Developing An Oral Communication Test in English for Business Communication using Evidence-Centered Design

Andrias Susanto[1]
[1]Iowa State University // andrias@iastate.edu

ABSTRACT: An effort to create assessments that are valid, reliable, and aligned with specific competencies is through the use of evidence-centered design (ECD) (Mislevy, 2006). This report details how to develop an oral communication test (OCT) for the English for Business Communication (EBC) course designed for International Business Management (IBM) students using ECD. The process involves several key steps: defining the competencies and skills to be assessed, identifying the specific evidence required to demonstrate these competencies, and designing tasks that elicit this evidence. This method ensures that the assessment accurately measures students' abilities to communicate effectively in professional contexts. In developing the OCT tasks, two scenarios are selected, namely providing an opinion on a specific business topic as well as creating an argument for a particular investment phenomenon. By grounding the test in real-world tasks and clearly defined criteria, ECD helps ensure that the assessment provides meaningful and actionable insights into students' oral communication skills, better preparing them for future roles in the business and technology sectors.

Keywords: Evidence-centered design, international business management, oral communication test, AI-generated media

Introduction

The oral communication test (OCT) developed in this report is to assess International Business Management (IBM) students' oral communication ability and measure their proficiency and skills as required in professional business communication. My work experience as an English lecturer teaching the English for Business Communication (EBC) course of an IBM study program in a medium-sized university in Indonesia has contributed to creating the test, which is implemented using evidence-centered design. Therefore, this model can serve as a reference for an ECD test model for future test developments of the EBC course or other similar courses, mainly English for Academic Purposes (EAP). This type of test can later be used to inform the course instructors and curriculum designers for future course and program curriculum developments. By understanding the current business communication abilities of IBM students, faculty members can make informed adjustments to better equip future students in the IBM program, as well as develop the curriculum for other courses offered in the program.

This test is part of the final test for the EBC course, complementing the written communication portion of the whole EBC test. The final exam for IBM students will determine if they meet the course's passing standards as outlined in the curriculum. This curriculum is supported by books (Chan, 2020; Davis, 2010; Gillett, 2024; Sweeney, 2023), journal articles (Bhatia & Bremmer, 2012; Lee, 2018), newspaper articles (e.g., Forbes,

The Week, Entrepreneur), and feedback from IBM alumni working in business industries that require English for professional communication. The feedback ensures that the course materials and assessment are relevant to the real business world. The test is designed to conform with the needed knowledge and skills based on the evidence obtained from the resources, such as the study program curriculum, literature review, and needs analysis questionnaire.

Although ideally, a test like this should be a face-to-face and human-human interaction, due to the limited time, resources, and personnel available, this EBC-OCT utilizes human-computer interaction. In this context, the test results will determine the test-takes' abilities against the Indonesian national standard of course grading table (Table 1). In the test scoring table, “Point” refers to the percentage of mastery, ranging from 0% for no mastery to 100% for full mastery. The 4-point grade represents a high level of proficiency, typically corresponding to a mastery percentage of 85 or above. The A to E classifications are used to categorize overall performance, with A indicating excellent performance and E indicating poor performance.

Table 1. Conversion Table of Sources

Category	Grade	Point
A	4	85 or above
A-	3.7	80 - 84
B+	3.3	75 - 79
B	3	70 - 74
B-	2.7	65 - 69
C+	2.3	60 - 64
C	2	55 - 59
C-	1.7	51 - 54
D	1	41 - 50
E	0	40 or below

This article provides a snapshot of the development of a language test that leverages the current advancements in technology within the evolving landscape of language assessment. As educational assessment adapts to meet modern demands and challenges, new methods and technologies play a critical role. ECD is a key approach in this evolution, offering a structured way to design and deliver assessments that capture complex skills and knowledge often overlooked by traditional methods (Behrens et al., 2010).

ECD is grounded in the principles of evidentiary reasoning, focusing on using evidence to make informed decisions about what and how to assess. This approach helps create assessments that reflect real-world applications and essential learning objectives (Mislevy, 2003; 2005; 2006). Unlike traditional tests, ECD incorporates interactive tasks and ensures that research and assessment work together for curriculum refinement (Dursun, 2023). Therefore, the OCT developed through the process described in this article aims to implement technology while upholding the principle of validity.

Methodology

Evidence-centered Design

ECD consists of four key parts: domain analysis, domain modeling, conceptual assessment framework (CAF), and delivery system model. The process begins with domain analysis to identify the abilities or knowledge to be assessed by defining relevant constructs in a real-life context. Next, domain modeling narrows these constructs to those that can feasibly be measured within the given time and context, ensuring the validity of the assessment. The CAF encompasses five models: the student model to define learner attributes, the evidence model to specify necessary data, the task model to describe what tasks to be completed, the assembly model to combine tasks into the assessment, and the presentation model to outline task presentation. Finally, the delivery system model details the operational procedures for administering

the assessment, from task delivery to response collection and result processing. This comprehensive framework ensures assessments are valid, reliable, and aligned with the intended constructs.

Domain Analysis

Domain analysis is the foundational step, meticulously defining the knowledge, skills, and competencies essential for effective oral communication in a business context (Mislevy, 2003; 2005; 2006). By systematically analyzing the domain, the test's structure and content can accurately reflect the real-world demands and learning objectives, ensuring validity and relevance in assessing students' proficiency in this critical skill set. The domain analysis was completed by conducting a needs assessment through questionnaires and a literature review. In this study, the domain analysis was completed through a data collection instrument, namely a needs analysis questionnaire, distributed to the target population where business communications occur.

Needs analysis questionnaire

The questionnaire is one of the instruments used for updating the course syllabus and the other two sources mentioned previously. In particular, participants of this questionnaire are alumni of the IBM program of the university who had previously completed the EBC course during their undergraduate studies and currently work in a business industry that requires English for professional business communication purposes. I intend to use the data from this questionnaire to inform decisions about curriculum development and improvements according to the needs analysis and OCT results. By understanding how these alumni perceived the relevance and applicability of the English knowledge and skills obtained from the course in their professional endeavors, informed adjustments can be made to ensure that future students are better equipped for success in the business world.

The procedures of creating the questionnaire include studying the IBM program documentation to understand the vision/mission and curriculum. I also checked the list of topics taught in the EBC program, and the following topics were the ones that I was familiar with:

1) Job Application Letters
2) Resume
3) Job Interviews
4) Describing Diagrams
5) Marketing New Products or Services
6) Business Negotiation
7) Conflict Management
8) Expressing Ideas in Meetings
9) Writing Meeting Notes

Next, I browsed the internet (websites, journal articles, and other universities' syllabi) to read some relevant literature. As a result, four additional topics were considered, namely:

10) Business Emails
11) Business Plan
12) Business Contracts
13) Business Ethics

Afterward, I put all of the above topics (13 in total) as the starting points to create the constructs to be included in the questionnaire, with three items designed for each topic. For the topic of “Job Application,” students can either write a job application letter or create a tailored resume. They might do mock interviews, describe complex diagrams, or develop a marketing plan. Other tasks include role-playing business negotiations, analyzing workplace conflicts, presenting ideas in meetings, writing meeting notes, drafting business emails, creating a business plan, reviewing a contract, and analyzing ethical dilemmas.

Then, I drafted the questionnaire items based on Dörnyei and Taguchi (2009) and conducted a pilot with my graduate-level classmates three times (twice in class and once via email). Fourth, I asked an expert in class some questions for parts I was unsure of and revised the questionnaire items based on the feedback from my peers and professor. Lastly, I finalized the draft and conducted a pilot study.

The type of questionnaire items selected is a multiple-choice question (MCQ) format to collect responses, with one item presented in an open-ended question format. The MCQ was chosen mainly for efficiency and standardization of responses to determine the relevance of each of these topics to real-world business communication from a large number of participants within a short time (Dörnyei and Taguchi, 2009). They are written based on the 13 topics mentioned in the previous section, and each topic has three similar items (instead of only one item) to check the consistency of the participants' answers on each topic. The only item written in an open-ended format was to provide a space for the participants to express their comments, suggestions, or additional information that had not been covered in the questionnaire but is considered necessary for the course material or curriculum development. There were four main sections of the questionnaire: eight items in Section 1 covering demographic questions, 39 items in Section 2 on participants' viewpoints about the current materials, and one open-ended question if participants wished to add more information.

Pilot Study

Participants

The questionnaire was piloted by six classmates as part of a class project: five female and one male participants. The participants' first languages vary, namely Arabic, English, Indonesian, and Spanish. One of them remarked that he was not working when the questionnaire was administered; this participant then opted out of participation. All five participants confirmed that they used English in their workplace. Nevertheless, responses from two participants could not be analyzed further as most of the items were left unanswered.

Meanwhile, the other three responses were complete and broadly varied in terms of the scales selected. Brief discussions with each of them showed that all of the questionnaire items were clear enough, and no single item was considered problematic. One suggestion was that some of the statements were too wordy, and there was a suggestion to revise these items to make them more efficient. I revised the suggestion based on the feedback provided. The final version of the questionnaire can be accessed through the following link: https://forms.gle/Hy1JoeqCXr38FtaY8.

Nevertheless, to illustrate a possible data analysis, I used a random number generator menu on Excel for 300 participants and SPSS to calculate the Cronbach's Alpha value that represents the reliability level of the questionnaire. Even though this data analysis is hypothetical because the numbers generated are not from actual participants, it can give us an idea of what it might look like when administered to 300 alums of the IBM study programs.

In general, the questionnaire is reliable in measuring the participants' perspectives on the usefulness of the EBC topics. Using SPSS Software version 29.0, the reliability coefficient of .798 indicates that the measure has good internal consistency. This suggests that the items within the test are consistently measuring the same underlying construct.

Table 2. Descriptive Statistics of the Participants' Responses (N=300)

Statistics	Value
Mean	3.55
Medium	4.0
Minimum	1.0
Max	5.0

As mentioned, all the questionnaire items were developed based on the 13 topics determined from the IBM study program documentation and professional/academic literature review. In general, most of the statements are supported by the participants as the most selected responses were "agree" (score 4) and "strongly agree" (score 5). Table 2 presents data indicating that respondents generally hold a neutral to positive stance regarding the relevance of the materials to their professional communication practices. Moreover, regarding the significance of each topic, participants exhibit differing levels of agreement, with "job interviews" garnering the highest consensus and "expressing ideas in meetings" receiving the lowest.

From Table 3, the levels of importance or relevance for each topic are rated on a scale from 0, indicating strong disagreement with the topic's importance, to 5, indicating strong agreement. Therefore, a score of 3.82, for example, means that the topic is generally considered to be of moderate importance or relevance.

Table 3. Participants' Responses on Each of the EBC Topics

Topic	Average Response Value
Job Application Letters	3.82
Resume	3.82
Job Interviews	4.08
Describing Diagrams	3.85
Marketing New Products or Services	3.97
Business Negotiation	3.76
Conflict Management	3.59
Expressing Ideas in Meetings	2.89
Writing Meeting Notes	2.98
Business Emails	3.08
Business Plan	3.00
Business Contracts	2.97
Business Ethics	3.00

It can be concluded from the results of the need-analysis questionnaires that most (10 out of 13) of the topics taught in the course were considered relevant by the time the responses were collected. The three topics with the lowest margins: "Expressing Ideas in Meetings," "Writing Meeting Notes," and "Business Contracts," can be discussed with the IBM Program Coordinator to decide whether or not they should be kept or discarded. The other ten would definitely be kept, and those with a value of 3.50 or below can be revisited to be more in line with the relevant business situations.

Next, the topics selected for the test are made based on the most popular topics, including "Marketing New Products/Services," "Business Negotiation," "Describing Diagrams," and "Conflict Management." The first two are used for the oral communication test (OCT), while the others are for the written communication test (WCT). In this project, only the OCT is described. The first three topics also obtained a high level of importance, but they are intended to prepare the students to apply for jobs. Therefore, the four selected topics are most relevant on this occasion, preparing them to use English in various business communication contexts.

Domain Modeling

Investigating what is useful in various business contexts helps determine what topics need to be taught to better prepare the students for what they need to perform in their future workplace. The four selected topics are prioritized to be mastered by the IBM program graduates, thus ensuring they acquire essential skills tailored to the demands of their prospective careers. Moreover, this targeted approach enhances graduates' marketability and cultivates a culture of innovation and adaptability essential for success in dynamic industries. This deliberate alignment of educational objectives with industry needs underscores the program's commitment to producing graduates who are academically proficient and well-equipped to make meaningful contributions to their respective fields. With this focused curriculum, students are expected to be familiar with each topic, gaining practical insights and hands-on experience. For instance, they learn how to provide arguments to evaluate investment opportunities, using standard English sentences and expressions pertinent to their work environment.

Data Analysis

This section explains the results of analyzing the questionnaire and will later be used for developing the test once the test constructs are determined. In order to develop the test, the following models are addressed to fulfill all the elements of CAF.

Conceptual Assessment Framework

CAF is a comprehensive blueprint that outlines the components and structure of an assessment system. It guides the design and development of assessments to ensure they are aligned with the intended constructs and validly measure what they are supposed to measure. The CAF consists of five interconnected models, and for this test, each of the following models is addressed.

Student Model

The EBC-OCT is intended to measure test-takers' oral communication skills necessary for success in professional business settings. Students will demonstrate the ability to understand messages delivered in a professional business communication environment by identifying the main points of the topic and responding suitably using acceptable relevance to the topic, grammar and vocabulary, fluency, and pronunciation. This will be indicated by their ability to (a) explain their viewpoint in a monologue form on the topic being discussed as a response to a colleague's opinion and (b) construct an argument to either accept or decline a proposal based on the information provided in the reading and listening materials.

Task Model

Oral communication tasks require test-takers to listen and speak, respond to prompts, and provide explanations and arguments. The tasks are structured to progress from simpler to more complex scenarios, evaluating students' ability to integrate information, respond with relevance, and demonstrate clarity in their speaking. Evaluation criteria include relevance, grammar, vocabulary, fluency, and pronunciation. The assessment aligns with real-world business contexts, presenting test-takers with materials such as videos, reading texts, and audio clips and analyzing their responses to provide a comprehensive view of their communication skills. Therefore, the test should include two integrated tasks: (1) combining listening and speaking and (2) integrating reading, listening, and speaking. These tasks allow test-takers to demonstrate their ability to articulate opinions in a conversational context and make decisions based on provided resources, reflecting common professional business scenarios. Test-takers are given a relatively short time for preparation and can take notes. The test can be recorded via a video conferencing platform for accurate rating, with further details available in the Presentation Model section.

Evidence Model

Rater norming

Before rating the EBC-OCT speaking section, both raters (course instructor and IBM program coordinator), who should possess at least a master's degree in TESOL/Applied Linguistics, must convene. They should systematically review scoring criteria and scales to establish a shared understanding of proficiency levels. They should also conduct a practice session, scoring three sets of speaking samples from the program's collection to compare evaluations, resolve discrepancies, and ensure consistent application of criteria. After a week, they are to meet again for a final practice session to confirm alignment in their rating approach. Throughout this process, they should discuss and justify their rating decisions to ensure accuracy and reliability in assessing student performances.

Scoring

The rating is conducted only after all the test-takers have completed the test. Each rater will have access to the recordings of all the students and then rate each performance in their own preferred time and space, but the final scores are to be submitted to the course instructor a maximum of two weeks after the test is administered. Next, because the assessment consists of two sections: this EBC-OCT and a writing test section, the score provided for this OCT accounts for 50% of the students' total final test score, with the remaining half being assigned from their performance in the writing test.

In regard to this, a holistic score is provided to help with the final score calculation, but a set of analytic scores is also designed so that the test-takers are able to identify specific areas for improvement. The test-

takers are then expected to understand their strengths and weaknesses in each aspect of their speaking proficiency. The same weights are assigned to each of the criteria: relevance to the topic (25%), grammar and vocabulary (25%), fluency (25%), and pronunciation (25%). The possible minimum score would be 0, and the possible maximum score is 100. As an illustration, Student A gets these two sets of scores:

From Rater 1:

Scoring	Relevance (25%)	Grammar/Vocab (25%)	Fluency (25%)	Pronunciation (25%)	Overall (100%)
Raw Score	5	5	4	4	-
Adjusted score based on percentage	1.25	1.25	1	1	4.5 = 90%

From Rater 2:

Scoring	Relevance (25%)	Grammar/Vocab (25%)	Fluency (25%)	Pronunciation (25%)	Overall (100%)
Raw Score	4	4	4	3	-
Adjusted score based on percentage	1	1	1	0.75	3.75 = 75%

From the above sets of scores, the holistic scoring assigned to Student A would be the average of the two scores of 90% and 75%, which is 82.5%.

Next, because performance-based assessment is sensitive to several factors (facets) that may influence score assignment, Many-Facet Rasch Measurement (MFRM) statistical analysis is then used to provide adjusted scores that are considered to be closely representative of the test-takers' true abilities (Ockey, 2021). Three facets taken into consideration are scores, raters, and tasks. For this test analysis, the rating scale model (RSM) is preferred over partial credit modeling (PCM). However, due to the article's limitations, this is not explained further; future researchers can follow procedures using MFRM.

Findings

Assembly Model

The OCT consists of two integrated tasks: (a) listening to speaking and (b) reading and listening to speaking. These tasks align with the construct outlined in the Student Model and aim to evaluate students' proficiency in articulating their viewpoints on a business topic. In Task 1, students respond to prompts from a humanoid video character, while in Task 2, they construct arguments based on provided reading material and audio. The tasks progress from simpler to more complex, transitioning from speaking based on one input to incorporating two sources. A single-speaking rubric assesses overall performance across both tasks.

Presentation Model

In this test, since it is estimated that a class consists of between 25 and 30 students in one cohort (1 class per academic year), two people are required to supervise the test (one is the course instructor and one non-academic staff). The program typically follows a test schedule that determines the dates and times allocated for specific tests.

The final version of the test can be accessed here: EBT-OCT. Prior to starting the test, the following procedures are to be closely followed by the test-takers. This set of rules is sent to each of the students' email one day before the test and is put on the computer lab's door for attention.

Test-takers should leave all personal items outside the computer lab, except for two pens and a transparent water bottle, if needed. Upon check-in, test-takers are checked by a room proctor and then seated in designated cubicles where they put on computer headsets. The test administration is centralized, with instructions given through the main lab computer, and test-takers are to start the test only when instructed. During the technical setup, test-takers must ensure their computers and audio systems are functioning properly, reporting any issues to room proctors immediately. The cubicles are designed to minimize noise interference, and headsets have noise-canceling features to allow test-takers to focus on speaking tasks without disturbance. The test lasts less than ten minutes, and once it begins, test-takers cannot leave the room until it is completed, except in an emergency. The scoring rubric for assessing performance is available online (link: EBC-OCT Scoring Rubric). The test summary can be seen below.

Figure 1. Test Summary

Total test time	12 minutes
Number of parts	2
Test-takers interact with	a video of a humanoid character generated by an artificial-intelligence (AI) tool for the first task (i.e. *HeyGen* app*)*, and a reading material and an audio that is generated from a text-to-speech (TTS) technology (i.e. *ElevenLabs* app).
Part 1 (Listening Speaking)	
Total time	5 minutes (30 seconds for asking the question, 30 second preparation time, 4-minute response time)
Number of question	1 question prompt
Task content	One-on-one interaction with the computer. The computer briefly explains a case then asks a question in relation to that and test-taker may take notes if they wish, test-taker prepare the answer for 30 seconds, and they are expected to speak for four minutes in response to the question
Part 2 (Reading and Listening Speaking)	
Total time	7 minutes (1 minute for reading the paragraph shown on the screen, 1.5 minutes for listening to the audio, a 30 second preparation time, 4-minute response time)
Number of question	1 question prompt
Task content	One-on-one interaction with the computer. The reading material are shown on the computer screen and the listening material is presented immediately afterwards and test-taker may take notes if they wish, test-taker prepare the answer for 30 seconds, and they are expected to speak for four minutes in response to the question

Delivery System Model

Although it is possible to conduct this test online, a detailed account of the process of administering the test in person is provided here. The delivery system model for administering the OCT involves five key steps. The test takers need to complete both tasks; thus, once the link is accessed, they need to follow the instructions and time lapse determined as shown on the monitor screen for both tasks. Next, as the test

takers directly speak on a microphone mounted to the headset, the oral responses are collected from their recorded speeches and stored in the main server computer. After that, once the test is completed, the Zoom app will be able to store each of the recordings and label them according to the test-takers' names and computer labels; all are stored in separate folders containing two files, namely video and audio recordings. For the summary scoring process, an Excel file will be created to list all the names and links to their speaking performances and will be distributed to the two raters separately to ensure independent scoring. Once the raw scores from both raters are collected into a single file with the two lists of scores next to each other, they are then analyzed by one of the examiners using MFRM statistical analysis, who will later update the scoring record. Scores are then combined with those collected from the Writing Exam, not discussed in this article, to determine if the students meet the scoring standards summarized in Table 1 using a criterion-referenced assessment framework. The full test instruction can be accessed here: EBC-OCT Test Instruction.

Conclusion

Employing ECD to develop an OCT for the EBC course ensures that the assessment is both valid and reliable, aligning with the specific competencies needed by IBM students. This approach involves defining the essential competencies, identifying the evidence required to demonstrate these skills, and designing tasks that effectively elicit this evidence. By incorporating real-world scenarios, such as expressing opinions on business topics and constructing arguments for investment phenomena, the OCT provides a practical and relevant measure of students' communication abilities in professional settings. This method not only enhances the accuracy of the assessment but also equips students with the necessary skills for their future careers in the business and technology sectors.

Additionally, ECD's structured framework allows for continuous improvement and adaptation of the assessment to meet evolving industry standards and educational goals. By regularly reviewing and updating the competencies and tasks, educators can ensure that the OCT remains relevant and effective in preparing students for the dynamic demands of the global business environment. This iterative process fosters a deeper understanding of the competencies required for success and supports the ongoing development of students' communication skills.

Limitations

This study faces several limitations, primarily due to the constrained time and resources available for its implementation in real classroom settings with actual participants. The development and validation of the OCT using ECD require extensive planning, coordination, and iterative testing, which were not feasible within the scope of this project. Additionally, the lack of access to a diverse group of students and instructors for pilot testing limits the ability to gather comprehensive feedback and make necessary adjustments to the assessment tasks. These constraints may affect the generalizability and applicability of the findings, as the OCT has not been fully tested in varied educational contexts. Consequently, further research and practical trials are needed to refine the assessment and ensure its effectiveness in accurately measuring students' oral communication skills in real-world business scenarios.

The Author

Andrias Tri Susanto is a double degree Ph.D. student in Applied Linguistics and Technology as well as Human-Computer Interaction at Iowa State University. He holds a master's degree in Teaching English to Speakers of Other Languages (TESOL) from the University of Leeds in the United Kingdom. He has been involved in the world of teaching English since 2007 with students from various academic backgrounds and

teacher training since 2014. His main interests include computer-assisted language learning (CALL), and language assessment, as well as mixed reality (XR) and artificial intelligence (AI) for second language learning.

References

Bhatia, V. K. & Bremmer, S. (2012). English for business communication. *Language Teaching 45*(4), 410-445.

Chan, M. (2020). *English for business communication*. Routledge.

Dörnyei, Z. & Taguchi, T. (2009). *Questionnaires in second language research*. Routledge

Davis, K. W. (2010). *The McGraw-Hill 36 hour course in business writing and communication* (2nd ed.). McGraw Hill.

Gillett, A. (2014). *Speak business English like an American: Learn the idioms and expressions you need to succeed on the job!* Language Success Press.

Lee, Y. (2018). Dynamics of symmetrical communication within organizations: the impacts of channel usage of CEO, managers, and peers. *International Journal of Business Communication 59(*1), 3-21.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design.* (Research Report 03-16), Princeton, NJ: Educational Testing Service.

Mislevy, R. J., & Risconscente, M. (2005). *Evidence-centered assessment Design: Layers, structures, and terminology.* SRI International.

Ockey, G. (2021). Item response theory and many-facet Rasch measurement. In G. Fulcher and L. Harding. *The Routledge handbook of language testing*, 2nd ed. pp. 462 – 476.

Sweeney, S. (2003). *English for business communication*. Cambridge University Press