar a investigar estas linhas de ação pedagógica.

Referências

BARBA MARTÍN Raúl (2019), *La investigación-acción participativa desde la mirada de las maestras participantes en un proceso de formación permanente del profesorado sobre educación inclusiva*. Tesis doctoral. Valladolid: Universidad de Valladolid.

BERNARDÉZ Paula, ALVÁREZ Reyes (2016), "Formación del profesorado de ELE y evaluación: construcción de una guía para autentificar pruebas de español lengua extranjera", *Porta Linguarum: revista internacional de didáctica de las lenguas extranjeras,* 25, 163-177.

BORDA Orlando (1970), *Ciencia propia y colonialismo intelectual,* México: Nuestro Tiempo.

BROWN, H. Douglas (2001), *Teaching by Principles: An Interactive Approach to Language Pedagogy*. London: Longman, 2nd Edition.

CASCANTE César, BRAGA Gloria (1994), "Una guía práctica. Investigación en la acción, práctica pedagógica", *Cuadernos de pedagogía*, 224, 20-23.

COHEN Andrew D. (2001), "Second Language Assessment". In Celce-Murcia M. (eds). *Teaching English as English as a Second or Foreign Language*. Heinle & Heinle Publishers, 3rd Edition.

ELLIOT John (1990), *La investigación-acción en educación*. Madrid: Morata.

FREIRE Paulo (1974), *Pedagogía del oprimid*o, Madrid: Siglo XXI.

GARCÍA CANO Maria Elena (2008), La evaluación por competencias en la educación superior. *Profesorado. Revista de curriculum y formación de profesorado*, 12(3), 1-16.

GARCÍA LABORDA Jesús, NAVARRO LABOULAIS Cristina (2008), "Automatizar los exámenes del Diploma de Español como Lengua Extranjera (DELE):¿ utopía o realidad?" *RLA. Revista de lingüística teórica y aplicada*, 46(2), 81-93.

HANSON F. Allan (1993), *Testing testing: Social consequences of the examined life*. Berkeley: University of California Press.

HUVER Emmanuelle, SPRINGER Claude (2011), *L'évaluation en langues*. Paris : Les Éditions Didier.

LEDESMA Alejo González López, ÁLVAREZ Guadalupe, BASSA Lorena (2018), "Educación virtual en la universidad: un estudio de Investigación-acción para la enseñanza de la asignatura Tecnología educativa", *Ciencia, docencia y tecnología*, 29(57), 201-234.

LLOSA Lorena (2017), Assessing Students' Content Knowledge and Language Proficiency. In: Shohamy E., Hornberger, N. (eds) *Language Testing and Assessment. Encyclopedia of Language and Education*. Springer: Cham, 3rd Edition.

LUSSIER Denise, TURNER Carolyn (1993), *Le point sur l'évaluation en didactique des langues.* Anjou, Québec : Centre Éducatif et Culturel.

MARTINS Margarida, MARQUES Isabelle Simões, GÖTTSCHE Katja, SETIÉN Ana, CHE-NOLL Antonio, CHILDS Jeffrey Scott, OLIVEIRA Susana (2020), *Inovação e Tecnologia no Ensino de Línguas: Reflexões e Perspetivas de Ação em Contextos Educacionais Diversos*. Lisboa: Universidade Aberta.

MENDES António Quintas, BASTOS, Glória, AMANTE Lúcia, AIRES Luísa Lebre, CARDOSO Teresa (2018), *Modelo pedagógico virtual: cenários de desenvolvimento*. Universidade Aberta.

MIGUÉLEZ M. M. (2000). "La investigación-acción en el aula". *Agenda académica*, 7(1), 27.

MOREIRA José António, HENRIQUES Susana, BARROS Daniela, GOULÃO Fátima, CAEIRO Domingos (2020), *Educação digital em rede: princípios para o design pedagógico em tempos de pandemia*. Lisboa: Universidade Aberta.

MOTHE, Jean-Claude (1975), *L'évaluation par les tests dans la classe de français*. Paris : Hachette.

NOËL-JOTHY Françoise, SAMPSONIS Béatrix (2006), *Certifications et outils d'évaluation en FLE*. Hachette : Paris.

PEREIRA Alda, MENDES António Quintas, MORGADO Lina, AMANTE Lúcia, BIDARRA José (2007), *Modelo pedagógico virtual da Universidade Aberta: para uma universidade do futuro.* Lisboa: Universidade Aberta.

RICHARDS Jack, LOCKHART Charles (1994), *Reflective Teaching in Second Language Classrooms.* Cambridge: Cambridge University Press.

SAMPIERI Roberto, COLLADO Carlos, LUCIO Baptista Pilar (1998) *Metodología de la investigación,* México, DF: McGraw-Hill, 233-426, vol. 1.

SAMPIERI Roberto, (2018). *Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta*. México, DF: McGraw-Hill.

SETIÉN Ana, MARTINS Margarida Pereira, MARQUES Isabelle Simões, CLARA Göttsche Katja (2019), "Criando uma comunidade virtual de aprendizagem para o ensino de LE na Universidade Aberta". In: Almeida J. Domingues, Tomé Auf der Maur S., Bacquelaine F., Lorenzo M. Barros, Duarte I. M., Hurst N., Kind A., Martínez P. Nicolás, Rodrigues S. Valente (orgs.), *Para lá da tarefa: implicar os estudantes na aprendizagem de línguas estrangeiras no ensino superior*. Porto: Faculdade de Letras da Universidade do Porto, 362-377.

SETIÉN Ana, MARTINS Margarida Pereira, MARQUES Isabelle Simões, CLARA Katja Göttsche (2019), "Oralidade e avaliação na prática de e-learning na Universidade Aberta". In: Costa, M. J. et al. (orgs.), *CNaPPES 2018*. Braga: Universidade do Minho, 55-60.

SHOHAMY Elana (1998), "Critical Language Testing and Beyond", *Studies in Educational Evaluation,* 24(4), 331-345.

SHOHAMY Elana (2017), “Critical Language Testing”. In: Shohamy E., Hornberger, N. (eds) *Language Testing and Assessment. Encyclopedia of Language and Education,* Springer: Cham, 3rd Edition.

TABA Hilda, NOEL Elisabeth (1957), *Action research: a case study*. Washington: ASCD.

TAGLIANTE Christine (2005), *L’évaluation et le cadre européen commun*. Paris : Clé international.

TAX Sol (1985), “The fox project”, *Human Organization*, 17(1), 17-19.

VELTCHEFF Caroline, HILTON Stanley (2003), *L’évaluation en FLE*. Hachette : Paris.