

POR Nível: Construção e validação de um teste de colocação para o Português Língua Estrangeira – resultados de um estudo-piloto¹

Carolina Gramacho^{}, Ana Madeira⁺, Cláudia Martins⁺, Nélia Alexandre^{*}, Jorge Pinto^{*} & Susana Correia⁺*

^{}Universidade de Lisboa, CLUL/⁺NOVA FCSH/CLUNL*

Abstract:

Placement tests are a form of assessment that aims at placing students in appropriate levels. Placement tests must, therefore, be based on validity and reliability arguments to ensure appropriateness and consistency of assessment. Despite the growing interest in Portuguese as a Foreign Language (PFL), both nationally and internationally, to our knowledge there is no validated placement test that informs on general language abilities of adult PFL learners across levels. The *POR Nível* project aims at designing and validating a placement test for PFL adult learners, from A1 to C1 CEFR levels in grammar, vocabulary, listening and reading components. Corpora of written productions of 15 learners of 3 native languages (English, Mandarin and Spanish) from certified exams of PFL provide the empirical basis for items construction. Also, items' specifications follow national and international guidelines (CEFR, Camões Profile for PFL), validity, reliability and practicality requirements, and also ALTE and ILTA recommendations. The results found thus far will provide empirical ground for item specification in a validated placement test and eventually contribute to the discussion of the contents of PFL teaching programs.

Keywords: placement test, validation, Portuguese as a Foreign Language, assessment

Palavras-chave: teste de colocação, validação, Português Língua Estrangeira, avaliação

1. Introdução

Os testes de colocação em nível de língua são uma forma de avaliação de diagnóstico que tem como finalidade garantir o posicionamento dos alunos em níveis de proficiência adequados. Por essa razão, um teste de colocação em nível de língua deve ser um instrumento válido e fiável, capaz de servir os seus propósitos de maneira consistente.

Apesar do interesse crescente na área do Português como Língua Estrangeira (PLE), quer nacional, quer internacionalmente, não existem testes exclusivamente construídos para colocação em nível de língua validados², que avaliem de forma segura e fiável as competências linguísticas de aprendentes adultos de PLE. As consequências lógicas deste facto, e face à pouca fiabilidade das competências autorreportadas, são a

¹ Este trabalho foi desenvolvido no Centro de Linguística da Universidade Nova de Lisboa, no âmbito do projeto cofinanciado pela NOVA FCSH (UID/LIN/03213/2013) e pela Fundação Calouste Gulbenkian. Devemos um agradecimento especial ao Professor Doutor Sérgio Moreira, pela colaboração em discussões metodológicas e pela realização da análise estatística dos dados do teste piloto. Agradecemos ao CAPLE pelo fornecimento de dados de *corpora*, essenciais à realização deste projeto, aos estudantes e docentes da Faculdade de Letras da Universidade de Coimbra e da Faculdade de Letras da Universidade do Porto pela colaboração no teste piloto e aos professores do CLCP da NOVA FCSH e do ICLP da Faculdade de Letras da Universidade de Lisboa pela colaboração no questionário de pilotagem do material inserido na versão piloto do teste.

² Referimos a existência do teste DIALANG, um teste de diagnóstico de competências linguísticas, validado para várias línguas, com o apoio da Comissão Europeia, inserido no programa SOCRATES, LINGUA 2. O DIALANG é um teste de diagnóstico que serve o propósito de estabelecer pontos fortes e pontos fracos detalhados sobre as competências linguísticas dos aprendentes, podendo ser usado opcionalmente para agrupar alunos em níveis de língua (Alderson, 2005). No entanto, o DIALANG apresenta uma estrutura complexa, constituída por cinco fases de realização opcional: I - escolha de um teste, II - teste de diagnóstico, III - autoavaliação, IV - teste linguístico e V - *feedback*: resultados e aconselhamento. Não se trata, portanto, de um teste exclusiva e idealmente desenhado para colocação em nível de língua, como se pretende que o POR Nível seja.



existência de uma considerável variabilidade na avaliação entre instituições de ensino e no perfil dos aprendentes colocados no mesmo nível de aprendizagem de língua.

Neste quadro, surge o projeto POR Nível, que visa não só construir e validar um teste de colocação em nível de língua para aprendentes de PLE, do nível A1 ao nível C1 do *Quadro Europeu Comum de Referência para as Línguas* (QEQR), nas dimensões de gramática, vocabulário, compreensão oral e compreensão escrita, como também permitir comparar as competências dos alunos de diferentes instituições de ensino e contribuir para o desenvolvimento de estratégias e métodos de ensino baseados no potencial informativo da aplicação de um teste de colocação validado, servindo as necessidades curriculares dos alunos.

A importância da utilização de instrumentos de avaliação validados tem sido atestada por diversos autores na literatura. Berry (2008) e Gardner (2006) destacam a importância da utilização de materiais de avaliação para a recolha de dados que auxiliem a atividade docente. As aplicações de instrumentos de avaliação auxiliam e informam o ensino, a aprendizagem e a pesquisa científica.

Um conceito de extrema importância no desenvolvimento de ferramentas de avaliação é o conceito de validade (McNamara, 2006; Messick, 1989). Havendo vários subtipos de validade, são referidos recorrentemente três subconceitos, em manuais internacionais de boas práticas para a construção de instrumentos de avaliação: (a) a validade de conteúdo; (b) a validade de construto; (c) a validade concorrente (ALTE, 2001; ETS, 2013). Entende-se por validade de conteúdo o grau de fidelidade dos itens relativamente ao que pretendemos medir ou avaliar. Uma das maneiras de garantirmos a validade do conteúdo é através da integração, na equipa responsável pela construção do teste, de um painel de peritos, selecionando apenas itens e tarefas que resultem de um consenso de opiniões. A validade de construto diz respeito às inferências que podemos fazer sobre a adequação dos itens, baseada na *performance* dos participantes, relativamente ao que pretendemos avaliar. Podemos garanti-la através da análise estatística, conduzindo, por exemplo, uma análise fatorial exploratória e confirmatória, para perceber até que ponto os itens avaliam aquilo para que foram construídos. Já a validade concorrente está associada ao poder relativo do teste, isto é, à relação entre os resultados que obtemos com o nosso teste e critérios independentes utilizados para avaliar as mesmas competências. Neste caso, falamos, por exemplo, do estudo e consulta de referenciais e da comparação do nosso teste com testes concorrentes.

Assim, a validade constrói-se como um conceito multidimensional que implica avaliação de peritos, tratamento estatístico de dados, recolha e análise de *corpora*, análises comparativas de instrumentos de avaliação concorrentes e, sobretudo, tempo e trabalho em equipa (Alderson & Benerjee, 2002; Chapelle, 1999; Cushing, 2017).

Outro conceito relevante no âmbito da validação de instrumentos de avaliação é o conceito de fiabilidade, que se relaciona com o de validade. A questão-chave acerca da validade dos instrumentos de avaliação é “Estarei eu a medir aquilo que pretendo medir?”, enquanto a questão que se coloca acerca do conceito de fiabilidade é “Se eu aplicar este teste mais do que uma vez, obterei os mesmos resultados ou resultados consistentes?”. Apesar de podermos diferenciar estes dois conceitos e de ambos subentenderem abordagens metodológicas distintas, existe uma forte inter-relação entre os conceitos de validade e fiabilidade. Com efeito, medir a validade de um instrumento é, muitas vezes, medir a sua fiabilidade (AERA, APA, & NCME 1996, 1999, 2014; Alderson & Banerjee, 2002; Black & William, 2006; CEFRL, 2001; Chapelle, 1999, 2008; Cureton, 1951; ETS, 2013; ITC, 2013; Messick, 1989).

A avaliação para colocação em nível de língua estrangeira pode assumir diferentes formas e funções. No que à forma diz respeito, podemos, numa avaliação longitudinal, ter uma avaliação pré e pós-programa (intrassujeito) ou podemos ter uma avaliação transversal (intersujeitos). No que à função diz respeito, a mesma pode ser apenas a de colocar aprendentes em níveis, programas ou cursos de língua adequados ou pode ter como objetivo a construção de fronteiras entre níveis ou descritores mais completos e realistas para cada nível. O objetivo é sempre o mesmo: contribuir para métodos e estratégias de ensino mais eficazes e para



sermos capazes de medir e comparar a proficiência dos aprendentes (Koizumi *et al.*, 2011; Sims, 2015; Tran, Griffin & Nguyen, 2010).

O objetivo deste trabalho é apresentar uma descrição metodológica do procedimento de validação de um teste de colocação em nível de língua para o Português Língua Estrangeira (PLE). Nas secções seguintes descrevemos a metodologia, com apresentação de resultados da análise de *corpora* e com a descrição do procedimento de validação dos materiais utilizados nas secções de compreensão oral e escrita. Apresentaremos ainda os resultados da aplicação de uma versão piloto do teste a aprendentes do Português.

2. Metodologia

O teste de colocação em nível de língua que está a ser desenvolvido no âmbito do projeto POR Nível é baseado em três tipos de dados: (i) documentos orientadores, perfis de língua e referenciais nacionais e europeus; (ii) testes de colocação validados desenvolvidos para outras línguas e (iii) análise de *corpora*.

No que aos documentos orientadores diz respeito, foram consultados os seguintes: *QECR, Referencial Camões PLE*, programas e perfis de outras línguas (*Plan Curricular del Instituto Cervantes, English Profile, Référentiel de l'Alliance Française, Principles of Good Practices for ALTE Examinations* (Out. 2001), *ALTE's Manual for Language Test Development and Examination* (Abr. 2011), *Educational Testing System (ETS) Guidelines* (2013), *Cambridge English Principles of Good Practice* (Maio 2011)). Nestes documentos, foi identificado e analisado o potencial conteúdo curricular a incluir no teste, para cada nível de proficiência.

Quanto aos testes de colocação para outras línguas, consultámos o teste do *Goethe Institut* para o alemão, o teste do *Instituto Cervantes* para o espanhol, o teste da *Alliance Française* para o francês e o teste para o inglês do *Cambridge* e do *British Council*. Aqui, procurou-se analisar a forma e o conteúdo de testes de colocação em nível, ou seja, analisar os formatos das perguntas e das respostas, e os conteúdos linguísticos e comunicativos.

O *corpus* consiste em textos produzidos em exames certificados do CAPLE, que foram usados com o objetivo de (i) fornecer uma base empírica para a especificação dos itens do teste e (ii) estabelecer um perfil de aprendente do PLE adulto, permitindo a identificação dos pontos fortes e fracos do aprendente em cada nível (cf. Alderson, 1996; Cushing, 2017; Egbert, 2017; e Taylor & Barker, 2008).

2.1. Primeira fase: análise de *corpora*

Os *corpora* analisados consistiram em setenta e cinco textos dos géneros “carta” e “texto argumentativo”, em registo formal e informal, retirados da parte de produção escrita de exames certificados do CAPLE, de candidatos nativos de inglês, mandarim e espanhol, três das línguas maternas mais representadas nos cursos de PLE em duas instituições de ensino superior nacionais (FLUL e NOVA FCSH). No total foram analisados vinte e cinco textos por língua (cinco textos por nível em cada língua), correspondendo a um total de 20.178 palavras.

As competências e dificuldades dos aprendentes foram codificadas manualmente numa análise em dois passos que consistiu:

i) na identificação de desvios e de usos convergentes com as estruturas-alvo, nos domínios de ortografia, morfossintaxe e vocabulário, tendo sido criado um *corpus* para cada uma das componentes referidas.

ii) na categorização, quer dos desvios, quer dos usos convergentes com as estruturas-alvo, sobretudo na morfossintaxe. Os critérios que presidiram à definição das estruturas-alvo têm como base o conteúdo curricular dos documentos orientadores mencionados no ponto 1 (*Referencial Camões PLE*) e dos programas dos cursos do Instituto de Cultura e Língua Portuguesa (ICLP-FLUL) e do Curso de Língua e Cultura Portuguesa (CLCP-NOVA FCSH). Estes programas foram utilizados como “bússolas” de expectativas de conteúdos programáticos.



Importa sublinhar que estes resultados têm como fonte um *corpus* de textos, resultantes da realização de tarefas de teste bem definidas, que limitam, evidentemente, a produção de estruturas gramaticais que não sejam de esperar pelas características do tipo de texto pretendido e pela limitação do número de palavras, estando o léxico também dependente do tema em enunciado. Apesar disso, estes dados permitiram-nos construir um perfil parcial de aprendente, com base também nos documentos orientadores consultados. As afirmações aqui feitas sobre os acertos e os desvios dos aprendentes devem, portanto, ser interpretadas como descrevendo um conhecimento gramatical parcial ou específico, uma vez que a análise é limitada a um *corpus* relativamente pequeno e com tarefas muito específicas. O facto de os textos terem sido produzidos em contexto de exame poderá ter afetado o uso de algumas formas e estruturas. Os textos analisados serviram-nos, como já foi dito neste trabalho, para informar a construção de itens e perceber possíveis efeitos de teto e de chão.

A análise de *corpora* resultou na construção de um quadro com um conjunto de estruturas gramaticais que, de acordo com os dados, parecem ser relevantes para avaliar e distinguir aprendentes de PLE. A Tabela 1 resume os resultados quantitativos obtidos na análise de *corpora*, que aqui não são mostrados em detalhe por limitações de espaço e de escopo³, mas que são traduzidos pelas etiquetas qualitativas *Usa Bem* e *Usa Mal*. Estas etiquetas devem ser entendidas como um reflexo das taxas de acerto e das taxas de desvio obtidas em cada uma das construções linguísticas observadas e têm o objetivo de traçar um perfil-resumo da progressão das aprendizagens, bem como o perfil do aprendente em cada nível.

	A1	A2	B1	B2	C1
Vocabulário					
Correspondência grafema-som					
Nasalidade					
Acentos gráficos					
Preposições e conectores					
Concordância nominal					
Determinantes e pronomes					
Conjunções					
Propriedades de seleção verbal					
Concordância negativa					
Estruturas passivas					
Estruturas relativas					
Morfologia Verbal					
Concordância sujeito-verbo					
Complementadores					

Tabela 1 - Perfil do aprendente de PLE

Usa bem ■
Usa mal ■

³ Os dados quantitativos detalhados foram apresentados na *2018 BAAL Annual Conference – Taking Risks in Applied Linguistics*, podendo os autores fornecê-los mediante pedido.



Nas secções que se seguem, apresentaremos uma descrição mais completa das competências linguísticas identificadas nos *corpora* analisados.

2.1.1. O perfil do aprendente de A1

Uma análise mais detalhada dos resultados para as línguas consideradas e para cada um dos níveis mostra-nos que, no nível A1, todas as L1 apresentam uma maior taxa de erro em morfossintaxe do que de nos domínios do vocabulário e de ortografia – e.g. *nas todas as quartas-feiras, felizes férias, dois vezes*. Os falantes nativos de espanhol apresentam uma taxa de erro na componente morfossintática de 51.3%, a taxa de desvio mais baixa dentro das línguas nativas consideradas.

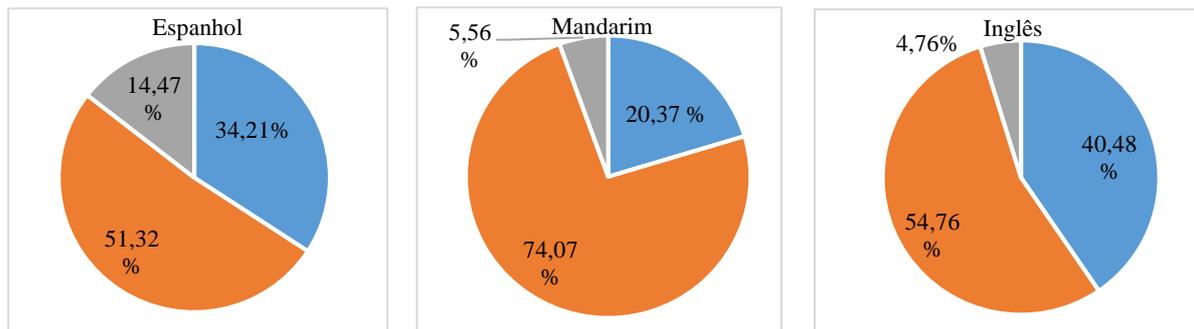
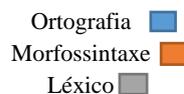


Figura 1 - Percentagem de desvios por domínio linguístico e língua materna no nível A1



A ortografia parece ser um problema maior para os nativos de espanhol e de inglês – e.g. *idea, boa, tambem* –, com taxas de erro de 34.2% e de 40.5%, respetivamente. De uma maneira geral, o vocabulário apresenta taxas de erro mais baixas quando comparadas com as taxas de erro das outras componentes. Apesar de o vocabulário de A1 ser limitado, por questões que dizem respeito ao tipo de tarefa requerido, é usado adequadamente. Existem ainda problemas na marcação gráfica da acentuação – e.g. *oculos, colegio, agua*.

No que à morfossintaxe diz respeito, os pontos fortes para os nativos de espanhol e de inglês parecem ser a concordância e a flexão nominais, estando os textos de inglês melhores no que diz respeito ao domínio da morfologia verbal. Já os falantes nativos de mandarim são melhores na morfologia e seleção verbal e no uso dos determinantes, pronomes e preposições. Os pontos fracos também mostram um contraste entre nativos de espanhol e de inglês, por um lado, e nativos de mandarim, por outro. Os primeiros têm problemas particularmente com as preposições – e.g. *tenho que ir no médico, das nove até da manhã* –, sendo que os ingleses têm problemas com os artigos. Os nativos de mandarim manifestam mais dificuldades com a concordância e a flexão nominal – e.g. *muito luz*.

Para este nível, as principais competências dos aprendentes parecem estar relacionadas com questões de correspondência entre grafema-som e com a representação da nasalidade. As maiores dificuldades parecem ser a marcação gráfica dos acentos e o uso das preposições.



2.1.2. O perfil do aprendente de A2

No nível A2, os falantes nativos de inglês e de mandarim apresentam uma maior taxa de erro de morfossintaxe (68% e 71.2%, respetivamente) do que de vocabulário e de ortografia – e.g. *vocês são o mesmo tamanho como nós, o razão, estou muito contente que recebi, estou esperar*. Por seu lado, os falantes nativos de espanhol apresentam uma maior taxa de erro na ortografia (32,6%) do que os falantes de inglês (18,6%) e de mandarim (21%) – e.g. *porqué, precisaram, som* (são). Em todas as línguas maternas analisadas, o vocabulário é a componente em que é registada menor taxa de erro, apesar de ter aumentado comparativamente à taxa de erro que se observou no nível A1, exceto para os falantes nativos de espanhol⁴.

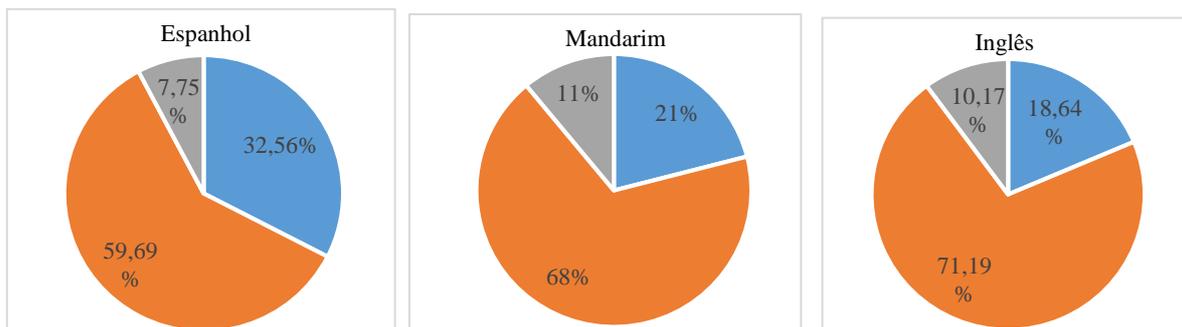


Figura 2 - Percentagem de desvios por domínio linguístico e língua materna no nível A2

Ortografia ■
Morfossintaxe ■
Léxico ■

Os pontos fortes na componente ortografia para todas as L1 analisadas parecem ser a representação da nasalidade e a correspondência grafema-som, sendo que os espanhóis têm alguns problemas com a correspondência grafema-som em vogais e ditongos – e.g. *primer*. Já os pontos fracos para os nativos de inglês e mandarim têm a ver com a marcação de acento gráfico – e.g. *aguas, Italia, ótima*.

No que à morfossintaxe diz respeito, os pontos fortes para os nativos de espanhol e de mandarim parecem ser a seleção das preposições. Os falantes de inglês parecem não apresentar grandes dificuldades nos usos de conjunções e dos conectores⁵. Os nativos de mandarim parecem ainda dominar o verbo *haber* com valor existencial, ao contrário dos nativos de inglês – e.g. *tem muitas flores aqui*⁶.

O vocabulário, apesar de limitado, por questões que dizem respeito ao tipo de tarefa que deu origem às produções textuais, parece ser usado adequadamente.

Assim, para este nível, as competências dos aprendentes parecem ser questões de correspondência grafema-som e preposições, e as principais dificuldades prendem-se com a marcação gráfica dos acentos.

⁴ Este facto pode ter a ver com o tipo de tarefa requerido em enunciado, uma pequena carta a um amigo.

⁵ Lembramos que nos referimos sempre, não às preposições e conectores em geral, mas às preposições e conectores esperados neste nível e contemplados nos documentos orientadores, nomeadamente no *Referencial Camões PLE*.

⁶ Dado o tamanho da amostra, não temos dados suficiente para excluir a hipótese de esta produção em particular ter como fundamento a exposição do candidato do exame à variedade brasileira do Português.



2.1.3. O perfil do aprendente de B1

No nível B1, os falantes nativos de inglês e de mandarim apresentam uma maior taxa de erro no domínio da morfossintaxe do que nos domínios do vocabulário e da ortografia – e.g. *Ninguém me ajudou com qualquer coisa, nunca pensei que a minha boa apetite será fazer*. No entanto, os falantes nativos de espanhol apresentam, novamente, uma taxa de erro mais alta relativamente à ortografia (47.2%) do que os falantes de inglês e de mandarim – e.g. *precisaban* -, imediatamente seguida da taxa de erro relativa à morfossintaxe (44.7%) – e.g. *não acreditar o que aconteceu depois*. Em todas as L1, o vocabulário é o domínio em que é registada menor taxa de erro, tendo esta diminuído comparativamente à registada para o nível A2.

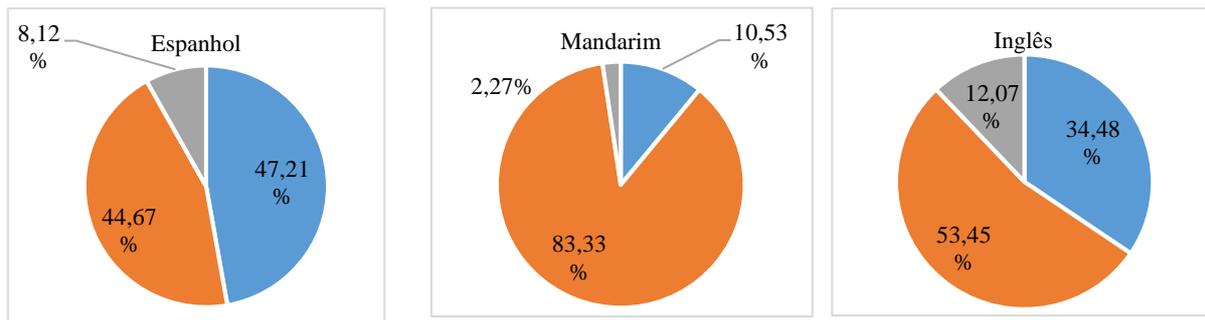
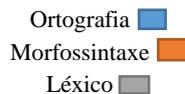


Figura 3 - Percentagem de desvios por domínio linguístico e língua materna no nível B1



Mais uma vez, os pontos fortes na componente de ortografia para todas as L1 analisadas parecem ser a representação da nasalidade e a correspondência grafema-som. A acentuação gráfica continua a ser um problema grande para os nativos de inglês e de espanhol. Os falantes nativos de mandarim parecem dominá-la, apresentando uma taxa de erro de apenas 10,5%, taxa muito baixa quando comparada com a dos nativos de inglês (34,5%) e a dos nativos de espanhol (47,2%).

Quanto à morfossintaxe, os pontos fortes para os nativos de espanhol e de inglês parecem ser a concordância nominal – e.g. *alunos antigos*. Por outro lado, os nativos de mandarim apresentam como pontos fortes o uso de determinantes, pronomes e conjunções, exibindo também um bom domínio da flexão verbal em tempo. Os pontos fracos na área da morfossintaxe são comuns aos sujeitos, falantes nativos das três línguas estudadas. Os aprendentes parecem ter dificuldades com a concordância negativa e com construções passivas – e.g. *Ninguém me ajudou com qualquer coisa, Fui cobrado 500€*.

Assim, para este nível, as principais competências dos aprendentes parecem ser questões de concordância nominal, uso de pronomes e determinantes (fenómenos intra-NP) e as principais dificuldades parecem ser a marcação gráfica dos acentos, a concordância negativa e as construções verbais compostas - *já lhe contei antes (já lhe tinha contado antes); como se passa? (como tem passado) ou com realinhamento sintático de papéis temáticos (passivas)*.

2.1.4. O perfil do aprendente de B2

Relativamente ao nível B2, a análise de *corpora* mostrou que os desvios de vocabulário são em percentagem semelhante à que se verificou nos níveis anteriores. Neste nível, verifica-se que existe, entre as



línguas maternas consideradas, uma maior dificuldade em questões de morfossintaxe – e.g. *A cidade de Maputo se desenvolve muito nestes dias; pedimos ao concelho municipal de intervir*. Os falantes nativos de espanhol apresentam uma taxa de erro na componente morfossintática de 54.5%, os falantes nativos de inglês uma de 64.3% e os falantes nativos de mandarim uma de 75.5%.

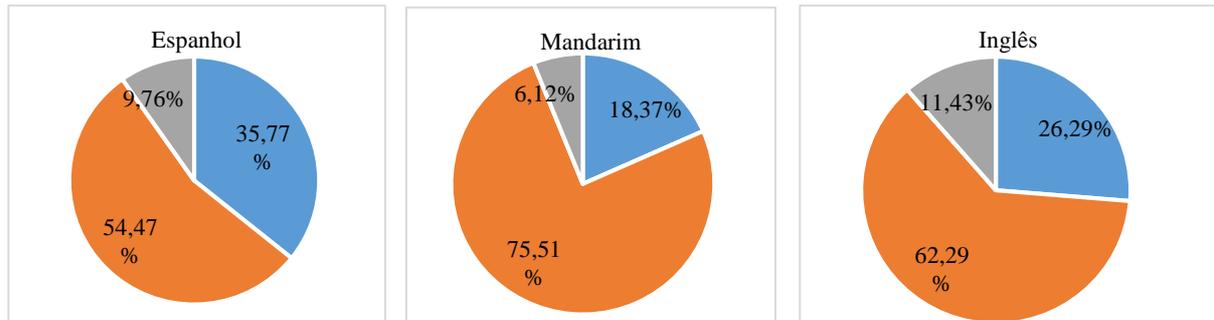
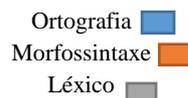


Figura 4 - Percentagem de desvios por domínio linguístico e língua materna no nível B2



Quanto à ortografia, os falantes nativos de espanhol continuam a apresentar uma taxa de desvio mais elevada (35,8%), seguidos dos falantes nativos de inglês (26,3%). Os falantes nativos de mandarim parecem ter menos problemas neste domínio, com uma taxa de erro inferior a 20%.

Quanto à morfossintaxe, os pontos fortes para os nativos de espanhol são a flexão verbal, a concordância nominal e o uso dos artigos – e.g. *o autor deste romance* (com correta explicitação do artigo, seguindo as normas de concordância). Já no que diz respeito aos falantes nativos de inglês e de mandarim, quer uns, quer outros apresentam, como pontos fortes, o uso das conjunções, conectores e preposições, além da flexão verbal e da concordância nominal. É de realçar que esses grupos de falantes não dominam o uso dos artigos tão bem como os falantes nativos de espanhol – e.g. *para comprar as frutas, a comida*⁷, (...). Por sua vez, os falantes de espanhol apresentam, como dificuldades, o uso das conjunções e dos conectores – e.g. *ruas muito sujas ainda entre a semana (mesmo durante a semana)*, exibindo inclusivamente problemas de seleção verbal – e.g. *as paredes eram sujas*. Os falantes nativos de espanhol e de inglês têm também como pontos fracos a construção de relativas ou o uso dos pronomes relativos, com taxas de erro acima dos 20% – e.g. *reunião de vizinhos na que*.

Assim, para este nível, as principais competências observadas estão relacionadas com a morfologia verbal e a concordância nominal. Já as principais dificuldades parecem ser a construção de relativas/uso de pronomes relativos, questões de seleção verbal e uso de preposições e conectores.

2.1.5. O perfil do aprendente de C1

No nível C1, os desvios de vocabulário apresentam taxas de erro muito semelhantes às observadas no nível B2. Neste nível também é possível verificar que existe, entre os falantes das línguas maternas

⁷ Embora, fora de contexto, seja possível a emergência de artigos nestas estruturas, este exemplo foi retirado de um contexto de listagem em que a omissão dos artigos é a única opção gramatical.



consideradas, uma maior dificuldade nas questões de morfossintaxe – e.g. *preocupar-nos-íamos pelos animais?*; *as imperfeições da nossa sociedade são múltiplos*. Os falantes nativos de espanhol apresentam uma taxa de erro na componente de morfossintaxe de 68.6%, os falantes nativos de inglês uma taxa de erro de 69% e os falantes nativos de mandarim uma taxa de erro de 67.4%.

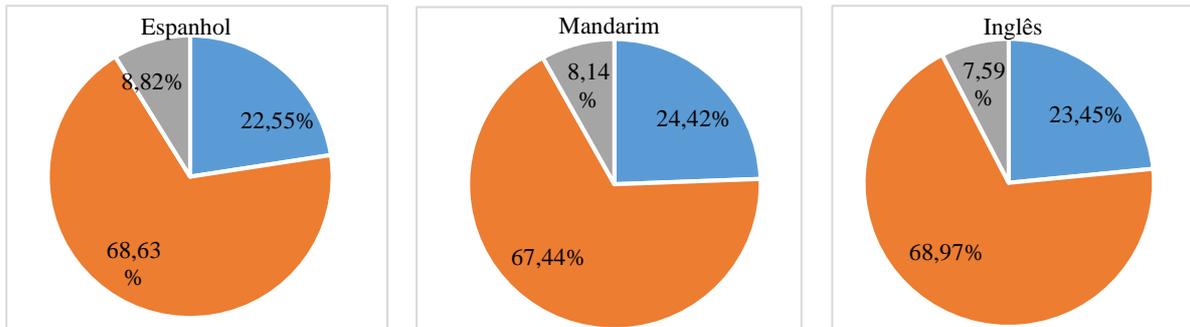


Figura 5 - Percentagem de desvios por domínio linguístico e língua materna no nível C1

Ortografia ■
 Morfossintaxe ■
 Léxico ■

Neste nível, verificamos ainda uma aproximação dos valores das taxas de erro relativas às competências ortográficas. Ao contrário do que acontece nos outros níveis analisados, estes números permitem-nos colocar a hipótese de que os desvios ortográficos dados pelos aprendentes não estão diretamente relacionados com as diferentes línguas nativas faladas. As taxas de erro situam-se entre os 20% e 25%.

No que à morfossintaxe diz respeito, os pontos fortes para os nativos de espanhol e de inglês são o uso de conjunções e de complementadores - *Espero que as minhas propostas*. Já no que aos pontos fracos diz respeito, os falantes de inglês e de espanhol parecem ter problemas com as preposições⁸ em geral e os falantes de mandarim parecem ter problemas particularmente com as preposições selecionadas pelos verbos e nomes – e.g. *nunca lhe faltou a paciência em comunicar, não se esforçam em, preocuparíamos pelos animais*.

2.1.6. Síntese

Considerando todos os dados analisados, os aprendentes de A1 parecem ter dificuldades no que diz respeito à utilização de preposições e conectores esperados para aquele nível, apresentando também dificuldades no que diz respeito à marcação gráfica do acento. Já no que diz respeito à correspondência grafema-som, à aquisição das marcas gráficas de nasalidade e à aquisição do vocabulário contemplado nos documentos orientadores, não são registados problemas de maior em nenhum dos níveis. Os aprendentes de A2 continuam a ter problemas com a acentuação gráfica, havendo uma melhoria no que ao uso das preposições esperadas para este nível diz respeito. No nível B1 parecem existir problemas com estruturas passivas e com a aquisição da concordância negativa e mantêm-se os problemas com a acentuação gráfica. No entanto, os aprendentes do nível B1 dominam a concordância nominal Nome-Adjetivo e sabem usar

⁸ Isto não implica um retrocesso na aprendizagem, sendo consequência, provavelmente, do tipo de texto requerido em exame – *carta formal às autoridades municipais em oposição à criação de um novo imposto e texto de opinião sobre a relevância de aprender línguas no estrangeiro* – e do consequente aumento da complexidade estrutural dos enunciados produzidos (lembramos que as preposições foram apontadas como um ponto forte para o grupo no nível B2).



determinantes e pronomes sem problemas de maior. As propriedades de seleção verbal só apresentam bons resultados estáveis no nível B2 e as estruturas relativas são usadas com mestria também nesta fase da aprendizagem. Os aprendentes de B2 manifestam dificuldades na morfologia verbal, nomeadamente no que diz respeito à seleção de modo e ao uso do infinitivo e ainda existem dificuldades na concordância sujeito-verbo, particularmente a longa distância ou em sujeitos recuperados. Os problemas com preposições e conectores arrastam-se até ao nível C1.

2.2. Segunda fase: pilotagem do material de compreensão escrita e de compreensão oral

A motivação para a conceção desta etapa como uma fase crucial no nosso projeto prende-se, por um lado, com a inexistência de bases de dados de materiais escritos e orais validados para a avaliação e o ensino de PLE e, por outro, com a necessidade de seguir as recomendações da ALTE - “*Materials should be commissioned for a particular test administration, or to add a new material to an item bank. [...] Test materials [...] must be checked for quality. This is done using expert judgement and by trying them out.*” (ALTE 2011:27, 28).

Como consequência da inexistência de materiais escritos e orais validados para PLE, no caso dos materiais para os níveis A1 e A2, embora conscientes de que não se tratava da situação ideal devido à importância da autenticidade, forjámos diálogos para a parte de compreensão oral, com situações do quotidiano (interação numa loja, compra de bilhetes de comboio, planeamento de uma visita a um museu). Para a compreensão escrita, houve apenas a necessidade de forjar o texto para o nível A1 (apresentação pessoal), sendo os restantes materiais autênticos com algumas adaptações (receita – A2, texto com instruções – B1, crónica de jornal – B2 e excerto de um texto literário - C1).

Após esta etapa de reunião e construção de materiais de compreensão escrita e oral, realizámos um questionário, na plataforma *Qualtrics*, com um *corpus* de textos e de material audiovisual (AV) para cada nível, para submeter a um painel de 15 peritos, constituído por professores de PLE do ICLP (ULisboa) e do CLCP (NOVA FCSH), selecionados por conveniência e contactados por *e-mail*. A tarefa dos peritos era a de ler/ouvir/visionar cada item (texto ou AV) e fazer uma correspondência entre esse item e o nível de língua mais adequado (A1, A2, B1, B2, C1 ou C2), tendo de, numa escala de *Likert* de 1 a 5, indicar o seu nível de confiança relativamente a cada resposta dada.

O questionário tinha três secções – uma parte introdutória em que eram dadas as instruções para a realização do teste e em que os professores tinham de referir a sua experiência em ensino de PLE e investigação; uma parte relativa aos itens de compreensão escrita; uma parte relativa aos itens de compreensão oral. A secção de compreensão escrita contava com 10 textos, um para cada subnível, e a secção de compreensão oral contava com 14 ficheiros AV, um para cada subnível, contando com os subníveis relativos ao nível A, cujos ficheiros audiovisuais foram manipulados para o controlo da variável *velocidade de fala*. Assim, em cada subsecção do nível A, tínhamos uma versão do ficheiro de áudio mais rápida e uma versão mais lenta. Em todas as respostas havia uma caixa de comentários em que se convidava os peritos a justificarem as suas decisões ou a fazerem comentários aos textos ou aos ficheiros AVs apresentados. Todas as questões eram de resposta única e não havia a possibilidade de voltar atrás no questionário. Os peritos tiveram como única instrução terem em consideração, para cada resposta, o seu conhecimento dos documentos orientadores (cf. *QEER*, 2001) e a sua experiência de ensino e investigação.

2.2.1. Resultados

A média de classificações para o grau de confiança na resposta foi de 3.64 o que constitui um indicador de fiabilidade das respostas dos peritos.



Em geral, parece ter havido uma boa discriminação entre os três níveis principais (A, B e C) e há uma relativa discriminação entre os diferentes subníveis (cf. Tabela 2).

Para a versão piloto, foram selecionados os materiais escolhidos pela maioria ($\geq 50\%$) dos peritos, para cada nível (cf. Tabela 2 – células preenchidas a cinza). Por exemplo, para o nível A1 foi escolhido o ficheiro AV na versão mais lenta, uma vez que 73% dos peritos o considerou mais adequado para esse nível. Para o nível A2, foi escolhido, também, o ficheiro AV na sua versão mais lenta, uma vez que 77% dos peritos o considerou adequado para esse nível.

	A1	A2	B1	B2	C1	C2	Total
AV - A1.1 (lento)	73%	20%	7%	0%	0%	0%	100%
AV - A1.1	40%	47%	13%	0%	0%	0%	100%
AV - A1.2 (lento)	68%	26%	6%	0%	0%	0%	100%
AV - A1.2	36%	50%	14%	0%	0%	0%	100%
AV - A2.1 (lento)	0%	77%	15%	8%	0%	0%	100%
AV - A2.1	0%	71%	29%	0%	0%	0%	100%
AV - A2.2 (lento)	28%	66%	6%	0%	0%	0%	100%
AV - A2.2	20%	60%	13%	0%	0%	7%	100%
Texto - A1.1	47%	47%	0%	0%	0%	6%	100%
Texto - A1.2	33%	47%	20%	0%	0%	0%	100%
Texto - A2.1	20%	60%	20%	0%	0%	0%	100%
Texto - A2.2	93%	7%	0%	0%	0%	0%	100%
AV - B1.1	0%	7%	73%	20%	0%	0%	100%
AV - B1.2	0%	0%	60%	33%	7%	0%	100%
AV - B2.1	0%	0%	40%	27%	27%	6%	100%
AV - B2.2	0%	0%	40%	27%	27%	6%	100%
Texto - B1.1	0%	13%	60%	7%	13%	7%	100%
Texto - B1.2	0%	0%	67%	27%	0%	6%	100%
Texto - B2.1	0%	0%	20%	67%	13%	0%	100%
Texto - B2.2	0%	0%	7%	47%	40%	6%	100%
AV - C1.1	0%	0%	14%	33%	33%	20%	100%
AV - C1.2	0%	0%	0%	13%	27%	60%	100%
Texto - C1.1	0%	0%	0%	21%	58%	21%	100%
Texto - C1.2	0%	0%	13%	60%	20%	7%	100%

Tabela 2 - Percentagem de correspondência entre nível previsto e nível proposto pelos peritos

Apesar de, na generalidade, ter havido facilidade em escolher o material mais adequado para testar compreensão escrita e compreensão oral, por haver percentagens $>50\%$, houve itens que suscitaram problemas. Os materiais problemáticos, destacados a negrito, dizem respeito a:

- i) um texto que foi considerado igualmente adequado para o A1 (47%) e para o A2 (47%)
- ii) dois ficheiros AV apresentados no questionário para o B2, que contêm percentagens iguais de resposta, com 27% dos peritos a acharem que esses ficheiros AV são igualmente adequados para B2 e para C1.



- iii) um ficheiro AV alinhado com o nível C1, que foi considerado pelos peritos adequado também para o nível B2 (com 33% dos peritos a considerarem-no adequado para o nível B2 e para o nível C1).

O resultado da pilotagem do material culminou na construção de uma primeira versão piloto do teste com (i) o encurtamento do ficheiro áudio que tinha sido apontado como adequado tanto para A1 como para A2 pelos peritos (cf. Tabela 2), (ii) a conversão para áudio apenas do ficheiro AV considerado igualmente adequado para B2 e para C1 em ficheiro de som (tratava-se de um ficheiro com informação multimodal em que apareciam símbolos, gráficos, números e tabelas, podendo isto dificultar o processamento e armazenamento da informação relevante) e (iii) a utilização de um ficheiro AV que tínhamos predito estar associado ao nível C1, mas que foi classificado por 33% dos peritos como correspondendo ao nível B2 e C1 (cf. Tabela 2), optando por aumentar o grau de dificuldade das perguntas.

2.3. Terceira fase: o teste piloto

A partir da análise destes materiais, foram definidas as especificações do teste. Posteriormente, procedeu-se à construção e validação dos itens de teste, que descrevemos em seguida.

A versão piloto do teste inclui três blocos, cada um deles correspondendo a cada um dos domínios linguísticos⁹, e cinco sub-blocos (subníveis de proficiência). O primeiro bloco testa o uso da língua no que ao conhecimento gramatical e ao vocabulário diz respeito; no segundo bloco, testa-se a compreensão escrita, e o terceiro bloco inclui a avaliação da compreensão oral (cf. Tabela 3 – Estrutura do teste). O piloto é composto por setenta e cinco itens de escolha múltipla/preenchimento de lacunas com quatro opções de resposta, divididos pelos três blocos (25 itens em cada bloco).

Gramática e Vocabulário	Compreensão Escrita (Leitura)	Compreensão Oral
<ul style="list-style-type: none">• A<ul style="list-style-type: none">○ 1 (5 itens)○ 2 (5 itens)• B<ul style="list-style-type: none">○ 1 (5 itens)○ 2 (5 itens)• C (5 itens)	<ul style="list-style-type: none">• A<ul style="list-style-type: none">○ 1 (5 itens)○ 2 (5 itens)• B<ul style="list-style-type: none">○ 1 (5 itens)○ 2 (5 itens)• C (5 itens)	<ul style="list-style-type: none">• A<ul style="list-style-type: none">○ 1 (5 itens)○ 2 (5 itens)• B<ul style="list-style-type: none">○ 1 (5 itens)○ 2 (5 itens)• C (5 itens)

Tabela 3 - Estrutura do teste

As instruções foram apresentadas de forma clara e sem enviesamentos de natureza cultural. Para prevenir a frustração, no caso de se tratar de um utilizador elementar ou iniciante de PE, os itens foram aleatorizados nos sub-blocos. Todos os itens eram de escolha múltipla.

No que à construção dos itens diz respeito, a equipa do projeto foi dividida em pares. Cada par tinha como tarefa fazer um conjunto de itens para todas as componentes de um dado nível. Só foram inseridos na versão piloto do teste os itens que resultaram de um consenso de opiniões dos membros da equipa do projeto.

Tal como foi dito na secção 2.1, os resultados da análise de *corpora* informaram a seleção dos conteúdos gramaticais que se pretendia testar. Por exemplo, os resultados da análise sugeriram dificuldades generalizadas (i.e., para os falantes de todas as línguas consideradas) na colocação de acentos até ao nível B2. Desta forma, construiu-se um item que pretendia avaliar a capacidade de acentuação gráfica (cf. (1)). Do

⁹ Por questões de exequibilidade, e uma vez que se trata de um teste de colocação e não de um teste de avaliação da proficiência, optou-se por não se incluir produção (escrita e oral).



mesmo modo, a Tabela 1, indicou-nos, também, efeitos de teto na morfologia verbal para o nível B2. Em termos programáticos, o nível B2 costuma integrar o ensino do modo conjuntivo, motivo por que se incluiu no teste um item que avalia o uso desse modo (cf. (2)).

(1) A _____ do Professor não me diz qual é o meu _____ de aluna.

Opções de resposta:

- (a) secretaria/numero
- (b) secretaria/número
- (c) secretária/numero
- (d) secretária/número

(2) – Então, boa viagem, Marina! Espero que te _____ muito e que _____ para descansar e passear!

Opções de resposta:

- (a) vais divertir/aproveita
- (b) divertes/aproveites
- (c) tenhas divertido/aproveitas
- (d) divirtas/aproveites

É importante referir que as estruturas gramaticais referidas na Tabela 1 não foram apenas testadas no bloco de gramática e vocabulário do teste, mas também nos blocos de compreensão escrita e oral.

2.3.1. A amostra

A versão piloto do teste foi realizada por 114 alunos dos cursos de PLE do ICLP e do CLCP, da Faculdade de Letras da Universidade de Coimbra e da Faculdade de Letras da Universidade do Porto. Houve 38 (33%) de respondentes excluídos por não terem completado o teste (17) ou por não cumprirem os tempos mínimos na reprodução dos ficheiros AV na parte de compreensão oral (21). Por esse motivo, foram incluídas, nesta análise de resultados, as respostas de apenas 76 alunos.

A versão piloto do teste contava, ainda, com uma parte em que os alunos deveriam autorreportar o seu nível de proficiência (cf. Tabela 4) e dizer a sua idade. A média de idades dos respondentes foi de 35 anos, com um desvio padrão de 16 anos.

A1 = 15%	A2 = 28%	B1 = 30%	B2 = 16%	C1 = 11%	C2 = 1%
----------	----------	----------	----------	----------	---------

Tabela 1 – Níveis de proficiência reportados pelos alunos

2.3.2. Dados sobre a realização da versão piloto do teste: validação e consistência

No que à *performance* dos itens diz respeito (cf. Figuras 6, 7 e 8), percebeu-se que houve 2 itens a apresentarem efeitos de chão (% de respostas corretas < 20) e 23 itens a apresentarem efeitos de teto (% de respostas corretas > 90). Assim sendo, apenas 50 dos 75 itens foram submetidos à Análise Fatorial Exploratória (AFE).



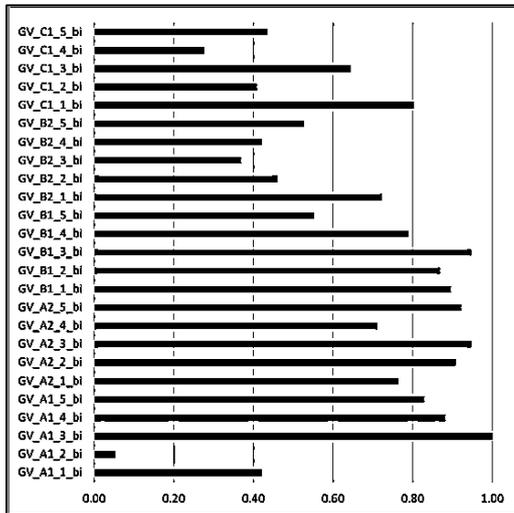


Figura 6 – Performance dos itens no bloco de Gramática e Vocabulário

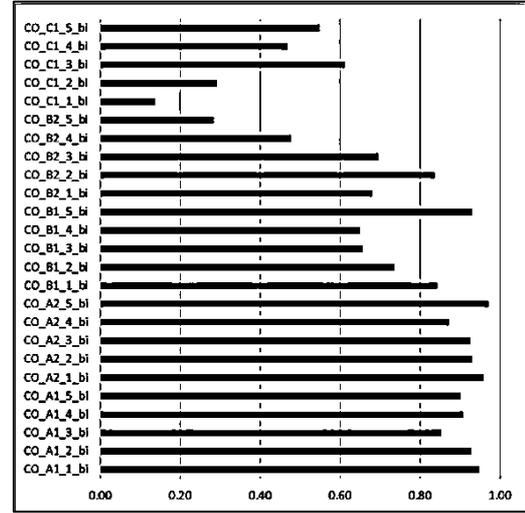


Figura 7 – Performance dos itens no bloco de Compreensão Oral

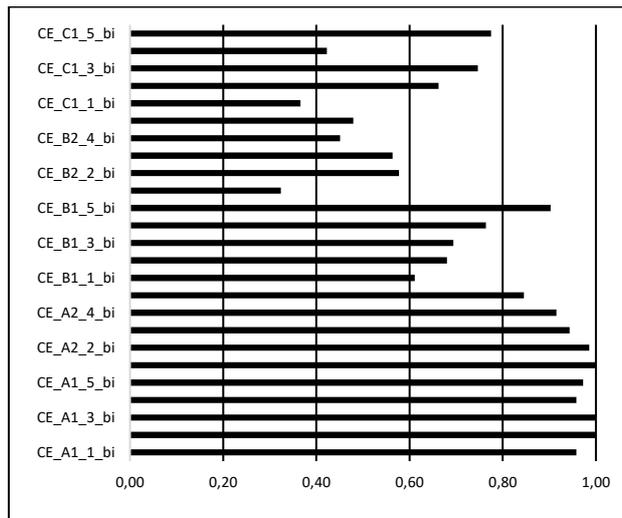


Figura 8 - Performance dos itens no bloco de Compreensão Escrita

Para testar a validade de conteúdo, corremos uma AFE. O tamanho da amostra não permitiu uma AFE única, pelo que foram corridas três AFE para cada tipo de item (gramática e vocabulário, compreensão oral e compreensão escrita), usando apenas itens para B1, B2 e C1 (respectivamente, 14, 13 e 14 itens).

Na parte da gramática e vocabulário, os resultados da AFE foram inconclusivos, devido à variabilidade na percentagem de respostas corretas em todos os níveis. Assim, prevê-se uma análise mais detalhada para este bloco, aquando de uma próxima versão do teste piloto.

Para a parte da compreensão escrita (Tabela 5), a AFE extraiu dois fatores (C1 e B2), com alguns itens de B1 sendo incluídos em C1. Estes dados sugerem que os itens de B2 e C1, na compreensão escrita, tiveram



desempenhos diferenciados, o que demonstra que estes itens, nos dois níveis, são discriminativos. No entanto, a AFE sugere que 2 itens do nível B1, na compreensão escrita, devem ser revistos.

	1	2
CE_C1_5_bi	0,69	-0,03
CE_B1_4_bi	0,53	-0,09
CE_C1_2_bi	0,51	0,12
CE_B1_3_bi	0,40	-0,04
CE_C1_1_bi	0,38	0,06
CE_C1_4_bi	0,30	0,06
CE_C1_3_bi	0,30	0,22
CE_B2_1_bi	-0,07	0,71
CE_B2_5_bi	-0,07	0,61
CE_B2_4_bi	0,24	0,34
CE_B2_2_bi	0,03	0,31
CE_B2_3_bi	0,17	0,30

Tabela 2 – Resultados da AFE para a parte de Compreensão Escrita

Na parte de compreensão oral (Tabela 6), a AFE extraiu três fatores (B1, B2 e C1), com um item de C1 sendo incluído em B2. Mais uma vez, o tamanho da amostra não permitiu mais testes na estrutura fatorial. Estes dados sugerem que muitos itens de B1, B2 e C1, na compreensão oral, tiveram desempenhos diferenciados, o que demonstra que estes itens, nos três níveis identificados, são discriminativos. No entanto, a AFE sugere que 1 item do nível B1, 2 itens do B2 e 4 itens no C1, na compreensão escrita, devem ser revistos.

	1	2	3
CO_B1_2_bi	0,71	-0,13	0,08
CO_B1_3_bi	0,58	-0,05	-0,04
CO_B1_4_bi	0,52	0,13	-0,07
CO_B1_1_bi	0,47	0,21	0,19
CO_B2_1_bi	-0,02	0,65	0,18
CO_B2_4_bi	0,03	0,62	-0,17
CO_B2_3_bi	0,24	0,42	-0,22
CO_C1_2_bi	0,09	-0,11	0,58
CO_C1_3_bi	-0,11	0,39	0,45

Tabela 3 - Resultados da AFE para a parte de Compreensão Oral

No que à consistência das respostas diz respeito (cf. Tabela 7), corremos uma análise do *alpha* de Cronbach para avaliar o tipo de item nos níveis com mais de dois itens. Os resultados (entre 0.40 *r* e 0.60 *r*) sugerem uma fiabilidade moderada. Foram corridas correções para cada tipo de item para os níveis com apenas dois itens. Os resultados (entre 0.30 e 0.60) sugerem uma correlação moderada entre os itens.



	Gramática e Vocabulário	Compreensão Oral	Compreensão Escrita
A1	-0.32	0.15 (r)	-
A2	-0.08 (r)	-	-
B1	0.27	0.68	0.52
B2	0.56	0.49	0.60
C1	0.46	0.39	0.62

Tabela 4 – Resultados da análise do *alpha* de Cronbach

3. Discussão e Conclusão

Tal como foi dito nas primeiras secções deste trabalho, as ferramentas de avaliação de diagnóstico para colocação num nível de língua apropriado são uma fonte de informação relevante para o ensino e para a investigação, uma vez que este tipo de teste nos permite medir as competências dos aprendentes.

A eficácia deste tipo de ferramentas e a relevância dos dados que da sua aplicação conseguimos retirar dependem do seu grau de validade e fiabilidade. Estes conceitos devem ser considerados para garantir a adequação e a consistência do teste. O projeto POR Nível pretende construir um teste de colocação válido e fiável para o PLE, determinando o nível de ensino de PLE em que os alunos devem ser colocados.

A análise de *corpora* foi crucial para desenhar um perfil, ainda que parcial, para o aprendente do PLE e informou-nos acerca dos conteúdos gramaticais que deviam ser testados em cada um dos níveis.

Depois da aplicação desta primeira versão piloto do teste, é necessário continuar o processo de validação do material para garantir a adequação dos itens, uma vez que os dados do teste piloto mostram que:

- i. os itens dos níveis A1 e A2 devem ser revistos e ajustados nos três construtos (gramática e vocabulário, compreensão escrita e compreensão oral);
- ii. alguns dos itens do nível B1 precisam de ajustes adicionais;
- iii. é crucial aumentar a amostra para uma análise estatística mais informativa e detalhada.

O trabalho e a pesquisa feitos no decorrer deste estudo mostraram-nos, tal como referido na literatura na área de *Language Testing and Assessment*, que a validação exige pilotagem e, frequentemente, os testes devem ser pilotados mais do que uma vez.

Comprometemo-nos, em investigação futura, a conduzir uma análise e, sempre que necessário, uma revisão individual dos itens, a substituir/ajustar os itens problemáticos, a aumentar a amostra e a fazer passar uma futura versão piloto por um painel de peritos.

Referências

- Alderson, J. C. (1996) Do corpora have a role in language assessment?. In J. A. Thomas & M. H. Short (Eds.). *Using Corpora for Language Research*. Longman, London.
- Alderson, J. C. & J., Banerjee (2002) State-of-the-Art Review. *Language teaching*, 35. 79-113.
- Alderson, J. C. (2005) *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. New York, USA: Continuum
- ALTE/Council of Europe (2011) *Manual for Language Test Development and Examination*. ALTE
- American Education Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME] (1966) *Standards for education and psychological tests and manuals*. Washington, D.C.: American Psychological Association (reviewed in 1974, 1985, 1999 e 2014).



- Association for Language Testing in Europe (2001) *Principles of Good Practices for ALTE Examinations*. ALTE.
- Berry, R. (2008) Basic Concepts of Assessment. In R. Berry (Ed.). *Assessment for Learning*. Hong Kong: Hong Kong University Press.
- Black, P. & Wiliam, D. (2006) The Reliability of Assessment. In J. Gardner (Ed.). *Assessment and Learning* (pp. 119-132). London: Sage Publishing.
- Cambridge English Language Assessment (2016) *Cambridge English Principles of Good Practice. Research and innovation in language learning and assessment. Validity, Reliability, Impact, Practicality*. Cambridge: University of Cambridge.
- Chapelle, C. (1999) Validity in language assessment. *Annual Review of Applied Linguistics*, 19, 254-272.
- Chapelle, C., Enright, M. & Jamieson, J. (2008) *Building a Validity Argument for the Test of English as a Foreign Language*. Routledge: New York and London.
- Cronbach, L. & R. Snow (1977) *Aptitudes and instructional methods*. New York: Irvington.
- Cureton, E. (1951). Validity. In E. F. Lindquist (Ed.). *Educational measurement* (pp. 621-694). Washington, D. C.: American Council on Education.
- Cushing, S. (2017) Corpus linguistics in language testing research. *Language Testing*, 34(4), pp. 441-449.
- Educational Testing Services [ETS] (2013) *Guidelines for best test development practices to ensure validity and fairness for international English language proficiency assessments*. Descarregado de https://www.ets.org/s/about/pdf/best_practices_ensure_validity_fairness_english_language_assessments.pdf.
- Egbert, J. (2017) Corpus linguistics and language testing: Navigating uncharted waters. *Language Testing*, 34(4), pp. 555-564.
- English Profile*, downloadable at <http://www.englishprofile.org/>
- Gardner, J. (2006) Assessment and Learning: an Introduction. In J. Gardner (Ed.). *Assessment and Learning* (pp. 1-6). London: Sage Publishing.
- International Test Commission [ITC] (2013) *Guidelines on Quality Control in Scoring, Test Analysis, and Reporting of Test Scores*. Downloaded from https://www.intestcom.org/files/guideline_quality_control.pdf
- Koizumi, R., Sakai, H., Ido, T., Ota, H., Hayama, M., Sato, M. & Nemoto, A. (2011) Development and Validation of a Diagnostic Grammar Test for Japanese Learners of English. *Language Assessment Quarterly*, 8(1), pp. 53-72.
- Krosnick, J. & Presser, S. (2009) Question and questionnaire design. In J. Wright & P. Marsden (Eds.). *Handbook of Survey Research*. San Diego, CA: Elsevier.
- McNamara, T. (2016) Validity in Language Testing: The Challenge of Sam Messick's Legacy, *Language Assessment Quarterly*, 3(1), pp. 31-51.
- Messick, S. (1989) Validity. In R. L. Linn (Ed.). *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education/Macmillan.
- Mondak, J. J., & Davis, B. C. (2001) Asked and answered: Knowledge levels when we will not take "don't know" for an answer. *Political Behavior*, 23(3), pp. 199-224.
- Plan Curricular del Instituto Cervantes*, downloadable at https://cvc.cervantes.es/ENSENANZA/biblioteca_ele/plan_curricular/default.htm
- Quadro Europeu Comum de Referência [CEFRL] para as Línguas: Aprendizagem e avaliação* (2001). Porto: Edições ASA/Conselho da Europa.
- Referencial Camões de PLE 2017*. Direção de Serviços de Língua e Cultura do Camões Instituto da Cooperação e da Língua (2017) Lisboa: Camões, Instituto da Cooperação e da Língua I.P.



- Référentiel de l'Alliance Française* - <https://www.cle-international.com/referentiel-de-l-alliance-francaise-pour-le-cadre-europeen-commun-livre-9782090352795.html>
- Sims, J. (2015) A Valid and Reliable English Proficiency Exam: A Model from a University Language Program in Taiwan. *English as a Global Language Education (EaGLE)*, 1(2), pp. 91-125.
- Taylor, L. & F. Barker (2008) Using corpora for language assessment. In E. Shohamy & N. H. Hornberger (eds.). *Encyclopedia of Language and Education, 2nd Ed., Vol. 7: Language Testing and Assessment*, pp. 241-254.
- Tran, H., Griffin, P. & Nguyen, C. (2010) Validating the university entrance English test to the Vietnam National University: A conceptual framework and methodology. *Procedia Social and Behavioral Sciences*, 2, pp. 1295-1304.
- Young, J., So, Y., Ockey, G. (2013) *Educational Testing System (ETS) Guidelines for Best Test Development Practices to Ensure Validity and Fairness for International English Language Proficiency Assessment*. ETS.

