



# Avaliação baseada em tarefas: subsídios para a formação continuada de professores de línguas

## Task-based assessment: subsidies for in-service language teacher education

Ana Lúcia Ferreira de MORAIS\*

Vanessa BORGES-ALMEIDA\*\*

**RESUMO:** Este artigo tem o objetivo de discutir características da avaliação baseada em tarefas em contextos de aprendizagem de língua estrangeira para identificar necessidades de letramento em avaliação de professores. O trabalho relata os resultados de uma investigação conduzida num centro de línguas público que recentemente havia adotado a metodologia de avaliação por tarefas. Fundamenta-se na literatura sobre ensino de línguas baseado em tarefas (Willis; Willis, 2007; Ellis, 2000; Janowska, 2014), avaliação de desempenho (McNamara, 1996; Coombe, 2018) e avaliação para/como aprendizagem (Black *et al.*, 2003; Brookhart, 2013; Carless, 2007; Coombe; Folse; Hubley, 2007; Earl, 2013). Participaram da pesquisa três professoras que lecionavam diferentes níveis de proficiência na escola. Os dados foram coletados por meio de observações de aulas, notas de campo, entrevistas e análise documental. Todos os instrumentos avaliativos utilizados pelas participantes foram analisados quanto a suas características, e os procedimentos empregados foram observados durante a totalidade das aulas lecionadas no período de um semestre letivo. Os resultados indicam o emprego de poucas tarefas de fato, escassez de autenticidade, e oferta de feedback essencialmente corretivo. Em decorrência disso, permitem apontar a necessidade de investimento no letramento em avaliação de professores de línguas para que possam refinar seus conhecimentos sobre tarefas e aprimorar suas práticas avaliativas, de modo a desenvolverem práticas, não apenas mais válidas e confiáveis, como também com maior potencial para orientar a aprendizagem.

**PALAVRAS-CHAVE:** Tarefa. Avaliação. Letramento em avaliação. Ensino de línguas.

**ABSTRACT:** This article has the objective of discussing characteristics of task-based assessment in foreign language learning contexts to identify teachers' assessment literacy needs. It reports the results of a case study conducted in a public language center that had recently adopted a task-based assessment methodology. It is based on the literature on task-based language teaching (Willis; Willis, 2007; Ellis, 2000; Janowska, 2014), performance assessment (McNamara, 1996; Coombe, 2018) and assessment for/as learning (Black *et al.*, 2003; Brookhart, 2013; Carless, 2007; Coombe; Folse; Hubley, 2007; Earl, 2013). Three teachers who taught different proficiency levels at school participated in the research. Data were

---

\* Mestra em Linguística Aplicada pela Universidade de Brasília. [aluciafm@gmail.com](mailto:aluciafm@gmail.com)

\*\* Doutora em Estudos Linguísticos pela Unesp São José do Rio Preto. Professora associada da UnB. [borgesalmeida@unb.br](mailto:borgesalmeida@unb.br)

collected through class observations, field notes, interviews, and document analysis. All the assessment instruments used by the participants were analyzed for their characteristics, and the procedures they employed were observed during the total amount of classes taught in the period of one school semester. Results indicate the use of very few tasks, a lack of authenticity, and the provision of essentially corrective feedback. As a result, they allow us to assert the need for investment in language teachers' assessment literacy so that they can refine their knowledge about tasks and improve their assessment practices, in order to develop practices which are, not only more valid and reliable, but also which have greater potential to promote learning.

**KEYWORDS:** Task. Assessment. Assessment literacy. Language teaching.

Artigo recebido em: 29.11.2023

Artigo aprovado em: 09.04.2024

## 1 Introdução

Cerca de cinco anos antes de esta pesquisa ser conduzida<sup>1</sup>, a direção de um centro de línguas autodeclarado comunicativo adotou uma metodologia considerada inovadora em seu contexto: decidiu que avaliariam por meio de tarefas. Embora os professores tivessem a liberdade de selecionar seus instrumentos de avaliação, era fortemente recomendado que empregassem tarefas sempre que possível. Provas tradicionais passaram a ser fortemente desencorajadas, em parte como uma das ações para alinhar a proposta pedagógica da escola aos documentos orientadores – o Currículo em Movimento da Educação Básica (Distrito Federal, 2013) e as Diretrizes Pedagógicas dos Centros Interescolares de Línguas (Distrito Federal, 2019) –, que preconizam a supremacia da função formativa da avaliação sobre a somativa. Para além disso, havia também o desejo dos professores dessa escola de solucionar um problema quanto ao que consideravam um baixo nível de desempenho oral dos aprendizes, e promover a autonomia dos estudantes.

---

<sup>1</sup> Este artigo apresenta um recorte da dissertação da primeira autora, defendida no Programa de Pós-graduação em Linguística Aplicada da Universidade de Brasília em 2022. O trabalho foi aprovado pelo Comitê de Ética em Pesquisa em Ciências Humanas e Sociais (CEP/CHS) da UnB (protocolo CAAE: 44711121.2.0000.5540).

Avaliação formativa é um termo guarda-chuva para se referir ao processo avaliativo que, de fato, é capaz de alimentar a aprendizagem e promover a autonomia do aprendiz, a partir da oferta de feedback de qualidade (Coombe; Folse; Hubley, 2007; Carless, 2007, 2015). Proposições mais recentes baseadas em dados empíricos robustos estão postas nas correntes da avaliação **para a** aprendizagem (Black *et al.*, 2003; Black; William, 2006), avaliação **como** aprendizagem (Earl, 2013) e avaliação orientada à aprendizagem (Carless, 2007, 2015), correntes que, no fundo, podem ser compreendidas como enfoques ligeiramente diferentes conforme o debate é trazido para um ou outro elemento constitutivo do processo, como os papéis de professor e aluno, as características do feedback de qualidade e das atividades avaliativas propostas, além do engajamento ativo do aluno para a promoção de sua aprendizagem e de sua autorregulação ou motivação, por exemplo (Shellekens *et al.*, 2021). Apesar de não ser recente, há uma generalizada má compreensão do conceito entre os professores brasileiros, que se observam nas contradições entre discurso e prática (Cunha, 2006).

De semelhante modo, embora tarefas sejam uma técnica já amplamente empregada no ensino de línguas em abordagens de natureza comunicativa, trata-se também de um termo polissêmico no ensino de línguas (Janowska, 2014), principalmente em língua portuguesa. Morais (2022) e Morais e Borges-Almeida (2022), por exemplo, relatam que, quando perguntados sobre o que são tarefas e sobre a sua implementação na prática pedagógica, professores desse mesmo centro de línguas demonstraram compreendê-las como atividades linguísticas autênticas e motivadoras, porém, complexas, difíceis de serem implementadas e compostas por diversas etapas que tomam considerável tempo. Mesmo em outros contextos, Barbirato e Carandina (2019, 2020) observaram conceitos equivocados ou parcialmente adequados imediatamente após um curso específico sobre tarefas na formação inicial de professores de línguas.

No nosso contexto, tarefas foram identificadas como potenciais promotoras de mudanças, a promessa de renovação, de um ensino mais autêntico e,

consequentemente, mais eficaz. Foram, entretanto, empreendidos poucos esforços no contexto para a formação dos professores antes e durante a implementação dessas mudanças na avaliação (Morais, 2022). Ora, é inegável a necessidade de promover o conhecimento sobre tarefas e o letramento em avaliação para que as ações pedagógicas e avaliativas por meio desse instrumento possam ser bem desenvolvidas e promotoras da aprendizagem (Scaramucci, 1999). Letramento em avaliação é um conceito que contempla os conhecimentos, habilidades e princípios para a condução consciente de boas práticas avaliativas e da adequada interpretação e uso de seus resultados (Inbar-Lourie, 2008; Fulcher, 2012; Quevedo-Camargo; Scaramucci, 2018).

O objetivo de identificar se havia e quais eram as necessidades de formação continuada no contexto local motivou a pesquisa maior de onde nasce este artigo. A discussão aqui presente expande as análises realizadas e apresentadas no trabalho original, e orienta-se pela seguinte pergunta de pesquisa: “quais são as características das tarefas e dos procedimentos avaliativos empregados pelos professores nesse centro de línguas e o que revelam sobre suas necessidades de letramento em avaliação?”. Seus resultados permitiram identificar necessidades de letramento e têm nos ajudado a desenhar ações pedagógicas de formação continuada para os professores no contexto local. Acreditamos que podem, igualmente, servir de base para outros contextos cujas necessidades sejam semelhantes.

Dividimos este artigo em quatro seções, além desta introdução. Como referencial teórico, discutimos os principais conceitos e características de tarefas para a avaliação em contextos de aprendizagem de línguas. Em seguida, apresentamos a metodologia do estudo. Por fim, discutimos os resultados a partir da caracterização dos instrumentos empregados e da descrição das ações avaliativas observadas.

## 2 Tarefas numa perspectiva comunicativa

Há, na literatura, duas principais perspectivas teóricas em se tratando de tarefas para ensinar e aprender língua, são elas: a **perspectiva psicolinguística** (cognitivista)

e a **sociocultural**. A perspectiva psicolinguística interessa-se pelos “[...] processos psicológicos tipicamente desenvolvidos quando alunos realizam tarefas”<sup>2</sup> (Skehan, 2003, p. 5); por isso, “[...] o *design* da tarefa é visto como potencialmente determinante das características do uso da língua e de oportunidades de aprendizagem que surgirão”<sup>3</sup> (Ellis, 2000, p. 193). Tendo isto posto, nota-se que para a perspectiva psicolinguística tipos diferentes de tarefas podem resultar em diferentes usos da língua e oportunidades de aprendizado, de modo que é possível elaborar tarefas que direcionem a usos e oportunidades predeterminadas por um programa de ensino institucional ou individual. As tarefas possuem centralidade, determinam os papéis exercidos, os contextos de realização e os tipos de interação que serão necessários.

Por sua vez, a perspectiva sociocultural, que se baseia nas premissas de Vygotsky (1896-1934), compreende a aprendizagem como sendo mediada pela interação com outras pessoas por meio de artefatos culturais (Paiva, 2014). Nela, o *design* da tarefa não ganha papel tão preponderante, pois o interesse está concentrado mais na interação entre os instrumentos e os usuários desses instrumentos:

os participantes co-constroem a “atividade” da qual participam ao realizar uma tarefa, de acordo com sua própria sócio-história e com objetivos determinados localmente, portanto é difícil fazer previsões confiáveis sobre as características do uso da língua e as oportunidades de ensino que surgirão<sup>4</sup> (Ellis, 2000, p. 193).

Nossa perspectiva é dialética: alinhamos a visão de linguagem a partir da teoria sociocultural à área de avaliação em contextos de línguas, para afirmar que o desenho do instrumento avaliativo – e, portanto, a atenção ao seu *design* – exerce influência

---

<sup>2</sup> No original: [...] on the psychological processes typically engaged in when learners do tasks.

<sup>3</sup> No original: [...] the design of a task is seen as potentially determining the kind of language use and opportunities for learning that arise.

<sup>4</sup> No original: [...] participants co-construct the ‘activity’ they engage in when performing a task, in accordance with their own socio-history and locally determined goals, and that, therefore, it is difficult to make reliable predictions regarding the kinds of language use and opportunities for learning that will arise.

sobre o resultado dessa interação, e, portanto, deve estar alinhado ao construto que se pretende avaliar. Doutra sorte, as conclusões e decisões que se tomem com base em tais conclusões ficam desprovidas de sustentação, podendo ser pouco eficazes ou, até mesmo, completamente equivocadas. Nesse sentido, é preciso considerar tanto o conceito de tarefa e suas características quanto o modo como são operacionalizadas em contexto de avaliação.

No que concerne ao conceito, embora diversos autores (por ex., Bachman; Palmer, 1996; Conselho da Europa, 2001; Janowska, 2014; Shekan, 2003) tenham definido tarefas de modo relativamente semelhante, gostamos da definição de Ellis (2003, p. 16, grifos nossos), por entendê-la como mais completa:

*Uma tarefa é um plano de trabalho que exige que os alunos processem a linguagem pragmaticamente para alcançar um resultado que pode ser avaliado em termos de se o conteúdo proposicional correto ou apropriado foi transmitido. Uma tarefa tem como objetivo resultar no uso de uma linguagem que tenha uma semelhança, direta ou indireta, à maneira como a linguagem é usada no mundo real. Como outras atividades de linguagem, uma tarefa pode envolver habilidades produtivas ou receptivas, orais ou escritas e vários processos cognitivos (Ellis, 2003, p. 16)<sup>5</sup>.*

Assim, a primeira característica de uma tarefa é a promoção do **uso significativo da língua alvo**, de modo que envolve processamento e produção de língua para a produção de sentidos – dimensão comunicativa da linguagem – e é significativa para os envolvidos, colocando os aprendizes no centro do processo e proporcionando maior motivação para seu engajamento ativo. Está, portanto, relacionada às proposições da abordagem comunicativa e todas as abordagens mais atuais a ela alinhadas (por ex., abordagem baseada em projetos, abordagem acional),

---

<sup>5</sup> No original: A task is a workplan that requires learners to process language pragmatically in order to achieve an outcome that can be evaluated in terms of whether the correct or appropriate propositional content has been conveyed. A task is intended to result in language use that bears a resemblance, direct or indirect, to the way language is used in the real world. Like other language activities, a task can engage productive or receptive, and oral or written skills and also various cognitive processes.

as quais se afastam de visões de língua como sistema. Assim, tarefas diferenciam-se de exercícios de manipulação da forma linguística à medida que priorizam os sentidos gerados a partir da forma (Long, 1997).

Além disso, ainda quando comparadas aos exercícios, tarefas apresentam mais alto grau de **autenticidade**, construto que está relacionado com a ideia de conexão entre as situações reais de comunicação e as atividades em contexto simulado, como no ambiente da sala de aula (Ellis, 2003; Long, 2015; Shekan; 1998; Willis, 1996). Assim como na vida real, elas oportunizam a interação e comunicação espontânea, e a negociação de sentidos.

Por fim, tarefas envolvem uma ação **orientada por e para uma meta/objetivo**. Portanto, delas resultam **produtos**, gerados a partir da interação linguageira que os aprendizes estabelecem com outros usuários da língua e textos orais ou escritos. O cumprimento dessa meta é prioridade e o sucesso pode ser avaliado a partir das amostras de uso da língua, que podem ser (a) produtos indiretos, gerados durante a interação para o cumprimento dos objetivos – como numa discussão em grupo para a tomada de uma decisão – ou (b) produtos diretos, que carregam em si mesmos o cumprimento da tarefa – como na produção de um *folder* de divulgação de um curso. É a qualidade desses produtos linguísticos que dá pistas e permite fazer inferências sobre a aprendizagem.

Tarefas podem se concretizar como diferentes tipos de atividades, que podem ser agrupadas a partir de suas características, objetivos e processos psicolinguísticos que demandam. A tipologia de tarefas de Willis e Willis (2007) agrupa tarefas em diferentes categorias: tarefas de listar, de ordenar e classificar, de buscar correspondência, de comparar, de compartilhar experiências, de resolver problemas e de desenvolver projetos e atividades criativas.

Há quatro dimensões relativas ao design de tarefas: (a) aberta ou fechada; (b) uni ou bi/multidirecional; (c) focada ou não focada; e (d) alvo/da vida real ou pedagógica (Willis, 1996; Leaver; Willis, 2004). **Tarefas fechadas** são bastante

estruturadas e permitem alta previsibilidade quanto aos recursos linguísticos que os aprendizes utilizarão para cumpri-las (Leaver; Willis, 2004, p. 24). Apesar disso, continuam diferenciando-se dos exercícios por terem foco principalmente no significado; assim, não são corrigidas com base num gabarito para identificar acertos e erros, mas podem ser corrigidas a partir de um rol de possibilidades. Do outro lado, estão as **tarefas abertas**, que permitem aos aprendizes maior controle tanto sobre os recursos linguísticos quanto sobre o tópico (Bygate, 1987 *apud* Leaver; Willis, 2004), o que requer ferramentas de correção qualitativas, que discutiremos mais adiante.

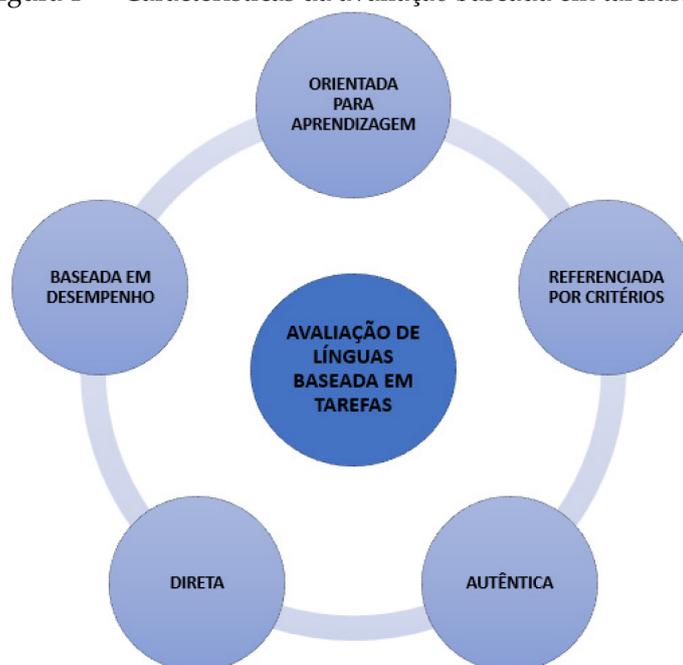
A **direcionalidade** refere-se ao fluxo da informação durante a realização da tarefa, e está relacionada ao padrão interacional proposto no *design*. Nas tarefas unidirecionais, a informação é controlada por um dos participantes, enquanto as tarefas bi/multidirecionais implicam interação entre duas ou mais pessoas, de maneira recíproca, em que todos possuem os mesmos direitos à participação (Leaven; Willis, 2004; Pica; Kanagy; Falodun, 1993). Tarefas que têm um eixo linguístico preestabelecido e explícito são denominadas **focadas**. Em oposição a elas, as tarefas não focadas não solicitam o uso de recursos linguísticos pré-determinados.

Por fim, **tarefas alvo** são aquelas que espelham consideravelmente as tarefas do “mundo real” (Long; Crooks, 1992 *apud* Leaver; Willis, 2004; Long, 2015; Nunan, 2004). Por terem um elevado grau de autenticidade, são chamadas de “tarefas autênticas” por alguns autores (Janowska, 2014). Por sua vez, tarefas pedagógicas são atividades segmentadas de modo a promover o ensaio das habilidades necessárias para realização da tarefa alvo em questão. Tarefas alvo são consideravelmente mais complexas do que tarefas pedagógicas, tanto do ponto de vista psicolinguístico quanto social (Janowska, 2014, p. 70). Apesar disso, conforme aponta Nunan (2004), toda atividade de linguagem trazida para a sala de aula com o propósito de aprendizagem passa a ser considerada pedagógica devido a sua finalidade. Ou seja, mesmo tarefas alvo continuam sendo uma **simulação** da vida real, a não ser em metodologias de abordagem por projetos, discussão que está fora do escopo deste artigo.

### 3 Avaliação de Línguas Baseada em Tarefas como Avaliação para a Aprendizagem

A avaliação de línguas baseada em tarefas é parte da abordagem mais ampla chamada de avaliação de desempenho, baseada no efetivo uso da língua em oposição ao conhecimento sobre a língua (Coombe, 2018). Aqui discutimos suas características em contextos de aprendizagem – e não de certificação. A figura 1 ilustra as principais características da avaliação baseada em tarefas.

Figura 1 – Características da avaliação baseada em tarefas.



Fonte: elaborada pelas autoras com base em Morais (2022, p. 67).

Avaliação de desempenho é sempre uma avaliação **direta**, o que equivale dizer que, se o objetivo é avaliar a produção escrita, solicita-se que o aprendiz produza um texto escrito, e não que resolva itens gramaticais, que responda uma sequência de perguntas abertas, ou, ainda, que insira sinais de pontuação num texto<sup>6</sup>. Assim, tarefas

---

<sup>6</sup> O conceito de texto poderá variar de acordo com a visão de linguagem que fundamenta o construto de uma avaliação. Por exemplo, numa perspectiva sociodiscursiva (abordagem acional que embasa, por exemplo, as propostas do QECRL (Conselho da Europa, 2001), não é autêntico pensarmos em texto como um produto linguístico isolado de um agir social, no qual operam de modo complexo as intencionalidades do falante, seus objetivos ao falar a um interagente específico, e os recursos

são um instrumento amplamente utilizados em avaliações de desempenho, podendo ser empregadas tanto em formatos de avaliação não-tradicional (um projeto, uma apresentação, um seminário) quanto compor partes de avaliações tradicionais, ou seja, seções em provas formais.

Como já discutido, a **autenticidade** é uma característica importante das tarefas. Mas o é ainda mais quando essas são utilizadas com propósitos avaliativos. A avaliação por meio de tarefas guarda forte relação com o uso da língua na vida real. Conquanto, avaliações de desempenho podem alinhar-se a uma perspectiva **fraca** – a tarefa é um meio de provocar um processamento e produção linguística, sendo considerada periférica a importância do conhecimento sobre o tema ou de possíveis decisões tomadas como produto da tarefa – ou forte – em que tanto o produto linguístico quanto o conhecimento são semelhantemente relevantes para os resultados da avaliação (McNamara, 1996; Wigglesworth; Frost, 2017). Se na hipótese fraca a tarefa é **um meio** para o uso da língua, na forte ela ocupa o centro da avaliação e a língua é um meio para realizá-la, sendo, portanto, esta última, exclusivamente baseada em tarefas alvo.

Outra característica da avaliação baseada em tarefas é ser **referenciada por critérios**, que balizam a qualificação da produção independentemente do resultado dos demais aprendizes<sup>7</sup>. Isso equivale dizer que na avaliação por tarefas é preciso “[...] mensurar desempenho e resultados usando um grupo predeterminado de critérios

---

linguísticos que se concretizam num gênero textual que precisa ser adequado para aqueles objetivos numa dada comunidade. Por outro lado, numa perspectiva comunicativa, porém não acional, pode-se ter uma compreensão de um texto a partir de sua organização interna, focada na articulação entre os recursos linguísticos em nível textual (coesão e coerência, organização retórica etc.). Como tarefas são instrumentos que podem ser utilizados em ambas as perspectivas, devido às limitações de espaço deste artigo e aos seus objetivos, não aprofundaremos essa discussão.

<sup>7</sup> O entendimento de avaliação referenciada por critério distingue-se da avaliação referenciada pela norma, na qual o desempenho de um aprendiz é comparado ao de outros, de modo a determinar o quanto este aprendiz está próximo ou distante da distribuição estatística dos resultados – o padrão estabelecido (Coombe, 2018), e cujo objetivo é ranquear e selecionar (Sawaki, 2016).

sobre os quais os aprendizes têm conhecimento [...]”<sup>8</sup> (Coombe, 2018, p. 16). Estes critérios servem tanto para qualificar os resultados da aprendizagem quanto para diagnosticar lacunas e promovê-la (Sawaki, 2016; Carless, 2007). Por isso mesmo, há grande alinhamento entre o uso de tarefas – principalmente de tarefas complexas – como instrumento mor na avaliação **para a** aprendizagem)<sup>9</sup>. Quando compartilhados e bem compreendidos por professor e aprendizes, esses critérios, que ajudam a estabelecer os padrões de qualidade esperados para uma determinada produção, funcionam como orientadores do processo avaliativo, aumentando a confiabilidade<sup>10</sup> na correção, estabelecendo ponto de partida comum para o feedback e oferecendo melhores informações para que o aluno possa gerenciar melhor seu próprio processo de aprendizagem, o que, tem alto poder para promover sua autonomia.

Devido a essa necessidade de qualificar as produções, toda avaliação baseada em tarefas precisa ser apoiada por **ferramentas de correção e feedback**, que geralmente apresentam-se como listas de verificação, escalas de classificação ou rubricas (Brookhart, 2013). Todas precisam manter relação direta com a tarefa e devem ser elaboradas concomitantemente. Listas de verificação são instrumentos dicotômicos (sim x não) e trazem os atributos que devem estar presentes na produção (por ex., inicia com letra maiúscula, apresenta uma justificativa), ao passo que escalas de classificação têm três ou mais níveis e a pontuação é distribuída entre todos os critérios e níveis. Estas últimas podem se apresentar como escalas de frequência ou de qualidade. Enquanto as primeiras denominam cada nível com advérbios de frequência (nunca, raramente, às vezes, sempre) e guiam o professor a identificar o quanto determinadas características estão presentes nas produções, as segundas utilizam

---

<sup>8</sup> No original: [...] measure student performance and outcomes using a fixed set of previously determined and concise written criteria which students are supposed to know [...].

<sup>9</sup> Sobre tarefas complexas, ver Carless (2007).

<sup>10</sup> Confiabilidade consiste no “grau de consistência de um instrumento ou método avaliativo. Confiabilidade do método refere-se diretamente aos examinadores ou corretores que atribuem as notas a uma resposta aberta a uma tarefa ou item. Trata-se do quanto uma avaliação é isenta de subjetividade [...]” (Borges-Almeida, 2023, p. 41).

adjetivos ou locuções adjetivas (precisa melhorar, satisfatório, muito bom, excelente). Tanto as escalas de frequência quanto as de qualidade apresentam limitações para a promoção da aprendizagem, pois seus resultados ainda são uma planificação dos impressionismos do professor, e quando apresentadas aos alunos, por serem desprovidas de informações concretas, pouco esclarecem as lacunas existentes e como essas podem ser preenchidas. Para que o feedback possa mediar a aprendizagem, o professor precisa, à medida que executa a correção, suplementar informações, produzindo anotações e comentários para cada aluno e por critério separadamente (Luoma, 2004; Brookhart, 2013). É uma ação trabalhosa e demorada.

Rubricas, por outro lado, são tipos de escalas que “[...] dividem uma atividade de acordo com seus componentes e oferecem **descrição detalhada** daquilo que constitui níveis aceitáveis ou inaceitáveis de desempenho para cada um desses componentes” (Stevens; Levi, 2005, p. 3, grifo das autoras). Por serem altamente descritivas, oferecendo estrutura para a realização da tarefa e tornando mais concretos os critérios de sucesso a partir dessas descrições das características esperadas do produto, elas têm alto potencial para orientar a avaliação de desempenho. Em vista disso, são ferramentas melhores para a mediação da aprendizagem e, após construídas, contribuem para acelerar também o trabalho de correção e subsidiar o feedback dado pelo professor (Brookhart, 2013; Panadero; Jonsson, 2013; Stevens; Levi, 2005).

#### 4 Metodologia

Esta investigação é um estudo de caso intrínseco (Stake, 1995), de natureza interpretativista, para o qual foram empregados métodos mistos. O contexto, como já mencionado, é um centro de línguas público que havia, dois anos antes do início da pesquisa, deliberadamente adotado a abordagem de avaliação por tarefas como parte do processo de avaliação formativa. Os dados analisados e discutidos neste artigo advêm da participação de três professoras de inglês que lecionavam diferentes níveis

(denominados 1B, 2A e 3B, respectivamente, pré-A1, A1 e B1, segundo o Quadro Europeu Comum), e foram coletados ao longo de um semestre letivo.

Foi conduzida análise documental dos instrumentos avaliativos empregados pelas participantes durante todo o semestre, foram observadas 10 aulas síncronas online e 37 aulas presenciais, e confeccionadas notas de campo. A tabela 1 elucida o quantitativo de aulas observadas e atividades assíncronas postadas no Ambiente Virtual de Aprendizagem (AVA) Google Classroom, uma vez que se tratou de um período com ensino híbrido devido à pandemia do Covid-19.

Tabela 1 — Observação de aulas<sup>11</sup>.

Profa.	Turma	1º bimestre			2º bimestre	
		Síncrona ( <i>on-line</i> )	Presencial	AVA	Presencial	AVA
Laís	1B	4	3	10	8	4
Cida	2A	3	4	8	8	4
Cris	3B	3	0	10	14	3
Total		10	7	28	30	11
Total de aulas síncronas online e presenciais observadas no semestre:						47
Total de atividades no AVA analisadas no semestre:						39

Fonte: adaptada de Morais (2022, p. 106).

As categorias para a análise documental foram inicialmente derivadas da literatura sobre *design* e componentes de tarefas, e alimentadas também numa abordagem de geração de categorias a partir dos dados (*data-driven approach*) por meio de triangulação por fontes e por informante (Stake, 1995).

## 5 Discussão dos resultados

A investigação revelou que a escola propõe o registro de notas separadamente por tarefa e por habilidades (produção oral, compreensão oral e produção escrita). Não há nenhuma proposta de integração entre habilidades, e, embora haja flexibilidade

<sup>11</sup> Cada aula online teve a duração de 50min, enquanto cada aula presencial foi de 1h40min. O curso todo desenvolve-se em 18 aulas por bimestre. A tabela 1 apresenta apenas aquelas observadas durante a coleta de dados.

para que o professor adapte a distribuição da pontuação, tampouco há clareza quanto a como proceder para que uma tarefa que porventura integre habilidades gere diferentes notas.

Efetivamente, observamos que, ao todo, foram utilizados **50 instrumentos** avaliativos durante o semestre letivo. Mesmo excluindo os exercícios que só precisavam ser entregues (portanto, não foram pontuados pelo índice de acertos), em média, cada professora empregou **15 instrumentos** diferentes para avaliar seus alunos durante o semestre letivo. Mesmo que este número esteja abaixo da proposta original da escola – em torno de 25 tarefas avaliativas diferentes por bimestre letivo por aluno, conforme discutido por Moraes (2022) –, ainda representa quantidade massiva de instrumentos avaliativos. Não à toa que Moraes (2022) e Moraes e Borges-Almeida (2022) observem que, nesse mesmo contexto, os participantes compreendem tarefas como uma metodologia complexa e trabalhosa. Esses resultados corroboram a discussão de Cunha (2006) de que professores (equivocadamente) compreendem que, uma vez que a avaliação para a mediação da aprendizagem tem como uma de suas características ser contínua, ela implica avaliar o tempo todo.

Durante a segunda entrevista, realizada após a observação das aulas, pedimos que as professoras indicassem quais de seus instrumentos avaliativos elas consideravam tarefas. As respostas evidenciaram conceitos equivocados, à medida que as participantes apontaram atividades que nem sempre apresentavam as características discutidas na literatura. Assim, havia uma confusão entre tarefas, exercícios e atividades de participação. Na tabela 2, apresentamos um resumo quantitativo desses resultados, e podemos observar que, embora, em média, as professoras acreditem que em torno de metade de seus instrumentos avaliativos eram tarefas de linguagem, foram identificadas apenas pouco mais de um décimo.

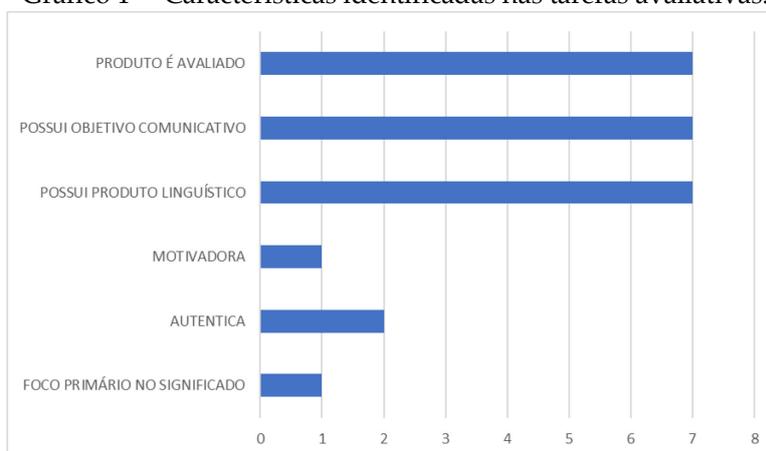
Tabela 2 — Instrumentos avaliativos evidenciando conceitos equivocados sobre tarefas.

Profa.	Total	T identificadas pela participante	%	T identificadas pelas pesquisadoras	%
Láís	16	7	43,75%	2	12,5%
Cida	14	4	28,6%	0	0%
Cris	20	20	100%	5	20%
<i>Total</i>	<i>50</i>	<i>31</i>	<i>62%</i>	<i>7</i>	<i>14%</i>

Fonte: elaborado pelas autoras.

Assim, a análise desses 50 instrumentos a partir da literatura indicou que apenas sete apresentavam características que nos permitiriam enquadrá-los como tarefas de fato<sup>12</sup>. Não obstante, a maioria dessas tarefas ainda gozava de baixo grau de **autenticidade**, e, durante sua aplicação, foram pouco motivadoras para os alunos e seu foco primário era, muito frequentemente, desviado da produção de sentidos – o significado – para a prática de formas linguísticas. As características mais observadas nas tarefas avaliativas foram: (a) serem atividades de produção, que (b) tinham um objetivo comunicativo claro, e que (c) geravam um produto linguístico (d) avaliado com base em critérios – embora essencialmente gramaticais, que não chegavam ao nível textual e muito menos ao discursivo (gráfico 1).

Gráfico 1 — Características identificadas nas tarefas avaliativas.



Fonte: elaborado pelas autoras com base em Morais (2022, p. 165).

<sup>12</sup> Quando utilizamos o termo tarefas nos referimos a esses sete instrumentos. Para fazermos alusão aos demais, mesmo que as professoras os denominem tarefas, preferimos empregar o vocábulo instrumento, um termo guarda-chuva que abarca todas as técnicas utilizadas com a finalidade de avaliação.

## 5.1 Tarefas avaliativas

Nesta seção, discutimos as principais características observadas nos sete instrumentos que foram considerados tarefas pelas pesquisadoras. Devido às limitações de espaço deste artigo, conduzimos a discussão a partir de duas tarefas que apresentam atributos observados na maioria delas.

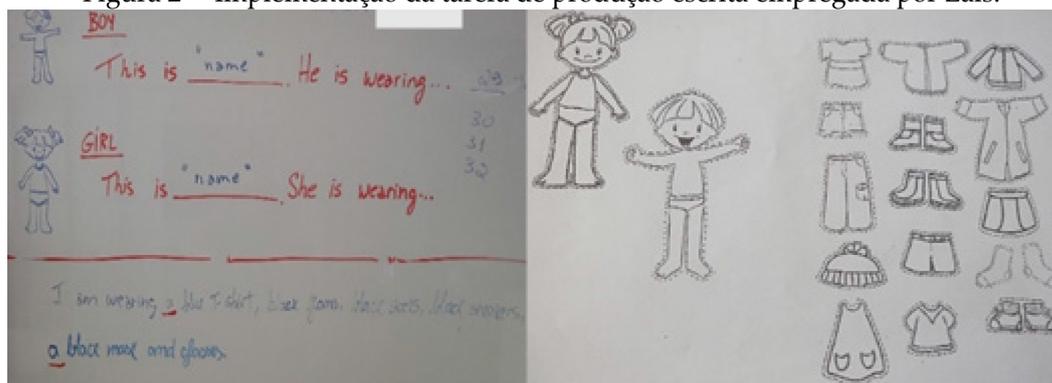
A primeira tarefa (T1.1B)<sup>13</sup>, aplicada por Laís, propunha que os alunos vestissem um personagem, a partir de uma atividade de recorte e colagem, e que descrevessem, por escrito, o que este personagem estava vestindo (figura 2). Dito desta forma, a atividade tinha um propósito comunicativo definido, parecia estar orientada para o significado, e proporcionou o engajamento dos alunos, que gostaram da atividade lúdica. Dela, também resultava um produto que poderia ser avaliado. No entanto, o que se observou na implementação dessa tarefa pela professora foi uma redução de seu potencial para construir sentidos, por ter sido conduzida de modo excessivamente controlado – embora houvesse espaço para algumas escolhas significativas. A figura 2 apresenta o quadro branco com o modelo a ser seguido (à esquerda) e o material entregue aos alunos (à direita), evidenciando que foi oferecido aos alunos um modelo de texto a ser seguido, e enfatizado o uso correto dos artigos (*she is wearing a dress*). O resultado é que a avaliação, proposta como uma avaliação de produção escrita, acabou se resumindo a um completar lacunas com vocabulário consideravelmente restrito sobre vestuário, afastando-se do conceito de avaliação de desempenho.

É de se compreender que, em níveis de proficiência mais baixos, devido ao limitado repertório linguístico dos alunos, haja maior necessidade de atividades de prática controlada, resultando em mais tarefas de tipologia fechada e muitas vezes focada, porém, neste caso, não é errado afirmar que os alunos seriam capazes de descrever de modo absolutamente mais rico as roupas de seus personagens, redigindo muito mais do que duas simples orações.

---

<sup>13</sup> A sigla adotada numera a tarefa (T1) e indica o nível em que foi empregada (1B). Esse mesmo método é adotado para a tarefa seguinte.

Figura 2 — Implementação da tarefa de produção escrita empregada por Laís.



Fonte: Morais (2022, p. 165).

Além disso, a tarefa também gerou uma segunda atividade avaliativa, que foi considerada pela professora como uma tarefa oral. Tratava-se da leitura em voz alta descrevendo o personagem, ou seja, uma oralização do pequeno “texto” anteriormente produzido. De fato, a oralização esteve presente na prática avaliativa de todas as participantes. Assim, este é um exemplo típico das características das tarefas orais utilizadas no contexto pelas três professoras, em detrimento de tarefas que requerem fala espontânea. Não podemos esquecer que uma das razões para a proposta da escola de utilizar tarefas na avaliação era promover o desenvolvimento da proficiência oral dos alunos, conforme mencionamos na seção introdutória deste artigo (ver também Morais, 2022). No entanto, tal como foram observadas as práticas em torno da avaliação de habilidades orais, há uma forte ameaça à validade de construto desses processos avaliativos (Scaramucci, 2009), pois toma-se o produto da oralização como evidência de proficiência oral espontânea.

Fica evidente que, para além dos equívocos em torno do que é uma tarefa, há questões mais fundamentais para professores de línguas que guiavam essas práticas: sua visão de língua e os conceitos de falar e escrever. Podemos dizer que oscilam entre uma compreensão de língua como uso e como sistema (visão estruturalista), em ações avaliativas que o professor “domina”, como quando se isolam variáveis (em geral, estruturas morfossintáticas) e se quantificam os erros, comumente sob a justificativa de uma suposta validade de conteúdo, pois importa avaliar “o que se ensinou”. Não à

toa é que a visão de linguagem e os conhecimentos sobre aquisição e pedagogia de línguas é componente intrínseco ao conceito de letramento em avaliação em contexto de línguas (Inbar-Lourie, 2008a, 2008b).

Por sua vez, nos níveis 2A e 3B, observou-se o uso de instrumentos que se assemelham a uma redação, ambos considerados tarefas pelas respectivas professoras regentes, (a) ora produzida em casa, baseada em pesquisa feita na internet, e lida em voz alta na sala (redação sobre o que se fez numa viagem, T3.2A), (b) ora produzida de modo tradicional como uma aplicação de prova (um texto dissertativo sobre a importância da arte para a humanidade, T5.3B), inclusive com as carteiras alinhadas em fileiras, um layout proposital para a avaliação, uma vez que no contexto as carteiras são dispostas em semicírculos.

No nível 3B, a professora também solicitou que os alunos pesquisassem sobre a vida de um ativista e a confeccionassem *slides* em *PowerPoint* e gravassem em áudio uma apresentação oral (T4.3B), que deveriam ser ambos postados no AVA *Google Classroom*, **note-se que a maioria** dos áudios postados pelos alunos apresentavam a leitura dos slides sem nenhuma ou com pouca informação adicional. De imediato, observamos que nos níveis um pouco mais altos, as tarefas são geralmente (a) abertas, (b) unidirecionais (o aluno escreve ou lê em voz alta), (c) não exigem colaboração entre pares, e (d) possuem baixa autenticidade. Em sua maioria também se apresentam como propostas de produção de textos descontextualizados, inclusive sem insumo que precise ser processado para que a tarefa possa ser cumprida.

Sobre o **insumo**, parece ser prática comum solicitar que os alunos façam uma pesquisa geral – assume-se que na internet – e escrevam sobre o que pesquisaram. Outra ação observada é apresentar, no momento da realização das tarefas ou na aula anterior, um texto padrão semelhante ao que a professora espera como produto. Via de regra, esses textos servem como modelos a serem copiados, o que observamos em todos os níveis e na prática de todas as professoras (excertos [1] e [2]).

[1] A sala está organizada em fileiras e os alunos escolheram onde se sentar, nesta aula eles terão que produzir um texto de três parágrafos sobre a importância da arte para a humanidade, os alunos foram informados que poderiam consultar dicionários, como nenhum deles trouxe de casa, Cris foi até a biblioteca da escola para buscar um para cada aluno. Antes do início da produção ela mostrou na TV um pequeno texto sobre arte e algumas citações sobre o tema de artistas e filósofos. Este texto ficou fixo na tela durante todo o período de produção do texto. Cris disse que era para inspirar-se, mas não copiar o texto mostrado. Ela então entregou uma folha impressa com as instruções de realização da produção e 27 linhas nas quais os alunos deveriam escrever seus textos, nas instruções os alunos foram orientados a dividir o texto em três partes e para cada parte Cris incluiu perguntas que deveriam orientar o conteúdo das mesmas. Os alunos terão toda a aula para realizar a produção do texto. A tarefa foi recolhida à medida que foram terminando e será corrigida em momento posterior pela professora, considerando coerência e coesão do texto, conteúdo e correção gramatical. (Nota de campo, observação de aula 3B, em 08/12/2021)

[2] Depois ela explicou que os alunos fariam uma tarefa da página 36 da apostila /.../ Ela explicou /.../ para fazer a atividade bastaria voltar às páginas anteriores e consultar a apostila. [...] ((Chamou minha atenção quando ela corrigiu quando alunos usaram somente um adjetivo, afirmando que teriam que usar dois por que no modelo tem dois, e que teriam que cumprimentar em todos os textos simulando situação de fala na escrita.)). (Nota de campo, observação de aula 1B, em 05/10/2021).

O insumo é componente importante que deve ser pensado e planejado ao avaliar. Para Willis e Willis (2007), o insumo não só contextualiza a tarefa, como oferece subsídios para seu cumprimento, e pode ser de tipos e formatos variados. Bachman e Palmer (1996) afirmam que o insumo é capaz de influenciar a resposta, e essa relação deve ser pensada e planejada tanto quanto todos os outros componentes de uma tarefa. Não obstante a sua importância, notou-se que o insumo acabava sendo pouco considerado, limitado ao estabelecimento do tema e, algumas vezes, à proposta de modelos a serem seguidos, ou seja, insumos nada autênticos ou variados, e, portanto, empobrecidos, se considerarmos a definição de Long (2015).

## 5.2 Correção e feedback

Numa perspectiva de avaliação como mediação da aprendizagem, são de fundamental importância os processos de correção e oferta de feedback aos alunos (Black *et al.*, 2003; Carless, 2007, 2015; William, 2011). Observamos considerável variabilidade no modo como as tarefas são **corrigidas**. Desde a ausência total de critérios e a simples atribuição de uma nota impressionista até o uso de grades com a nomeação dos critérios e distribuição da pontuação em níveis, configurando especialmente escalas de frequência ou de qualidade (tabela 2). Entre um e outro, há também as práticas de mencionar uma lista de critérios — geralmente morfosintáticos, fonológicos, e grafológicos — e de redigir comentários qualitativos para cada trabalho. O uso de rubricas com critérios e descritores por nível não foi identificado na prática de nossas participantes.

Tabela 2 — Exemplo de escala de frequência de avaliação escrita.

Crítérios	não	raramente	às vezes	usualmente	sim
Cumprimento da tarefa (se mantém no tema, oferece informação suficiente, usa linguagem apropriada, coesão e coerência)	0	0,2	0,4	0,6	0,8
Uso do vocabulário estudado	0	0,1	0,2	0,3	0,4
Uso correto do inglês (forma de palavra, tempo verbal, sem interferência do português)	0	0,1	0,2	0,3	0,4
Uso apropriado de construção de frases	0	0,05	0,1	0,15	0,2
Mecânica (uso correto de maiúsculas, escrita correta, pontuação, paragrafação, letra legível e escrita organizadas)	0	0,05	0,1	0,15	0,2
Comentários:					

Fonte: elaborado em inglês pela participante Cida; tradução nossa.

Entendemos que a motivação para utilizar **grades de correção** advém da consciência do professor de que a principal função da avaliação em contextos educacionais deve ser a promoção da aprendizagem. Não obstante, grades sem descritores não oferecem ou oferecem pouca informação qualitativa para os estudantes. Então, mesmo que o docente tenha construído uma imagem mental das características linguísticas típicas, digamos, do terceiro nível de qualidade numa escala impressionista de quatro níveis, para o aluno essa imagem mental inexistente. Daí, é que,

o professor que percebe essa situação, necessita envolver-se por horas na geração de comentários qualitativos para que o feedback faça sentido para o aluno e possa ter algum impacto na aprendizagem, como fez a professora Cris (excerto [3]).

[3] Somente seis alunos estavam presentes; ela devolveu as mini-tarefas escritas ((esse foi o modo como Cris chamou as atividades que entregava para os alunos)) que foram avaliadas como ACE (Avaliação contínua escrita), um total de cinco tarefas valendo 0,2 cada e nas quais ela marcou com + ou +/- e corrigiu os erros com uma caneta colorida, indicando em alguns comentários e dicas sobre o que podia ser melhorado e o que deveriam estudar. Ela também entregou o teste que fizeram valendo 2 pontos e deu a todos feedback detalhado questão a questão e a escala preenchida da tarefa de apresentação oral sobre dois mistérios, indicando os erros de pronúncia, estrutura e o que pode ser estudado e revisado para corrigir tais problemas. (Nota de campo, observação de aula 3B em 15/12/2022).

Assim, no contexto investigado, percebemos que as grades servem muito mais para a atribuição de pontuação do que como ferramenta para auxiliar na mediação da aprendizagem. A literatura aponta, de fato, que o feedback qualitativo fornecido como comentários tem alto potencial de direcionar a aprendizagem, em oposição ao feedback por notas (Black *et. al*, 2003). No entanto, toma um tempo considerável do professor quando não apoiado por outras ferramentas, como rubricas. Defendemos que, para além do uso de grades para a manutenção da confiabilidade da avaliação, há a necessidade de garantir que essas tenham atributos que permitam que de fato sejam eficazes no suporte ao fornecimento de feedback qualitativo relevante para indicar pontos fracos e fortes, e sinalizar caminhos de aprendizagem para os alunos.

Ainda com relação ao uso dessas ferramentas, observamos que, quando são utilizadas, as grades são geralmente apresentadas aos alunos antes da realização da tarefa, o que está em linha com práticas formativas. Há momentos em que, entretanto, tais critérios foram compartilhados imediatamente antes da apresentação do produto da tarefa, principalmente no caso das apresentações orais. O compartilhamento de

critérios — e, num nível ideal, sua co-construção entre alunos e professores — é uma vertente dos princípios éticos envolvidos na avaliação: é justo que os alunos compreendam como foram avaliados. Nesse sentido, compartilhar os critérios após a realização da tarefa é uma forma de atender a esses princípios (Carless, 2015). Porém, em tarefas abertas e complexas, para que tais práticas possam ter significativo impacto na aprendizagem, ainda é preciso avançar, compartilhando tais critérios **antes** do ciclo de realização da tarefa, pois a compreensão dos critérios que indicam a qualidade dos produtos orienta os alunos durante a preparação, o estudo, o engajamento e a efetiva realização da tarefa, que é o que, de fato, configura avaliação **como** aprendizagem (Earl, 2013; Schellekens *et al.*, 2021). Portanto, com essa fotografia, percebemos que as professoras têm consciência da importância do compartilhamento de critérios numa prática avaliativa formativa, porém, empregam procedimentos que nem sempre contribuem para a aprendizagem.

Outro dado que reforça o que afirmamos, é que o *feedback* oferecido é invariavelmente do tipo corretivo (ver Lyster; Saito; Sato, 2013) entre as três docentes. Por exemplo, nas avaliações escritas, Cida (2A) faz a marcação de erros, escrevendo a forma correta ao lado, e devolvendo os textos para serem “reformulados e corrigidos” pelos alunos. Entretanto, a reformulação implica somente passar os textos a limpo, incluindo, na nova versão, as formas corrigidas escritas no texto pela própria professora. Dito de outra forma, a versão final do produto escrito é editada pela própria professora. Observamos que o feedback, em geral, acaba funcionando mais como uma justificativa para as notas do que como informações para orientar o aluno (excerto [4]).

[4] Ao final Laís mais uma vez chamou de um a um para mostrar os resultados ((notas)) e dizer quais foram os erros de pronúncia e gramática. (Nota de campo, observação de aula 1B, em 07/12/2021).

Entretanto, a professora do nível 3B, preferiu realizar aulas inteiras específicas para oferta de *feedback* mais detalhado (excertos [5] e [6]). Esse retorno aos alunos aproxima-se mais do conceito de *feedback* como andaime, ou *feedforward* (Sadler, 2010), pois aponta não apenas o quê pode ser melhorado, mas como.

[5] sobre as notas, atividades que foram feitas e as que não, e no caso de alunos abaixo da média, Cris indicou tarefas avaliativas que deveriam ser feitas ou refeitas, buscando um melhor resultado. Ela deu dicas sobre como essa melhoria poderia ser alcançada. Aos alunos com resultados na média ou acima, Cris também realizou o mesmo processo, e fez observações sobre problemas com pronúncia que ela anotou das tarefas avaliativas orais. Foram atendidos sete alunos, os outros por algum motivo não puderam participar. (Nota de campo, observação de aula 3B em 11/11/2022).

[6] Somente seis alunos estavam presentes; ela devolveu as mini-tarefas escritas [...] nas quais ela marcou com + ou +/- e corrigiu os erros com uma caneta colorida, indicando em alguns comentários dicas sobre o que podia ser melhorado e o que deveriam estudar. (Nota de campo, observação de aula 3B em 15/12/2022).

Apesar disso, essas duas aulas aconteceram ao final de cada bimestre e foram uma atividade “extra” opcional; o resultado é que poucos alunos estiveram presentes. Contraditoriamente, ao mesmo tempo em que essa prática evidencia a importância que a professora atribui ao feedback, sua oferta tão tardiamente contribui para que os alunos percebam o feedback como irrelevante, o que não contribui para a autonomia e a motivação do aluno, e acaba por atrasar a aprendizagem (Black *et al.*, 2003; Carless, 2015; Hamp-Lyons, 2017). Além disso, observamos uma tensão entre a intenção de utilizar a avaliação para promover a aprendizagem e a famigerada prática da “recuperação”, como uma segunda chance para alcançar a média. Embora providas de boas intenções, essas práticas alimentam um ciclo vicioso e contribuem para mascarar práticas avaliativas essencialmente somativas, resultando, paradoxalmente, no reforço à cultura de testar, em oposição à de avaliar (Inbar-Lourie, 2008b; Shepard, 2000).

### 5.3 Atando os nós: recomendações para o letramento de professores

Diante dos resultados, foi possível identificar padrões e lacunas relacionados: (1) ao planejamento e *design* das tarefas avaliativas; (2) a seus procedimentos de aplicação; e (3) à oferta de *feedback*.

Quanto ao *design*, observamos um padrão de uso de tarefas abertas, pedagógicas, unidirecionais e focadas, e pouca variedade tipológica - em sua maioria, do tipo que envolve compartilhar conhecimentos e experiências (Willis; Willis, 2007). Sem tarefas alvo, sem tarefas que, por natureza, promovam a integração de habilidades linguísticas, e com tamanha escassez de tarefas que demandem interação e negociação de sentidos entre os aprendizes, esses processos avaliativos acabam ficando desprovidos de autenticidade. Conseqüentemente, se são os produtos dessas atividades as amostras selecionadas para serem avaliadas, eleva-se enormemente o risco de que as inferências feitas a partir delas não representem, de fato, a realidade da aprendizagem dos alunos e de sua capacidade para usar a língua, uma ameaça à validade da avaliação pela sub-representação do construto (Messick, 1989; Scaramucci, 2009). Em última instância, as decisões pedagógicas são tomadas com base em informações frágeis. Uma dessas decisões pode advir do reforço de uma percepção de que “tarefas são muito trabalhosas e não valem a pena o esforço, já que os resultados não surtem efeito”, culminando no desencorajamento de seu uso, o que entendemos que poderia ser um impacto negativo de um processo com ainda muitas lacunas sobre a cultura de avaliar e também de ensinar na escola.

Se por um lado, pelo *design* a maioria das tarefas são abertas, são também essencialmente focadas e seus **procedimentos de aplicação e correção** são demasiadamente controlados, o que proporciona um esvaziamento da tarefa, que acaba por exercer a mesma função dos itens avaliativos (avaliar o conhecimento ou capacidade de usar recursos linguísticos específicos pré-determinados pelo professor). Dessarte, as práticas avaliativas são guiadas predominantemente por uma visão de língua instrucionista e com ênfase no sistema linguístico (Leffa; Irala, 2014), refletida

no modo como a produção é altamente controlada: compreende-se produção escrita como a reprodução ou substituição de grupos nominais, *chunks* ou orações curtas (geralmente coordenadas) a partir de um modelo, e produção oral como a memorização e oralização de textos previamente redigidos.

Compreensível que o *feedback* seja muito frequentemente apenas corretivo, gerando um contexto de carência de feedback qualitativo. Ainda que exista a boa intenção, o discurso e genuínas tentativas de avaliação formativa, o que identificamos no contexto foi uma avaliação somativa fracionada, em que muitos instrumentos são empregados e as pontuações são somadas ao final (ou é calculada a média). Assim, a avaliação mostrou-se essencialmente ainda somativa e centrada na nota.

Em suma, podemos afirmar que essas ações avaliativas (a) se pautam por fragmentos de preceitos da abordagem comunicativa, filtrada pelas suas visões de língua das professoras, (b) são alimentadas por pouco conhecimento formal sobre o instrumento tarefas e (c) estão fundamentadas na compreensão incompleta do conceito de feedback. Entendemos que é necessário investir em ações de formação continuada a partir desses eixos para que possam agir de fato de modo mais alinhado com as epistemologias que fundamentam ações avaliativas formativas em contextos de línguas.

## 6 Considerações finais

Neste artigo, descrevemos as práticas avaliativas em um centro de línguas que adotou a avaliação baseada em tarefas. Reconhecemos e louvamos o interesse, o esforço, a dedicação dos professores dessa escola para promover melhores práticas pedagógicas. Esperamos verdadeiramente que tais ações frutifiquem e inspirem profissionais em outros contextos. Porém, tal mudança exige investimento, e este trabalho evidencia lacunas importantes sobre as quais tais ações de formação – de professores e, também, de aprendizes – podem e devem se debruçar.

A despeito de ter sido um estudo de caso, a confiabilidade de nossos resultados constrói-se sobre os ecos de outras pesquisas que indicam a necessidade de formação em questões semelhantes. Temos total confiança de que o único caminho é a educação continuada, com vistas ao desenvolvimento de um letramento em avaliação em contextos de línguas que contemple a des/re-construção de visões de língua como uso e o refinamento da compreensão sobre a dimensão formativa da avaliação. Além disso, é fundamental um componente formal de conhecimento sobre design, aliado a ações situadas de uso de tarefas para o ensino e a avaliação. Recomendamos fortemente ações que contemplem essas questões.

### Referências

BACHMAN, L. F.; PALMER, A. **Language testing in practice: designing and developing useful language tests**. Oxford, UK: Oxford University Press, 1996.

BARBIRATO, R. de C.; CARANDINA, M. P. Elaboração de tarefas comunicativas por professores em formação inicial: concretizando princípios, criando sentidos. **Revista Horizontes de Linguística Aplicada**, v. 18, n. 2, p. 139–166, 2019. DOI <https://doi.org/10.26512/rhla.v18i2.25147>

BARBIRATO, R. de C.; CARANDINA, M. P. Compreensões de professores acerca do termo tarefa comunicativa. **Revista Horizontes de Linguística Aplicada**, v. 19, n. 2, p. 91–117, 2020. DOI <https://doi.org/10.26512/rhla.v19i2.26694>

BLACK, P. *et al.* Putting the ideas into practice. *In: BLACK, P. et al. Assessment for learning: putting it into practice*. Nova Iorque: Open University Press, 2003. p. 30-57.

BLACK, P.; WILLIAM, D. **Assessment for learning in the classroom. Assessment and learning**. Londres: Sage Publications, 2006.

BORGES-ALMEIDA, V. Confiabilidade. *In: GONTIJO, S. B. F.; LINHARES, V. L. de C. N. (org.). Dicionário de avaliação educacional*. 1 ed. Brasília: Editora IFB, 2023. p. 41-41.

BROOKHART, S. M. **How to create and use rubrics: for formative assessment and grading**. Alexandria: ASCD, 2013. DOI <https://doi.org/10.4135/9781452218649.n15>

CARLESS, D. Learning-oriented assessment: conceptual bases and practical implications. **Innovations in Education and Teaching International**, v. 44, n. 1, 2007. DOI <https://doi.org/10.1080/14703290601081332>

CARLESS, D. Exploring learning-oriented assessment processes. **Higher Education**, v. 69, n. 6, p. 963-976, 2015. DOI <https://doi.org/10.1007/s10734-014-9816-z>

COOMBE, C. (ed.). **An A to Z of second language assessment: how language teachers understand assessment concepts**. Londres: British Council, 2018.

COOMBE, C.; FOLSE, K.; HUBLEY, N. **A practical guide to assessing English language learners**. Ann Arbor: MI: The University of Michigan Press, 2007.

CONSELHO DA EUROPA. **Common European framework of reference for languages: learning, teaching, assessment**. Cambridge: Cambridge University Press, 2001.

CUNHA, M. C. C. Nem só de conceitos vivem as transformações: equívocos em torno da avaliação formativa no ensino/ aprendizagem de línguas. **Revista Brasileira de Linguística Aplicada**, v. 6, n. 2, 2006. DOI <https://doi.org/10.1590/S1984-63982006000200004>

DISTRITO FEDERAL. Secretaria de Educação. **Currículo em movimento da Educação Básica**. Pressupostos teóricos. Brasília, 2013.

DISTRITO FEDERAL. Secretaria de Educação. **Diretrizes pedagógicas dos Centros Interescolares de Línguas (CILs)**. Brasília, 2019.

EARL, L. M. **Assessment as learning: using classroom assessment to maximize student learning**. 2 ed. Thousand Oaks: Corwin Press, 2013.

ELLIS, R. Task-based research and language pedagogy. **Language Teaching Research**, v.4, n. 3, Sage Publishers, 2000. DOI <https://doi.org/10.1177/136216880000400302>

ELLIS, R. **Task-based language teaching and learning**. Oxford: Oxford University Press, 2003.

FULCHER, G. Assessment literacy for the language classroom. **Language Assessment Quarterly**, v. 9, n. 2, 2012. DOI <https://doi.org/10.1080/15434303.2011.642041>

HAMP-LYONS, L. Language assessment literacy for language learning-oriented assessment. **Papers in Language Testing and Assessment**, v. 6, n. 1, p. 88-111, 2017. DOI <https://doi.org/10.58379/LIXL1198>

INBAR-LOURIE, O. Constructing a language assessment knowledge base: a focus on language assessment courses. **Language Testing**, v. 25, n. 3, 2008a. DOI <https://doi.org/10.1177/0265532208090158>

INBAR-LOURIE, O. Language assessment culture. *In*: SHOHAMY, E.; HORNBERGER, N. H. (ed.). **Encyclopedia of language and education**. 2 ed. v. 7, Language Testing and Assessment. Boston: Springer, 2008b. p. 285-300.

JANOWSKA, I. Tarefa: um conceito-chave da perspectiva acional. Tradução de Eduardo Nadalin e José Carlos Moreira, Universidade Federal do Paraná. **Revista X**, v. 2, 2014. DOI <http://dx.doi.org/10.5380/rvx.v2i0.2014.40330>

LEAVER, B. L.; WILLIS, J. R. **Task based instruction in foreign language education: practices and programs**. Washington: Georgetown University Press, 2004.

LEFFA, V. J.; IRALA, V. B. O ensino de outra(s) língua(s) na contemporaneidade: questões conceituais e metodológicas. *In*: LEFFA, V. J.; IRALA, V. B. (org.). **Uma espiadinha na sala de aula: ensinando línguas adicionais no Brasil**. Pelotas: EDUCAT, 2014. p. 21-48. Disponível em: [leffa.pro.br/textos/trabalhos/livro\\_espiadinha.pdf](http://leffa.pro.br/textos/trabalhos/livro_espiadinha.pdf)

LYSTER, R.; SAITO, K.; SATO, M. Oral corrective feedback in second language classrooms. **Language Teaching**, v. 46, n. 1, p. 1-40, 2013. DOI <https://doi.org/10.1017/S0261444812000365>

LONG, M. H. Focus on form in task-based language teaching. *In*: LONG, M. H. **Annual McGraw-Hill Teleconference in Second Language Teaching**, 1997.

LONG, M. **Second language acquisition and task based language teaching**. West Sussex: John Wiley and Sons, 2015.

LUOMA, S. Speaking scales. *In*: LUOMA, S. **Assessing speaking**. Nova Iorque: Cambridge University Press, 2004. p. 59-95. DOI <https://doi.org/10.1017/CBO9780511733017.005>

McNAMARA, T. **Measuring second language performance**. Londres; Nova Iorque: Longman, 1996.

MESSICK, S. Validity. *In*: LINN, R. L. (ed.). **Educational Measurement**. 3 ed. Nova Iorque: American Council on Education; Macmillan, 1989. p. 13-103.

MORAIS, A. L. F. **A tarefa de usar tarefas**: práticas avaliativas em um Centro Interescolar de Línguas público. Dissertação (Mestrado em Linguística Aplicada) – Instituto de Letras, Universidade de Brasília, Brasília, 2022.

MORAIS, A. L. F.; BORGES-ALMEIDA, V. Compreensão de professores sobre tarefas de ensino de língua estrangeira em um Centro Interescolar de Línguas público. **Revista Desempenho**, v. 1, n. 33, p. 41-67, 2022. Disponível em: <https://periodicos.unb.br/index.php/rd/article/view/44467>

NUNAN, D. **Task based language teaching**. Cambridge: Cambridge University Press, 2004. DOI <https://doi.org/10.1017/CBO9780511667336>

PAIVA, V. L. M. O. Teoria Sociocultural. *In*: PAIVA, V. L. M. de O. **Aquisição de segunda língua**. São Paulo: Parábola, 2014. cap. 8.

PANADERO, E.; JONSSON, A. The use of scoring rubrics for formative assessment purposes revisited: a review. **Educational Research Review**, v. 9, n. 0, 2013. DOI <https://doi.org/10.1016/j.edurev.2013.01.002>

PICA, T.; KANAGY, R.; FALODUN, J. Choosing and using communication tasks for second language instruction and research. *In*: CROOKES, G.; GASS, S. (ed.). **Tasks and language learning**: integrating theory and practice. Clevedon: Multilingual Matters, 1993. p. 9-34.

QUEVEDO-CAMARGO, G.; SCARAMUCCI, M. V. R. O conceito de letramento em avaliação de línguas: origem e relevância para contexto brasileiro. **Linguagem: Estudos e Pesquisas**, v. 22, n. 1, 2018. DOI <https://doi.org/10.5216/lep.v22i1.54474>

SADLER, D. R. Beyond feedback: developing student capability in complex appraisal. **Assessment & Evaluation in Higher Education**, v. 35, n. 5, p. 535–550, 2010. DOI <https://doi.org/10.1080/02602930903541015>

SAWAKI, Y. Norm-referenced vs. criterion referenced approach to assessment. *In*: TSAGARI, D.; BANERJEE, J. (ed.). **Handbook of second language assessment**. Berlin; Boston: De Gruyter Mouton, 2016. p. 45-60. DOI <https://doi.org/10.1515/9781614513827-006>

SCARAMUCCI, M. V. R. Avaliação de leitura em inglês como língua estrangeira e validade de construto. **Calidoscópico**, v. 7, n. 1, 2009. DOI <https://doi.org/10.4013/cld.2009.71.04>

SCARAMUCCI, M. Avaliação: mecanismo propulsor de mudanças no ensino/aprendizagem de língua estrangeira. **Revista Contexturas**, v. 4, p. 75-81, 1999.

SCHELLEKENS, L. H. *et al.* A scoping review on the notions of Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL). **Studies in Educational Evaluation**, v. 71, 1010094, 2021. DOI <https://doi.org/10.1016/j.stueduc.2021.101094>

SHEPARD, L. A. The role of assessment in a learning culture. **Educational Researcher**, v. 29, n. 7, p. 4-14, 2000. DOI <https://doi.org/10.3102/0013189X029007004>

SKEHAN, P. **A cognitive approach to language learning**. Oxford: Oxford University Press, 1998. DOI <https://doi.org/10.1177/003368829802900209>

SKEHAN, P. Task-based instruction. **Language Teaching**, v. 36, p. 1-14, 2003. DOI <https://doi.org/10.1017/S026144480200188X>

STAKE, R. E. **The art of case study**. Thousand Oaks: SAGE Publications, 1995.

STEVENS, D. D.; LEVI, A. **Introduction to rubrics: an assessment tool to save grading time, convey effective feedback, and promote student learning**. Canada: Stylus, 2005.

WIGGLESWORTH, G.; FROST, K. Task and performance-based assessment. *In*: SHOHAMY, E.; IAIR, G. O.; MAY, S. (ed.). **Language Testing and Assessment**. 3. ed. Cham: Springer Cham, 2017. DOI [https://doi.org/10.1007/978-3-319-02261-1\\_8](https://doi.org/10.1007/978-3-319-02261-1_8)

WILLIAM, D. What is assessment for learning? **Studies in Educational Evaluation**, v. 37, n. 1, p. 3-14, 2011. DOI <https://doi.org/10.1016/j.stueduc.2011.03.001>

WILLIS, D.; WILLIS, J. **Doing task-based teaching**. Nova Iorque: Oxford University Press, 2007.

WILLIS, J. A flexible framework for task-based learning: an overview of a task-based framework for language teaching. *In*: WILLIS, D.; WILLIS, J. (ed.). **Consciousness-raising activities**. London: Macmillan, 1996.