Assessing speaking through multimodal oral presentations: The case of construct underrepresentation in EAP contexts

Louise Palmour [ID]

University of Southampton, UK; University of Glasgow, UK

Abstract

This article explores the nature of the construct underlying classroom-based English for academic purpose (EAP) oral presentation assessments, which are used, in part, to determine admission to programmes of study at UK universities. Through analysis of qualitative data (from questionnaires, interviews, rating discussions, and fieldnotes), the article highlights how, in EAP settings, there is a tendency for the rating criteria and EAP teacher assessors to sometimes focus too narrowly on particular spoken linguistic aspects of oral presentations. This is in spite of student assessees drawing on, and teacher assessors valuing, the multimodal communicative affordances available in oral presentation performances. To better avoid such construct underrepresentation, oral presentation tasks should be acknowledged and represented in rating scales, teacher assessor decision-making, and training in EAP contexts.

Keywords

Construct validity, English for academic purposes, multimodality, oral presentation, qualitative research

Within English-medium degree programmes at universities, oral presentations commonly feature as learning and assessment tools (Huxham et al., 2012). English for academic purposes (EAP) instruction seeks to prepare students for English-medium university study and, therefore, endeavours to mirror elements of the target language use domain (Bachman & Palmer, 1996). Due to such emphasis on authenticity in EAP learning and assessment (Harwood & Petrić, 2011; Hyland & Shaw, 2016), academic oral

Corresponding author:
Louise Palmour, School of Modern Languages & Linguistics, University of Southampton, Southampton SO17 1BJ, UK.
Email: Louise.palmour@glasgow.ac.uk

presentation (AOP) assessments form part of many EAP courses' assessment suites. Compared with research conducted on large-scale EAP tests (e.g., International English Language Testing Service – IELTS; Test of English as a Foreign Language – TOEFL), there has been a dearth of studies investigating what is being assessed in in-house, classroom-based, EAP assessments (Schmitt & Hamp-Lyons, 2015).

This article provides closer understandings of the official and unofficial constructs underlying EAP summative oral presentation assessments. This is achieved by capturing student and teacher decision-making processes linked to EAP AOP assessment events via questionnaires, interviews, and fieldwork. The research draws on Macqueen's (2022) distinction between theoretical, stated, perceived, and operationalized constructs. Macqueen's framework is used to compare theories of the abilities being assessed (theoretical constructs), rating scales (stated constructs), the views (perceived constructs), and practices (operationalized constructs) of assessors and assessees. The study sheds light on theoretical, stated, perceived, and operationalized constructs from multiple stakeholder perspectives in EAP settings, in order to facilitate higher levels of construct validity and fairness (Macqueen, 2022) in future assessments.

Literature review

After outlining the socially oriented approach to the exploration of assessment constructs adopted in this study, the literature review discusses theories of strong and weak senses of performance assessments and their application to English for Specific Purposes (ESP) contexts. Then, the concept of multimodality and its implementation in language assessment is discussed. Finally, the case is made for the multi-perspectival qualitative approach taken in this article to investigate EAP assessment constructs.

Socially oriented approach to construct dimensions

Critical reconceptualisations of language in the social and multilingual turns in applied linguistics have prompted calls for expanded constructs in language assessments which reflect the sociolinguistic reality in specific communicative contexts (McNamara & Roever, 2006; Shohamy, 2011). In ESP testing, it has been argued that the assessment construct must reflect domain-specific content and communicative and professional competences (Douglas, 2000; Knoch & Macqueen, 2019) and that such core features of the domain be reflected in rating scales (Messick, 1995). Investigating ESP assessment contexts and target domains in a manner which is sensitive to social action is required to contribute to understandings of theoretical, stated, perceived, and operationalized assessment constructs currently espoused.

A socially oriented approach views assessment constructs as fluid and dynamic, and often elusive, because people's conceptualisations and practices vary, are nuanced (Jamieson, 2013), and are subject to shifts. This is because the construct may be thought of and presented differently depending on the sphere in which a stakeholder (e.g., teacher assessor, student assessee) operates at given moments in time. Macqueen (2022) proposed conceptualising assessment construct activity as "spheres of activity," addressing the situated and multiperspectival nature of assessment events. Spheres of activity

include "curriculum documents," "test specifications," "actual performance," and "stakeholder discourse" among others (p. 242). Furthermore, to facilitate the exploration of dynamic assessment constructs across these spheres of activity, Macqueen outlined a framework consisting of four construct dimensions: theoretical, stated, perceived, and operationalized. Theoretical construct denotes the ability assessed as theorised in literature or from experience. A stated construct is that which is communicated in official discourse, such as in curriculum documents and rating scales. Perceived constructs are the perceptions and interpretations of the assessment construct, for instance assessor and assessee interpretations of rating scales. Finally, the operationalized construct refers to aspects such as the actual assessees' performance and assessors' rating practices. Operationalized constructs constitute construct activity, which is put into practice by eliciting samples of, and making judgements on, observable performance, often through the use of a task, such as an AOP. Operationalized constructs are particularly complex because, as Macqueen noted, context is embedded in this activity. In fact, the four construct dimensions and the spheres of activity themselves are "overlapping and messy" (p. 243). Macqueen's construct dimension categories, nonetheless, prove useful in drawing out the nature of construct activity in EAP contexts.

Strong and weak senses of performance assessment

EAP encompasses a focus on discourse and academic literacy development (Hyland, 2018). For instance, EAP syllabi often focus on aspects such as delivering academic presentations and writing persuasively rather than on grammar and pronunciation. The expanded notion of EAP as academic literacy development may be realised to different degrees in theoretical, stated, perceived, and operationalized assessment constructs in particular EAP assessments.

The EAP AOP assessments under investigation in this study are second language (L2) performance assessments (McNamara, 1996). The assessments involve L2 users of English engaging in a task that taps into both linguistic and non-linguistic skills, rather than abstract demonstration of linguistic knowledge. McNamara provided a conceptual distinction between strong and weak senses of performance assessment, which is useful when ascertaining the role and treatment of linguistic and non-linguistic aspects. Weak language performance assessment is conceptualised as having criteria that are heavily linguistic in focus, whereas the stronger sense of the term performance assessment includes or focuses on aspects of task fulfilment, with language forming only part or none of the criteria used to make judgements of performance (McNamara). In strong performance assessments, language ability is viewed not so much as an object of assessment, but rather as the vehicle for task performance, with performance being judged according to overall effectiveness.

In ESP assessment, empirical investigations have been undertaken, ascertaining the ability of particular ESP assessments to encompass what matters in the target language use domain. A strong view of performance assessment has been advocated in a number of ESP spheres. For example, Kim (2018) underscored the vital role of aspects of professional knowledge and behaviour in ascertaining what constitutes successful communication in radiotelephony communication in aviation settings. In a health care

context, O'Hagan et al. (2016) demonstrated that raters could be trained to implement an expanded speaking construct, which included professionally relevant criteria of "clinician engagement" and "management of interaction," in an occupational English test.

Similarly, criticisms of construct underrepresentation have been levelled at another branch of ESP: EAP assessment. Khabbazbashi et al. (2022) argued that EAP assessment tasks do not reflect contemporary multimodal tasks and literacies required in higher education. Effective communication and knowledge sharing in academic settings is shaped by a multitude of processes and practices at the epistemological and literacy rather than purely linguistic level (Lea & Street, 1998). Thus, AOP assessments used, in part, to determine transition onto English-medium degree programmes are strong candidates for the stronger sense of performance assessment. Oral presentations involve a range of competencies in EAP settings and in the target domains on disciplinary modules. In the context of an EAP course in Malaysia, Januin and Stephen (2015) reported three types of discourse competence: public speaking, structure, and linguistic knowledge. On disciplinary modules, Heron's (2019) study highlighted how content (e.g., research skills, organisation, and coherence) and delivery (e.g., responses to questions, visuals, timing, and audience engagement) were emphasised, albeit to different degrees, on two business studies disciplinary modules. This study builds on the existing research base by investigating the constructs underlying EAP AOP assessments and probing claims that the multimodality of academic discourses is underrepresented.

The terms linguistic and non-linguistic, and strong and weak senses of performance assessment, facilitate understandings of EAP assessment constructs in this study. These categories serve to indicate the extent to which particular assessment constructs target academic language and/or academic literacy. This notwithstanding, the categories of linguistic and non-linguistic are too broad to capture and analyse the complex character of EAP AOP assessment constructs. The AOPs investigated in the current study all involve using gestures, slides, and speech, resulting in a complex interplay between linguistic and non-linguistic skills (e.g., text and images on slides accompanying speech in real time). Teachers and students in the EAP contexts featured in this research referred to a range of abilities involved in the AOP assessments on their programmes. These included physical skills (e.g., clarity in pronunciation, gestures, eye contact), linguistic skills (e.g., use of grammar, variety of vocabulary, appropriate register), cognitive skills (e.g., content, self-regulation), socio-emotional skills (e.g., confidence, responding to questions, ability to integrate feedback), and the use of multimedia skills (e.g., use of images, font size, ability to digest slides, interplay between speech and gestures). Therefore, an adapted version of an Oracy Skills Framework (Mercer et al., 2017) was used to capture the multiple modes available in the AOP tasks, shown to be important to the assessment construct activity. Table 1 shows the adapted version of the Oracy Skills Framework.

Multimodality

Multimodality "describes approaches that understand communication and representation to be more than about language, and which attend to the full range of communicational forms people use including image, gesture, gaze, posture and so on—and the relationships

Table 1. Adapted oracy skills framework (Mercer et al., 2017).

Physical	Linguistic	Cognitive	Social and emotional	Use of multimedia (added section)
Voice (pace, tonal variation, clarity of pronunciation, and voice projection) Body language (gestures and postures, facial expression, and eye contact)	Vocabulary (appropriate vocabulary choice) Language (register, grammar) Rhetorical techniques (e.g. metaphor, humour, irony, and mimicry)	Content (choice of content to convey meaning and intention, building on the views of others) Structure (structure and organisation of talk) Clarifying and summarising (seeking information and clarification through questions/ing, summarising) Self-regulation (maintaining focus on task and time management) Reasoning (giving reasons to support views, critically examining ideas and views expressed)	Working with others (guiding or managing interactions, and turn taking) Listening and responding (listening actively and responding appropriately) Confidence in speaking (Self assurance, liveliness, and flair) Audience awareness (Taking account of level of understanding of the audience)	Use of text, images, graphics, videos (legible and engaging, interplay between multimedia and other aspects supports communication of main message)

between these" (Jewitt, 2017, p. 15). Kress (2017) included layout, writing, and moving images in his definition so as not to exclude multimedia and technology. EAP AOP assessments, with required use of slides, entail delivering a communicative performance that taps into a complex, fluid arrangement of modes.

Progress on including paralinguistic and non-linguistic facets of communication in speaking tests has been notably slow (Plough, 2021). Recent studies that call for an expanded speaking construct focus on eliciting interactional competence in speaking tasks such as interviews and roleplays (Salaberry & Burch, 2021; Youn & Chen, 2021). There is a dearth of research on how linguistic performance is treated in reference to the communicative affordances available in oral presentation tasks. Research conducted on AOPs has included studies addressing the role of multimedia in improving performance (Rowley-Jolivet, 2002), but such research is rarely undertaken in assessment contexts. Assessments can enhance construct validity by eliciting dynamic languaging practices and "*multimodal* and *multisensorial* assemblances" (Shohamy & Pennycook, 2019, p. 36), which recognise resources such as images and movement. Extracting spoken production from other semiotic resources is not only artificially restrictive but raises issues of fairness when the task affords such assemblances. With the prominent visibility of multiple semiotic resources assembled in real-time during AOP performances, these sites were ideal for probing the treatment of multimodality in EAP assessment.

Multiperspectival qualitative approach

In classroom-based performance assessments, local cultures and people can be highly influential, making a qualitative, ethnographically oriented methodology a fruitful research strategy. Qualitative methods have been used to investigate rater behaviour previously, such as through verbal reports and interviews (Kim, 2015; Orr, 2002). Ethnographic studies, which gain immediate access to assessment sites, performances, and score-reaching talk between assessors, are less common. An exception is Kalthoff's (2013) ethnographic study of grading practices in German high schools that gained access to score-reaching dialogue between teacher assessors on oral assessments, highlighting the socially constructed dimension of assessment practice.

Canvassing assessor and assessee perspectives is vital in evaluating whether the assessment is being implemented as officially stated by assessment developers and is capturing the cognitive and contextual factors that shape performance. Studies employing qualitative methodology have found that teachers in practice supplant or supplement rating scale criteria with their own ideas on what counts (Orr, 2002), underscoring the need to investigate teachers' stated beliefs and practices. Student conceptions of oral presentation tasks are also diverse (Joughin, 2007), and teacher and student understandings of the same assessment rubrics may differ (Li & Lindsey, 2015). Illuminating similarities and disparities across stakeholders is crucial, as shared understanding can enhance assessment quality (Carless, 2009) and learning experiences.

The study design

This article is based on part of a larger project (Palmour, 2020) which applied a blend of constructivist grounded theory and ethnographic principles (Charmaz, 2014;

Charmaz & Mitchell, 2001; Hammersley & Atkinson, 2019) to guide the qualitative study of AOP assessment practices in higher education institutions in the United Kingdom. The study comprised of two phases: the survey study and the fieldwork study. The survey study encompassed analysis of website pages and questionnaires for both EAP teachers and teaching staff on disciplinary degree modules at UK universities. The fieldwork study took place on two EAP modules on programmes at two different UK universities: field site 1 (FS1) and field site 2 (FS2). Fieldwork involved my (first author's) observation of teaching and assessment events; audio-recorded interviews with student assessees and teacher assessors; audio-recorded score-reaching discussions between teacher assessors; fieldnotes; and document analysis. The larger project covered comprehensive discussion of teaching and assessment processes linked to EAP AOP tasks as well as accounts of AOP practices on degree programmes. This article focuses on data pertaining to EAP settings only. It explores the nature of the construct underlying EAP AOP assessments and particularly how multimodality factors into EAP AOP assessment construct activity. Thus, in this article, the research questions investigated include:

1. What are the theoretical, stated, perceived, and operationalized oral presentation assessment constructs in EAP contexts at UK universities?
2. What tensions exist (if any) related to multimodal communication between these theoretical, stated, perceived, and operationalized constructs?

This article reports on data obtained on stated constructs from responses to a questionnaire for EAP teachers (see Supplementary Materials Section 1 for the EAP Teacher Questionnaire). The questionnaire included three items asking for background information and an additional three closed items and 11 open-ended questions. In this article, theoretical, perceived and operationalized constructs are shared from questionnaires, rating discussions between EAP teachers, and semi-structured interviews with EAP teachers and EAP students (see Supplementary Materials Sections 2 and 3 for Interview Plans).

Research context

The questionnaire responses related to pre-sessional and in-sessional AOP assessments within EAP programmes at UK universities. Fieldwork took place within programmes that required students to complete an oral presentation to meet language entry requirements in order to transition to their prospective undergraduate and postgraduate degree programmes. The fieldwork was conducted on two EAP modules at two different UK universities: one EAP module on a pre-Master's Programme and one on an International Foundation Programme (IFP). The pre-Master's module is referred to as field site 1 (FS1); while the IFP module is referred to as field site 2 (FS2). Document analysis and ethnographic observation showed that both modules included academic literacy elements in the syllabi and teaching and assessment practice rather than teaching academic English language proficiency. Both the IFP and pre-Master's programmes also contained content pathway modules on topics loosely linked to the students' degree courses. In the

EAP sites accessed in the current study, the teachers and students worked together for an academic year. This article reports and compares findings across the pre-Master's and IFP programmes to display practice across diverse sites. The sites represent different institutions, levels of study, and types of oral presentations (group and individual tasks).

Participants

Calls to participate in the EAP strand of the wider research project were sent via the British Association of Lecturers in English for Academic Purposes (BALEAP) mailing list. The target respondents were EAP teachers and assessment designers engaged in summative oral presentation activity within EAP programmes at UK universities. The EAP questionnaire respondents ($n=14$ respondents from 11 institutions) held a variety of roles at UK universities, including EAP course leaders ($n=6$) and EAP tutors ($n=8$). For the fieldwork, convenience sampling was used, and participating fieldwork sites did not necessarily complete the questionnaire. The questionnaire served to indicate phenomena of interest to pursue the fieldwork study. Data were triangulated from fieldwork methods (interviews, documents, rating discussions, and fieldnotes at the sites). In FS1, within the pre-Master's EAP module, there were 7 main participants: 2 teacher assessors, and 5 students out of the 14 students in the class. In FS2, on the IFP EAP module, there were 5 main participants: 2 teacher assessors, and 3 students out of 10 in the class. The main participants are classed as those who consented to interviews as well as classroom and assessment observation.

In FS1, Adam, (pseudonyms used throughout), the main teacher assessor, taught the lessons within the pre-Master's EAP module and marked the presentations with his co-assessor Daniel. Daniel had taught the student participants different modules within the pre-Master's programme. In FS2, at a different UK university to FS1, Georgina was the main teacher of the IFP EAP module, covering syllabus content within the module. She worked with Tracey when rating the AOPs. Tracey had taught student participants other modules within the IFP programme. At both sites, the teachers (>10 years of EAP teaching experience) had established working relationships, partly from having marked presentations together previously.

The groups of EAP student and teacher participants had diverse linguacultural backgrounds and experiences, which shaped their actions related to oral presentation events. One student had studied in English-medium education program for a year and had delivered AOPs in English, while others had not studied in English medium programs and never delivered an AOP in English before. The nature of my interaction in the field sites meant that I was co-constructing the reality under investigation with the participants. There were times when I was an active member and periphery member in interactions (Adler & Adler, 1998). During classroom observations, I interacted with students and teachers in lessons, while at the summative AOP performances and rating discussions, I positioned myself on the periphery of interactions so as not to influence score decisions. I had previously worked alongside one of the gatekeepers at one institution where fieldwork took place. Therefore, I had some familiarity with the processes and the colleagues of participants at one site. However, I had no previous close working relationships with any of the participants. Due to my background as an EAP teacher, I also entered sites

with notions of EAP and AOPs; for example, from previous rating scales I had used. I exercised reflexivity, recognising the need to identify and suspend such judgements and adopt an ethnographic "mode of looking" (Hammersley & Atkinson, 2019, p. 239).

The assessment format and event

At both sites, the formal assessed presentations took place live and in-person with students assesses, the teacher assessors, and me (first author) as the researcher in the audience. Teacher assessors within these courses planned to conduct AOP assessments and assign marks in rating discussions prior to the research study; therefore, the observation conducted on these programmes was happening in naturalistic settings. Table 2 provides the core details about the assessment context.

In both settings, teachers video recorded the AOP assessment performances. In FS1, the teachers referred to video clips of performances, their real-time notes, and their memories of events when assigning marks to the pre-Master's students in a discussion two days after the performances. In FS2, teachers assigned marks directly after they had viewed a series of IFP presentations, referring to their real-time notes and their memories, with no reference to the video recordings.

Teacher assessors used analytic scoring rubrics, designed in-house with colleagues. The matrix in each EAP site contained prose descriptors (e.g., "use of grammar significantly affects intelligibility") in sections (e.g., "content," "fluency and intelligibility"). Descriptors for each section were assigned a band of marks (e.g., 0–20 out of 100). (See the AOP rating scales sections in Table 2 and an analysis of these in the section "Theoretical and stated assessment constructs" that appears below in the Data Analysis section). Parts of the rating scale were reworded and summarised for research dissemination purposes only to safeguard the identify of institutions and participants.

Data analysis procedures

Charmaz (2014) highlighted the progressive focussing approach adopted in constructivist grounded theory studies by bringing "key scenes closer and closer into view" (p. 27). In phase 1 of the project, the survey study of AOP practice was conducted to identify such key leads to pursue in phase 2 of the project: the fieldwork study. All data from the survey study were analysed before the fieldwork study commenced. In both phrases of the project, I analysed data in written form: rater conferences and interviews were transcribed. I coded all data using constructivist grounded theory (Charmaz, 2014) as an analytic framework in an iterative process. Table 3 summarises the data set used in this article.

Line-by-line initial coding, focused coding and categorising, constant comparison, memo-writing and theoretical sampling were conducted, using NVivo 12 to manage the process. Initial codes were collapsed into broader focused codes. Then, building memos on codes and triangulating raw data from all methods anchored ideas, and ultimately, balanced evidence with theoretical argument (Charmaz & Mitchell, 2001). Initial fieldnotes, coding, and writing memos engendered the identification of codes and categories requiring further evidence and exploration.

Table 2. Field site contextual details.

Category	Field Site 1 (FS1)	Field Site 2 (FS2)
Type of programme	Pre-Master's Programme	International Foundation Programme
Modules on the programme	EAP module and content pathway modules	EAP module and content pathway modules
Module under study in the research	EAP module	EAP module
Students	14 students, IELTS 5.5 on entry L2 users of English	11 students, IELTS 4.5 on entry L2 users of English
Target disciplines	Mixed	Mixed
AOP assessment task	Group presentation on group project findings 7 mins per presenter followed by 5 mins Q&A	Individual presentation on topic of student's choice 10 mins per presenter followed by 5 mins Q&A
Time in the field	Semester one	Semester two
AOP assessment weighting	Potentially 10% of final mark. One of two presentations completed by students, the best performance was then included in the portfolio. The programme portfolio also consisted of reflective writing tasks, an essay, report writing, a seminar discussion, and a listening exam.	Contributed to 25% of final mark The programme portfolio also included a written task, listening test and reading test
AOP rating scale sections (reworded to uphold higher degree of anonymity)	1. Content 2. Fluency and intelligibility 3. Accuracy and range 4. Presentation skills 5. Group work	1. Overall Competence in Academic English 2. Content 3. Fluency and coherence 4. Vocabulary 5. Grammar 6. Pronunciation

Table 3. Dataset.

Survey study data	Fieldwork study data: Field site 1	Fieldwork study data: Field site 2
14 responses to EAP Teacher questionnaire	Documents: Module year-round syllabus Rating scale for oral presentation task	Documents: Module description Rating scale for speaking tasks
	Interviews: Two interviews with one teacher Member check interviews with two teachers Two interviews with four students One interview with one student	Interviews: Two interviews with one teacher Member checks with two teachers Two interviews with three students
	Rating discussions: one rating discussion (allotting marks for 14 students)	Rating discussions: one rating discussion (allotting marks for 10 students) Fieldnotes:
	Fieldnotes: Notes taken during time at the site, including at one rating discussion and observations of three lessons	Notes taken during time at the site, including at one rating discussion and observations of three lessons

I conducted theoretical sampling to provide fuller categories and richer analysis. For example, interviews and member checks included specific questions related to focused codes such as "difficulty noticing language use" and "developing rather than showcasing language" to collect sufficient data. The process of coding and constant comparison determined which theory applied to the data, meaning the Oracy Skills Framework was admitted into the analysis process after some initial and focussed coding took place. Charmaz (2014) emphasised researching how "consistent" the categories are with the experiences under study. Coding was conducted by the first author only; however, interpretations of the data were shared with teacher participants in member checks.

Data analysis

In the analysis that follows, first the EAP AOP theoretical and stated assessment constructs are presented from questionnaire and rating scales data. After this, the perceived and operationalized constructs are provided from interviews with teacher assessors and student assessees, rating discussions, and fieldnotes. Factors underpinning the tensions between the perceived, operationalized, and the stated constructs are outlined through the discussion of the following themes: teachers' challenges in noticing linguistic performance; teachers' expanded constructs; teachers' tensions implementing the stated constructs; students developing rather than showcasing linguistic performance; and students' multimodal approach.

Theoretical and stated assessment constructs

Questionnaire respondents were asked "What aspects did you assess and what was the weighting for each component?" (see the EAP Teacher Questionnaire in the supplementary materials"). Some of the rating scales are strongly linguistic in focus (see response from EAP Teacher 10 below), while in others, non-linguistic skills feature prominently (see response from EAP Teacher 5 below).

[EAP Teacher 10, EAP Questionnaire]

Equally weighted: grammar, vocabulary, interaction, pronunciation, fluency

[EAP Teacher 5, EAP Questionnaire]

40% linguistic features, 20% Spontaneity in Responses, 20%Presentation Skills, 20% Presentation Content and Preparation

The data from the EAP teacher questionnaire showed that the majority of rubrics included at least 40% linguistic features. In EAP Teacher 10's reporting of the rating scale sections, the construct components include interaction, which on the surface is an expanded language construct. However, given the communicative resources that can be elicited in an AOP, the construct remains somewhat narrow. EAP Teacher 7 provided a justification for such an approach by stating that the prime objective is language development; therefore, the assessment criteria must reflect this.

[EAP Teacher 7, EAP Questionnaire]

We have had many discussions in our team about the weighting and keep coming back to the argument that the module is a language module and the weighting needs to reflect this.

The target features of "grammar" and "vocabulary" contained in the EAP Teacher 10's rubric demonstrate a prime focus on language rather than discourse and communication more broadly. This approach is at odds with contemporary theoretical constructs of EAP (e.g., Hyland, 2018) and calls for expanded language assessment constructs.

The questionnaire data showed the majority of EAP AOP assessments to be L2 performance assessments in the weak sense (McNamara, 1996), due to their heavy focus on discrete linguistic skills. Teachers also widely referred to AOPs as "speaking" assessments, their broader theoretical construct of the AOP assessment. Despite this, a good proportion of assessment rubrics contained non-linguistic features, similar to EAP Teacher 5's task. Non-linguistic features were often under categories such as "content" and "presentation skills." Descriptors included reference to "structure," "analysis," "use of slides," "body language," and "eye contact."

The documents, including rating scales, from the fieldwork study provided similar information to the questionnaire, on the official representation of AOP assessments in EAP contexts. The sections in FS1 pre-Master's analytic scoring rubric were Content; Fluency and Intelligibility; Accuracy, Appropriacy and Range; Presentation Skills; and

Group Work. The proportion of the weighting of the linguistically oriented sections (Fluency and Intelligibility; Accuracy, Appropriacy and Range) at 40% against the non-linguistic skills (Content; Presentation Skills; Group work) at 60% indicates marginally greater emphasis on the non-linguistic features of the AOP performances in the rating scale.

Comparatively, in FS2 on the IFP, descriptors pertaining to linguistic performance comprised the majority of the rating scale. The sections included Overall Competence in Academic English; Content; Fluency and Coherence; Lexis; Grammar; and Pronunciation. The weighting for each section was not clearly stipulated on the criteria; the main EAP teacher stated in interviews that the sections equally contributed to the final score. The number of linguistically oriented sections (Overall Competence in Academic English; Fluency and Coherence; Lexis; Grammar; and Pronunciation) outweighed the non-linguistic criteria (in the Content section).

In answer to RQ1, the survey study and the fieldwork study data analysis revealed a tendency in theoretical and stated EAP AOP constructs to emphasise discrete spoken linguistic skills. In many settings, cognitive skills (e.g., content) and in some settings multiple modes (e.g. slides, gestures) are referenced in stated constructs. However, how performances in the EAP contexts are, in fact, mapped (or not) to these rating scales complicates the narrative and notions of the constructs underlying the EAP AOP assessments that feature in this project.

Teacher assessors' perceived and operationalized constructs

In relation to RQ2, the EAP teachers' perceived and operationalized constructs – from questionnaire and fieldwork data – revealed pitfalls associated with stated constructs. The following sections share teachers' cited challenges in noticing linguistic performance in real-time oral presentations. The teachers have expanded notions, compared to the stated constructs, of what the assessment involves. Furthermore, there were challenges with implementing aspects of the stated construct in practice.

Teachers' challenges noticing linguistic performance. EAP questionnaire respondents were asked "What aspects of marking and giving feedback on AOPs have you found challenging?" In this open-ended question (See Supplementary Materials Section 1 for EAP teacher questionnaire), 3 out of 14 respondents noted difficulty making judgements about linguistic competence.

[EAP Teacher 6, EAP Questionnaire]

All because there is so much to think of in one go.

Probably language (use of appropriate vocab and grammar) as easier to reflect back on content/argument and easy to notice delivery.

[EAP Teacher 12, EAP Questionnaire]

Trying to distinguish what consists of "poor use of grammar"—as it is quite hard to notice grammatical errors in speech (much easier when marking essays)

[EAP Teacher 13, EAP Questionnaire]

It is sometimes hard to give feedback on individuals grammatical errors because there are other aspects including content and presentation skills, which require more focus to mark.

Discrete linguistic components of vocabulary and grammar were not easily recalled for EAP Teacher 6 and less noticeable in speech for EAP Teacher 12 when marking AOPs. EAP Teacher 13 suggested, somewhat contrary to EAP Teacher 6, that the non-linguistic areas of content and presentation skills require dedicated attention, but like EAP Teacher 6, reported difficulty in processing a particular mix of linguistic and non-linguistic skills in real-time marking. All data extracts show that the oral presentations are considered cognitively demanding to rate. EAP teachers stated that reaching judgements proved challenging due to a difficulty in noticing the target linguistic skills in their AOP tasks.

This cited challenge faced by teachers is particularly noteworthy because it raises an important validity concern related to confidence with making judgements on the target construct (Knoch & Macqueen, 2019). Difficulty in noticing linguistic competence was a phenomenon which was taken forward to ethnographically orientated fieldwork on two EAP modules.

During the fieldwork at two EAP sites, teachers were asked what areas they wished to change about the assessment. Adam, similar to questionnaire respondents, spoke about the difficulty in judging and marking linguistic form in oral presentation assessments:

[Adam, Field Site 1, Interview]

With the accuracy and the range as we talked about before, I find it very hard to assess you know based on particular errors, and it really I think it has to be errors which affect intelligibility or it has to be kind of also just the range of language being used. I think it's quite hard to listen for "complex language or complex grammar structures."

Adam, in interviews, advocated a fluency-based approach. Parts of the rating scale were compatible with Adam's stated priority, with descriptors such as "which may impede understanding" prioritising intelligibility in the pre-Master's EAP AOP rating scale. Contained within the rubric were also descriptors such as "attempts at using complex language," however. Listening for particular complex grammatical structures was difficult for Adam, suggesting that a strong focus on form is challenging to achieve when rating AOPs – an extended discourse-level speech task – in real-time. Therefore, fluency-oriented elements, unlike accuracy-based elements, of parts of the stated construct in the rubric are more compatible with Adam's perceived construct.

Similarly, in FS2, Georgina stated that sections of the rubric on the IFP EAP module could be tailored more closely to the assessment task.

[Georgina, Field Site 2, Interview]

So this (the language section in the rubric) needs changing. Because this bit here is based on the IELTS marking criteria.

Compared to FS1, in FS2 the language criterion was heavily based on IELTS criteria, with the wording very close to the IELTS speaking rating scale (IELTS, 2023). Like in the IELTS speaking band, there was no reference to presentation skills such as use of slides, gestures, and handling questions. The rating scale in FS2 contained a section on content, which is a departure from the IELTS speaking descriptors. Georgina admitted that the linguistic skills in the rating scale "needs changing" to be more tightly connected to what the oral presentation task in fact involves. There was hesitancy displayed at both field sites regarding adapting criteria. Therefore, the readily available and familiar speaking criteria from large-scale standardised tests such as IELTS continue to inform assessment design and implementation of tasks that involve drawing on different communicative resources.

In relation to RQ2, at both sites, the teacher assessors cited issues with stated construct components related to linguistic performance, but to a far greater degree in FS2 than in FS1.

Teachers' expanded constructs. In FS1 and FS2, pertaining to RQ1, the teachers' theoretical construct was academic speaking, whereas their perceived and operationalized constructs were broader notions of readiness for academic study. When assigning marks, teacher assessors not only referred to features of the oral presentation task, but also commented on wider qualities.

[Tracey, Field Site 2, Interview]

If you're passing them, are you sure that they are going to be able to perform on an undergraduate programme.

Daniel made a similar comment in FS1 stating that one student would be "impressive in a seminar" on their target Master's course. Thus, Daniel indicated this perceived readiness for the target domain should be reflected in the AOP mark.

In fact, addressing RQ1, the teacher assessors in score-reaching talk discussed a mix of cognitive, linguistic, use of multimedia, and social and emotional abilities. This indicates that an "ability to perform" in their target degree programmes is part of the operationalized construct. All teachers heavily referenced "cognitive" aspects of the Oracy Skills Framework (Mercer et al., 2017), such as "choice of content" and sources, "structure" and "focus on task," which are primarily contained in the "content" sections in the rubric at both sites. The teacher assessors also mentioned "Social & Emotional" dimension of practices (e.g., "responding appropriately" to questions and "confidence"). Some of these qualities are contained in the FS1 pre-Master's rubric in the "teamwork" and "presentation skills" section, but none of these is covered in FS2 IFP rating scale. (See Supplementary Materials, Section 4 for more data on teachers' assessment constructs at both sites).

In sum, the way that teachers navigated score-reaching decisions in rating discussions revealed tensions with operationalising the given stated constructs in their contexts. The teachers valued not only the linguistic dimension of performances, but also content abilities, socio-emotional skills, and multimodal competencies.

Teachers' tensions implementing stated constructs. When rating the AOPs, the teacher assessors in the field sites were engaged in a range of multidimensional and complex overlapping processes. In both sites, the teachers reached score decisions in pairs, describing and evaluating performances and negotiating scores with their colleagues.

In FS2, Tracey and Georgina made reference to the cognitive elements in the rating scale when reaching score decisions, particularly noting "descriptive rather than analytical" approaches. This showed alignment with the stated sub-construct labelled "content." They also valued performance on presentation skills such use of slides which were not contained in the rating scale:

[Georgina, Field Site 2, Interview]

Basically read from whatever slides she had on.

[Tracey, Field Site 2, Interview]

Tell you one thing I did like though, I liked her slides. Very clear. I liked the colour coordination.

[Tracey, Field Site 2, Interview]

Although she's not you know as an accomplished a presenter as some of the others. Once she sorts out presentation style I think you know and gets her head in order with sorting things out I think she'll go from strength to strength.

During the rating discussions in FS2, Georgina and Tracey valued coherence and slide design elements. The interplay between the speech, gestures, and slides was only explicitly mentioned when students read from slides. This was seen as ineffective presenting. Georgina did not make reference to the use of gestures and eye contact when reaching score decisions.

The implementation of linguistic related sub-constructs also proved challenging. Stated constructs at both sites involved extracting linguistic evidence in real-time that did not impact intelligibility. For instance, the descriptor in the middle band of the "fluency" section in the FS1 rating scale prompts raters to detect "mispronunciation of some words" that "generally do not affect intelligibility." In the score-reaching dialogue, Adam and Daniel referred to notes they had made, pertaining to this descriptor on pronunciation, during the AOP performance. The teachers noted that language use extracted during a performance may not be a fair reflection of a student's overall AOP performance and readiness for academic study, indicating tensions that can arise when operationalizing constructs (addressing RQ2).

[Field Site 1, Rating Discussion]

Adam: I'd written down "questioneer" so I think a pronunciation of questionnaire was slightly . . .

Daniel: But didn't you think that was because it was generally her fluency and accuracy of pronunciation is quite good and then there'll be ONE WORD that is completely wrong and you kind of pick on it.

(. . .)

Adam: Well I've gone 10 or 11. But maybe I don't know 11 [top of that band].

Daniel: [I think we're] distracted by the "questioneer."

Adam: Ok 11 then. We shouldn't be distracted by one 'questioneer'!

The student pronunciation of the word questionnaire was recognisable to the teacher assessors but not considered a standard pronunciation. The teachers noticed that this close-up is not a fair reflection of the student's communicative ability and so discounted this example. Adam and Daniel were attuned to this potential pitfall with the representativeness of evidence selected and highlighted a vital checking strategy that assessors can implement in EAP AOPs.

This episode raised further questions related to judging and marking AOPs. The word "questionnaire" was used multiple times during the assessment, as the task involved designing a questionnaire and reporting the findings. The word was evidently clear from context and was written on the student assessee's slides in multiple stages of the task. It is important to consider why such evidence entered the decision-making dialogue and was not discounted on the grounds of taking the full range of multimodal resources deployed by the student assessee into account.

The student approaches to the task offer crucial insights into how isolating the spoken mode can be problematic, and how some assessees' perceived and operationalized constructs risk being at odds with stated constructs, particularly regarding the role of multimodality.

Student assessees' perceived and operationalized constructs

During interviews, after the formal assessment performance, students in FS1 and FS2 were prompted to describe what was important to successful AOP assessments within their module (the final interview prompt in Supplementary Material, Section 3). The students had referred to the rubric used on their respective modules, but their notions of what should be assessed diverged from this key artefact. The first student interviewed in FS2 offered percentages for sub-constructs they believed should be included in the assessment rubric. Therefore, the remaining IFP students in FS2 were then also asked to provide percentages to aid comparison. Table 4 presents the stated and students' perceived constructs in FS2.

Table 4. Field site 2 rubric and student assessment constructs.

FS2 Rubric Stated Construct	FS2 Student Perceived Construct
1. Overall competence in Academic English	Content 50% Presentation Skills 35%–40% Language 10%–15%
2. Content	(Marina, interview)
3. Fluency and coherence	Content 50% Presentation Skills 30% Language 20%
4. Vocabulary	(Tara, interview)
5. Grammar	Content 70% Presentation Skills 5% Language 25%
6. Pronunciation (FS2 rating scale sections)	(Aimee, interview)

The three student participants in FS2 diluted the emphasis on language compared with the rating scale (stated construct). In the rating scale, 5/6 of the criteria are devoted to language and 1/6 to content. Conversely, all student participants in FS2 assigned 3/6 or more of the weighting to content. For Marina and Tara, FS2 students, presentation skills also accounted for a larger percentage of the grade than language. In FS1, interview data suggested that the EAP pre-Master's students' perceptions of the assessment task were more tightly aligned to the stated construct. One dominating factor here is that the rating scale in FS1 covers presentation skills, which included physical and use of multimedia dimensions (see the dimensions in the Oracy Skills Framework in Table 1). One exception to the focus on content over language was student Ben's approach to the task in FS1. He was reportedly "improvising" in his assessment so that he could practise his speaking abilities. See Supplementary Materials Section 5 for data from FS1 and FS2 on students' assessment constructs.

Students provided reasons behind emphasising aspects other than spoken linguistic performance in their AOP assessments. When Aimee, a student in FS1, provided her marking criteria, she highlighted the difference between the in-house educational EAP assessments from her year-long module and IELTS:

[Aimee, Field Site 2, Interview]

So this course's aim is to improve language so language might be more higher. I would put for a presentation assessment on an IFP, I think 80% is a bit heavy, so 70% for content and 25% language and 5% presentation skills. I think IELTS is 80% on language and 20% on content!

Although Aimee recognised that the "course's aim is to improve language," she did not deem it a priority in the assessment task. By stating that 80% of the IELTS test is language and 20% content, Aimee characterised IELTS Academic as having a far weaker

sense of performance assessment than the classroom-based, oral presentation EAP assessment. In her verbatim quote, Aimee highlighted a shift from a focus on language to content that she and many of her fellow students encounter from IELTS Academic to EAP educational programmes. Reasons underlying this shift were provided in interviews and are explored next.

Students developing rather than showcasing linguistic performance. The majority of students focused energies on developing and showcasing academic literacies rather than their linguistic performance. Ivan shared why students might opt to focus attention on content-related competences.

[Ivan, Field Site 1 Interview]

I think I can't improve my English in such a short time, but I can concentrate on making a good content and at least will probably improve my English.

For Ivan, the rationale provided for dedicating attention to improving and showcasing content was that language ability was developed over time by completing the course and assessments.

Coupled with a focus on content, students downplayed the importance of language due to viewing verbal communication in the AOP as predominantly a communicative medium rather than the target of the assessment. From analysis of interview data, when referring to "language" in the extracts below, the students seemed to be referring to grammar and vocabulary in spoken production.

[Fiona, Field Site 1, Interview]

I won't be worried about language. Because language is just part of the presentation. We should concentrate more on what we're talking about. Not what kind of language we choose to talk.

[Marina, Field Site 2, Interview]

Even if their (fellow students') English isn't perfect, if we can understand what they're trying to say that's ok. It has to be near enough. It's the effort that they make to answer the question.

[Aimee, Field Site 2, Interview]

So if the grammar was wrong, it not matter. I think it is not much big issue for the presentation I think. It's not big issue I think.

Marina's, Aimee's, and Fiona's quotes above suggest that they viewed presenters using spoken English as a "vehicular language" (Mauranen, 2006, p. 124) and lingua franca, not as the target of the assessment. The students reported not attending to features with low communicative valency. They would not focus on certain non-standard usage of English in the AOP task even when it is assessed. However, some students were less accommodating of non-standard use in written text on slides.

[Marina, Field Site 2, Interview]

Spelling mistakes because all the audience look at the PowerPoint these are mistakes that we would all notice a lot so I think it's important on the PowerPoint.

It materialised that when students were downplaying the importance of "language," they were referring to the accuracy of the spoken delivery. The speech of a presentation is a central part of the communicative act in oral presentations and considered a privileged property of many AOPs. However, the importance of the spoken delivery was diluted in students' reports. The majority of students viewed it as a vehicle by which to convey a message while drawing on other modes of communication available (See Supplementary Materials Section 5 for further data on students' assessment constructs). Relating to RQ2, this indicates that students' operationalized constructs diverged from the stated construct, which centred more heavily on linguistic components as the target of the assessment. Instead, the majority of students viewed language as functioning more as a medium of communication.

Students' multimodal approach. The assemblage of multimodal and multisensorial (Shohamy & Pennycook, 2019) resources available in the AOP tasks meant that students regarded the spoken monologue as one tool in their communicative repertoires. In fact, Aimee and Marina, along with other students, put emphasis on the multimedia PowerPoint presentations in conveying their messages.

[Aimee, Field Site 2, Interview]

The main message they many times saying and the main message in the presentation (slides).

[Marina, Field Site 2, Interview]

So the PowerPoint will show what they want to tell as well I think. I think we can understand what they want to say.

During interviews, students shared that they devoted considerable energy to preparing PowerPoint slides. Anna in FS1 was an exception, stating that the slides were prompts for her, and she felt it important to have "a good balance between" the slide content and spoken delivery. Anna, however, also thought her group should have used better bar charts and images to improve the communication in their assessed performance.

All students shared what other non-verbal communicative resources were important to them. The holistic act of communication involved drawing on the affordances offered in spoken utterances, but also the "physical" (gestures, eye contact) and "use of multimedia" dimensions (graphics on slides) of the Oracy Skills Framework (see Table 1 for the Oracy Skills Framework):

[Marina, Field Site 2, Interview]

They (eye contact and body language) are a massive thing. I think it's really important.

[Henry, Field Site 1, Interview]

So the first important thing is to stand face to the audience and your eyes need to look around. This is very important. Probably like your hand have a little bit of action and when you show a graph probably use shaped hand.

It is important not to overstate nor understate the role of non-verbal communication. Teacher assessors in FS2 were reticent to assess gestures because they feared that judgements could be discriminatory. The students made the case that non-verbal communicative resources are available and used to convey content. Therefore, arguably these cannot be ignored when reaching judgements on successful situated communication, underscoring a tension between how language is extracted in teachers' stated and operationalized constructs and how students approach the task (addressing RQ2).

In both field sites, the students' perceived and operationalized constructs were at odds, in parts, with official stated constructs, but more starkly in FS2. The students prioritised cognitive aspects (content-related abilities in the stated constructs) and overall tended to downplay the emphasis on spoken delivery compared to stated constructs. In response to RQ2, the findings strongly indicated that students were taking advantage of integrated transmodal communication in their set assessment task; however, this was not mirrored in the stated constructs and to some degree not mirrored in teachers' operationalized constructs.

Discussion

This article investigated the construct underlying EAP AOPs at UK universities, focussing on the role of multimodality. The presentation of data has demonstrated that, in response to RQ1, theoretical and stated constructs of EAP AOPs are often performance assessments in the weak sense (McNamara, 1996), in that they report to assess primarily linguistic features of performance. However, a number of EAP AOPs assess a variety of non-linguistic facets such as physical, cognitive, social, and emotional aspects (Mercer et al., 2017), and the use of multimedia. This is the case even when the non-linguistic dimensions do not account for the majority of the weighting in rating scales. Furthermore, teacher assessors and student assessees dilute, and diverge from, these official linguistic focused representations of AOP assessment tasks. Their expanded perceived and operationalized constructs are closer to stronger senses of performance assessment.

In response to RQ2, tensions between theoretical, stated, perceived, and operationalized constructs surfaced due to the stated constructs' inability, in cases, to reflect communicatively salient features of the AOP tasks. In survey responses and during fieldwork, teacher assessors relayed challenges in noticing and judging particular linguistic descriptors. Teachers at field sites who believed in a fluency-oriented approach continued to extract information on linguistic features. However, at times, they checked the fairness of the linguistic evidence sampled from the oral presentation. Taking a multimodal approach (e.g., considering the use of spoken, written language and graphics in conjunction) may aid the decision-making as to whether intelligibility is affected. The teacher assessors in FS2, on the IFP module, also acknowledged the inadequacy of descriptors pertaining to language in the AOP rating scale, which were heavily based on the IELTS speaking criteria. This echoes

concerns highlighted in previous studies regarding the relevance of linguistically oriented descriptors in proficiency tests when mapped to classroom-based integrated EAP assessments (Green, 2005; Green, 2019; Uludag & McDonough, 2022). This study highlights tensions when assessing language proficiency and academic achievement through an AOP on EAP educational courses. This is because these courses in fact develop academic literacies, and academic literacy tasks involve non-linguistic skills.

At the two fieldwork sites, EAP teacher assessors diluted the focus on linguistic features, due to a focus on aspects deemed crucial to AOP task fulfilment and the teachers' notions of readiness for academic study. The teachers valued physical, cognitive, social, and emotional skills (Mercer et al., 2017) as well as the use of multimedia. Notable emphasis was placed on learning abilities within the course, which fall under the cognitive and social and emotional dimensions of Mercer et al.'s Oracy Skills Framework. Some of these abilities were not contained in rating scales. This corroborates findings that teacher assessors in classroom assessments refer to criteria outside the rating scales in their decision-making processes (Orr, 2002) and refer to information from outside sources in reference to predicting future performance (Bonner, 2013).

The majority of student participants conceptualised AOP communication as an assemblage of semiotic resources. Canagarajah (2018) described an assemblage perspective as when "all modalities including language work together and shape each other in communication" (p. 39). Students regarded communicative proficiency as shaped by artefacts (Kuby, 2017) when describing the modality of PowerPoint slides as central to relaying "the main message" in the AOP. This emphasis on the use of multimedia was at odds with the stated construct in these "speaking assessments" in FS2, and, to a much lesser degree, in FS1. In FS1, greater alignment between the stated, perceived, and operationalized constructs was detected, as the rating scale included a section on presentation skills, which recognised modes such as slides and gestures. The slides were evaluated in terms of how well they supported the presentation and gestures for improving audience engagement. This is encouraging, as it better reflects the resources assessees use.

The students' perspective coheres with findings that visuals and gestures mediate and shape language use, with spoken and visual components forming an integrated performance in AOPs (Canagarajah, 2018; Rowley-Jolivet, 2002). Therefore, caution must be exercised to ensure not to discount the use of particular modes of communication when evaluating aspects such as intelligibility and content in AOPs. An integrated treatment of slides and verbal communication needs to be negotiated and clearly communicated in stated AOP assessment constructs. Overall, both students and teachers seem to underestimate the interplay between gestures, spoken language, and slides in their AOP tasks. The spoken monologue is but part of a complex multimodal and embodied activity of the AOP.

Conclusion

In the EAP contexts featured in this study, the theoretical and stated AOP assessment constructs diverged from perceived and operationalized constructs to different degrees. EAP AOP events are stronger senses of performance assessments (McNamara, 1996) in practice than reflected in the rating scales in that teacher assessors and students value non-language-related aspects. Focusing on showcasing content rather than linguistic

performance was the majority of students' main priority at the two field sites. The official discourse, including the rating scales, nudged assessors to retain a focus on linguistic features; however, teachers and students raised fundamental concerns with aspects of this approach. These important user perspectives demonstrate that in some settings, contemporary understandings of language assessment constructs require a determined effort to recognise the mediums and modalities available.

The scope of the current project was an investigation of the treatment of communicative competence, with minimal reference in the analysis to aspects such as content and group work in EAP AOP tasks. Further research in this area would be welcomed. The study's limitations include the lack of a second coder to enhance consistency in the data analysis process and member checks conducted with teachers only. Although the current study took place within a small number of EAP programmes, it is hoped that the level of description communicated may enable others to determine that the findings in these cases hold relevance in their own contexts (Lincoln & Guba, 1985). Research studies that offer a thick description of AOP assessment practices, including the interplay between modes, are a vital next step to build on this study's modest contribution.

This article contributes to understandings of multimodal classroom assessment in EAP settings. The findings are relevant to EAP assessment researchers, developers, teachers, and students with stakes in these assessments. The study's findings strongly indicate that using an AOP, which expects assessees to use multimedia while having discrete spoken linguistic features as the target construct, is ill-advised. Rating scales and assessment decisions which acknowledge and integrate spoken and written modes of communication with use of multimedia would be welcomed. This is to ensure that student assessees who draw on semiotic resources available in tasks are not disadvantaged. To best safeguard against this, students' approaches to assessment tasks should be a consideration in shaping the curriculum and assessment.

Investigating tensions in artefacts, perspectives, and practices has the power to promote more congruence between stated constructs and perceived constructs (Macqueen, 2022). AOP events, although complex, are ideal sites in which to implement expanded constructs of language and successful communication, which effectively elicit "multimodal and multisensorial assemblances." The teacher assessors' and student assessees' perceived and operationalized constructs reported in this article could encourage EAP assessment developers and teachers to consider using an AOP performance assessment to (continue to) assess an effective assembler of communicative resources while assessing and developing content and learning abilities.

Author's note

Acknowledgements

Author contributions

Louise Palmour: Conceptualisation; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualisation; Writing—original draft; Writing—review & editing.

Declaration of conflicting interests

Funding

ORCID iD

Louise Palmour (iD) https://orcid.org/0000-0002-3426-8414

Supplemental material

Supplemental material for this article is available online.

References

Adler, P. A., & Adler, P. (1998). Observational techniques. In N. K. Denzin & Y. S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (pp. 377–392). SAGE.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford University Press.

Bonner, S. M. (2013). Validity in classroom assessment: Purpose, properties and principles. In J. H. McMillan (Ed.), *Sage Handbook of research on classroom assessment* (pp. 87–106). SAGE.

Canagarajah, S. (2018). Translingual practice as spatial repertoires: Expanding the paradigm beyond structuralist orientations. *Applied Linguistics*, *39*(1), 31–54. https://doi.org/10.1093/applin/amx041

Carless, D. (2009). Learning-oriented assessment: Principles, practice, and a project. In L. H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P. M. Johnston, & M. Rees (Eds.) (education.). *Tertiary assessment and higher education student outcomes: Policy, practice, and research* (pp. 79–90). Ako Aotearoa & Victoria University of Wellington.

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). SAGE. https://doi.org/10.1002/9781405165518.wbeosg070.pub2

Charmaz, K., & Mitchell, R. G. (2001). Grounded theory in ethnography. In P. Atkinson, A. Coffey, S. Delamont, J. Lofland, & L. Lofland (Eds.), *Handbook of ethnography* (pp. 160–174). SAGE.

Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge University Press. https://doi.org/10.1017/CBO9780511732911

Green, A. (2005). EAP study recommendations and score gains on the IELTS Academic Writing test. *Assessing Writing*, *10*(1), 44–60. https://doi.org/10.1016/j.asw.2005.02.002

Green, A. (2019). Restoring perspective on the IELTS test. *ELT Journal*, *73*(2), 207–215. https://doi.org/10.1093/elt/ccz008

Hammersley, M., & Atkinson, P. (2019). *Ethnography: Principles in practice* (4th ed.). Routledge. https://doi.org/10.4324/9781315146027

Harwood, N., & Petrić, B. (2011). English for Academic Purposes. In J. Simpson (Ed.), *Routledge handbook of applied linguistics* (pp. 243–258). Routledge.

Heron, M. (2019). Making the case for oracy skills in higher education: Practices and opportunities. *Journal of University Teaching and Learning Practice*, *16*(2), 1–16. https://doi.org/10.53761/1.16.2.2

Huxham, M., Campbell, F., & Westwood, J. (2012). Oral versus written assessments: A test of student performance and attitudes. *Assessment & Evaluation in Higher Education*, *37*(1), 125–136. https://doi.org/10.1080/02602938.2010.515012

Hyland, K. (2018). Sympathy for the devil? A Defence of EAP. *Language Teaching*, *51*(3), 383–399. https://doi.org/10.1017/S0261444818000101

Hyland, K., & Shaw, P. (2016). Introduction. In K. Hyland & P. Shaw (Eds.), *Routledge handbook of English for academic purposes* (pp. 1–15). Routledge. https://doi.org/10.4324/9781315657455

IELTS. (2023). *IELTS scoring in detail*. https://www.ielts.org/for-organisations/ielts-scoring-in-detail

Jamieson, J. (2013). Defining constructs and assessment design. In A. J. Kunnan (Ed.), *The companion to language assessment* (*vol. 2*, part 7, pp. 1–17). Wiley Blackwell.

Januin, J., & Stephen, J. (2015). Exploring discourse competence elements in EAP class presentations through document and ethnographic analyses. *Social & Behavioural Sciences*, 208, 157–166. https://doi.org/10.1016/j.sbspro.2015.11.192

Jewitt, C. (2017). An introduction to multimodality. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 15–30). Routledge.

Joughin, G. (2007). Student conceptions of oral presentations. *Studies in Higher Education*, *32*(3), 323–336. https://doi.org/10.1080/03075070701346873

Kalthoff, H. (2013). Practices of grading: An ethnographic study of educational assessment. *Ethnography and Education*, *8*(1), 89–104. https://doi.org/10.1080/17457823.2013.766436

Khabbazbashi, N., Chan, S., & Clark, T. (2022). Towards the new construct of academic English in the digital age. *ELT Journal*, *77*, 207–216. https://doi.org/10.1093/elt/ccac010

Kim, H. (2018). What constitutes professional communication. Is language proficiency enough for testing purposes? *Language Testing*, *35*(3), 403–426. https://doi.org/10.1177/0265532218758127

Kim, H. J. (2015). A qualitative analysis of rater behavior on an L2 speaking assessment. *Language Assessment Quarterly*, *12*(3), 239–261. https://doi.org/10.1080/15434303.2015.1049353

Knoch, U., & Macqueen, S. (2019). *Assessing English for professional purposes*. Routledge. https://doi.org/10.4324/9780429340383

Kress, G. (2017). What is a mode? In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 60–75). Routledge.

Kuby, C. (2017). Why a paradigm shift of "more than |human ontologies" is needed: Putting to work poststructural and posthuman theories in writers' studio. *International Journal of Qualitative Studies in Education*, *30*(9), 877–896. https://doi.org/10.1080/09518398.2017.1336803

Lea, M., & Street, B. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, *23*(2), 157–172. https://doi.org/10.1080/03075079812331380364

Li, J., & Lindsey, P. J. (2015). Understanding variations between student and teacher application of rubrics. *Assessing Writing*, *26*, 67–79. https://doi.org/10.1016/j.asw.2015.07.003

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE. https://doi.org/10.1016/0147-1767(85)90062-8

Macqueen, S. (2022). Construct in assessments of spoken language. In T. Haug, U. Knoch & W. Mann (Eds.), *Handbook of language assessment across modalities* (pp. 233–250). Oxford University Press. https://doi.org/10.1093/oso/9780190885052.003.0020

Mauranen, A. (2006). Signalling and preventing misunderstanding in english as lingua franca communication. *International Journal of the Sociology of Language*, 177, 123–150. https://doi.org/10.1515/IJSL.2006.008

McNamara, T. (1996). *Measuring second language performance*. Longman.

McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Blackwell.

Mercer, N., Warwick, P., & Ahmed, A. (2017). An oracy assessment toolkit: Linking research and development in the assessment of students' spoken language skills at age 11-12. *Learning and Instruction*, *48*, 51–60. https://doi.org/10.1016/j.learninstruc.2016.10.005

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, *14*, 5–8. https://doi.org/10.1111/j.1745-3992.1995.tb00881.x

O'Hagan, S., Pill, J., & Zhang, Y. (2016). Extending the scope of speaking assessment criteria in a specific-purpose language test: Operationalizing a health professional perspective. *Language Testing*, *33*(2), 195–216. https://doi.org/10.1177/0265532215607920

Orr, M. (2002). The FCE speaking test: Using rater reports to help interpret test scores. *System*, *30*, 143–154. https://doi.org/10.1016/S0346-251X(02)00002-7

Palmour, L. (2020). *Capturing the human dimension of performance assessment: An investigation of decision making in academic oral presentation assessment tasks on English for academic purposes courses at UK universities* [Unpublished doctoral dissertation]. University of Southampton.

Plough, I. (2021). A case for nonverbal behaviour: Implication for construct, performance and assessment. In M. R. Salaberry & A. R. Burch (Eds.), *Assessing speaking in context: Expanding the construct and its applications* (pp. 50–69). Multilingual Matters. https://doi.org/10.21832/9781788923828-004

Rowley-Jolivet, E. (2002). Visual discourse in scientific conference papers: A genre-based study. *English for Specific Purposes*, *21*(1), 19–40. https://doi.org/10.1016/S0889-4906(00)00024-7

Salaberry, M. R., & Burch, A. R. (2021). Assessing speaking in context. Expanding the construct and applications. In M. R. Salaberry & A. R. Burch (Eds.), *Assessing speaking in context: Expanding the construct and its applications* (pp. 3–22). Multilingual Matters. https://doi.org/10.21832/9781788923828-002

Schmitt, D., & Hamp-Lyons, L. (2015). The need for EAP teacher knowledge in assessment. *Journal of English for Academic Purposes*, *18*, 3–8. https://doi.org/10.1016/j.jeap.2015.04.003

Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *The Modern Language Journal*, *95*(3), 418–429. https://doi.org/10.1111/j.1540-4781.2011.01210.x

Shohamy, E., & Pennycook, A. (2019). Extending fairness and justice in language tests. In C. Roever & G. Wigglesworth (Eds.), *Social perspectives on language testing: Papers in honour of Tim McNamara* (pp. 29–45). Peter Lang.

Uludag, P., & McDonough, K. (2022). Exploring EAP instructors' evaluation of classroom- based integrated essays. *TESOL Journal*, *13*, e670. https://doi.org/10.1002/tesj.670

Youn, S. J., & Chen, S. (2021). Investigating raters' scoring processes and strategies in paired speaking assessment. In M. R. Salaberry & A. R. Burch (Eds.), *Assessing speaking in context: Expanding the construct and its applications* (pp. 107–131). Multilingual Matters. https://doi.org/10.21832/9781788923828-006