Research Article

Analytic Assessment of English as a Foreign Language Oracy and Student Perceptions in Teacher Education

Miriam Elizabeth Cid Uribe[1], Gabriel Arturo Farías Rojas[2], Francisco Javier Orellana González[3]

[1]PhD, Department of Linguistics and Literature, Faculty of Humanities, Universidad de Santiago de Chile (USACH), Chile, Email: miriam.cid.u@usach.cl, ORCID: https://orcid.org/0009-0003-0315-569X
[2]PhD, Faculty of Communications and Arts, Universidad de Las Américas (UDLA), Chile, Email: gfarias@udla.cl, ORCID: https://orcid.org/0000-0002-7642-1237
[3]MA, Faculty of Letters, Pontificia Universidad Católica de Chile (PUC), Chile, Email: fjorella@uc.cl, ORCID: https://orcid.org/0000-0003-4992-3327

Abstract

Objective: To evaluate dimension-level patterns in English as a foreign language oral assessments and students' perceptions of a paired dialogue task. Methods: We examined a descriptive synthesis of individual rubric assessments and a separate valuation chart from a classroom activity involving 20 university teacher education students. The task was a three-minute dialogue requesting a medical appointment. The analytic rubric covered ten dimensions and four achievement categories; the valuation scale contained ten statements and five response options. Results: Organization was summarized at category four; mechanics, segmental production, vocal delivery, and discourse participation at category three; grammar, vocabulary, and fluency at category two; and prosody and connected speech at category one. The valuation chart showed predominantly favorable responses, especially for communication and confidence. The rubric synthesis did not report its aggregation procedure, and the valuation chart did not provide a verified respondent denominator. Conclusion: Positive perceptions of the dialogue task coexisted with uneven achievement categories in the descriptive summary. These forms of evidence are complementary but cannot be linked at the individual level. The findings support criterion-specific feedback and further investigation; they do not demonstrate proficiency gains or causal effects.

Descriptors: Speech; Educational Measurement; Communication; Teaching; Students

Resumen

Objetivo: Evaluar las dimensiones de la oralidad en inglés como lengua extranjera y las percepciones estudiantiles sobre una actividad dialógica. Métodos: Se examinó una síntesis descriptiva de evaluaciones individuales con rúbrica y un gráfico independiente de valoración, correspondientes a una actividad con veinte estudiantes universitarios de formación docente. La tarea consistió en un diálogo de tres minutos para solicitar una cita médica. La rúbrica comprendió diez dimensiones y cuatro categorías; la escala contenía diez afirmaciones y cinco opciones. Resultados: Organización se situó en la categoría cuatro; mecánica, producción segmental, emisión vocal y participación discursiva en la tres; gramática, léxico y fluidez en la dos; prosodia y habla conectada en la uno. El gráfico mostró valoraciones predominantemente favorables, especialmente sobre comunicación y confianza. No se dispone del procedimiento de agregación de la rúbrica ni del denominador verificado de la escala. Conclusión: Las percepciones positivas coexistieron con categorías heterogéneas de desempeño en la síntesis descriptiva. Ambos instrumentos aportan evidencias complementarias, pero no vinculables individualmente. Los resultados orientan retroalimentación específica y nuevas investigaciones; no demuestran mejoras lingüísticas ni efectos causales.

Descriptores: Habla; Evaluación Educacional; Comunicación; Enseñanza; Estudiantes

Resumo

Objetivo: Avaliar as dimensões da oralidade em inglês como língua estrangeira e as percepções dos estudantes sobre uma atividade dialógica. Métodos: Examinamos uma síntese descritiva de avaliações individuais por rubrica e um gráfico independente de apreciações, referentes a uma atividade com vinte universitários em formação docente. A tarefa consistiu em um diálogo de três minutos para marcar uma consulta médica. A rubrica continha dez dimensões e quatro categorias; a escala incluía dez afirmações e cinco alternativas. Resultados: Organização recebeu categoria quatro; aspectos de apresentação, produção segmental, emissão vocal e participação discursiva, categoria três; gramática, vocabulário e fluência, categoria dois; prosódia e fala encadeada, categoria um. O gráfico indicou avaliações predominantemente favoráveis, sobretudo sobre comunicação e confiança. O procedimento de agregação da rubrica e o denominador da escala não estão disponíveis para verificação. Conclusão: Percepções positivas coexistiram com categorias heterogêneas de desempenho na síntese descritiva. Os instrumentos oferecem evidências complementares, sem permitir

correspondências individuais. Os achados orientam devolutivas específicas e novas pesquisas; não demonstram ganhos de proficiência nem efeitos causais.
Descritores: Fala; Avaliação Educacional; Comunicação; Ensino; Estudantes

Introduction

Oral communication in English as a foreign language teacher education extends beyond grammatical correctness. Speakers must coordinate linguistic resources, pronunciation, temporal organization, and interaction in context. Cognitive scholarship emphasizes attention, memory, and the organization of knowledge, while constructivist accounts locate learning in meaningful activities and interaction.[1,2,3,4] Analytic assessment can make distinct dimensions visible, but a single assessment does not reveal the cognitive mechanisms that produced performance.

Paired role-play can provide a purposeful context for negotiated meaning, coordination of turns, and observation of peers. Accounts of constructive and networked learning emphasize active participation, while social-learning theory draws attention to attention, retention, practice, and motivation.[5,6,7] These are reasons to consider dialogue pedagogically relevant, not proof that participation produced measurable gains in this setting.

The perceived value of speaking activities can also be shaped by the climate in which participation occurs. Scholarship on socioemotional learning and communities of practice identifies constructive dialogue, trust, inclusion, and collective reflection as relevant to educational participation.[8,9,10,11,12] These themes support looking at student appraisals alongside, but not in place of, criterion-based observations. They do not imply that the present study measured motivation, anxiety, interpersonal equality, or social-emotional development directly.

The communicative contexts that learners encounter extend beyond a single classroom. Connectivist scholarship addresses links among people and information sources, while critical accounts caution against assuming that access to networks guarantees meaningful learning.[13,14,15] Perspectives on media and cultural industries likewise foreground the unequal availability of resources and the diverse linguistic and cultural experiences learners bring to teaching.[16,17,18] These contributions provide possible directions for future oral-language instruction; no online activities or effects of educational technology were evaluated here.

Neuroeducation addresses cognition, memory, attention, emotion, and variation among learners.[19,20] Such knowledge can inform the sequencing and accessibility of classroom practice, but it does not allow a researcher to infer a particular neurological difficulty from a rubric score. Likewise, a descriptive judgment about pronunciation should not be interpreted as a diagnosis of the cognitive processes underlying that performance.

The research problem concerns how criterion-level summaries of oral-performance assessments can be interpreted alongside students' evaluations of a dialogic task. Performance judgments and perceived usefulness are different constructs; agreement with a statement about a task does not demonstrate improvement in speaking ability. The research question asks which oracy dimensions are emphasized by a descriptive synthesis of individual assessments, how these patterns compare with students' appraisals, and what the comparison implies for assessment practice. Our provisional inductive proposition was that generally favorable appraisals may coexist with uneven achievement categories across dimensions of speaking.

The objective was to evaluate the oral-performance dimensions summarized from individual assessments and their relationship, at a descriptive level, to perceived educational value. Specifically, we sought to identify the assessment dimensions, describe the displayed patterns, apply educational concepts, analyze conceptual convergences and discrepancies, and reflect on pedagogical and methodological implications. We did not aim to measure learning gains or establish correlations between individual ratings and questionnaire responses.

Methods

We conducted a descriptive, nonexperimental classroom study involving 20 students in the second and third years of an English teacher education program (seven men and thirteen women). The students' English proficiency was characterized as intermediate; a specific standardized placement level was not established in the analytic record. Students prepared a three-minute dialogue in pairs to request a medical appointment and performed it before peers and the instructor at the following class.

Individual oral performances were assessed using an analytic rubric with ten criteria: organization and coherence; mechanics and supporting materials; grammar; vocabulary; segmental production; prosody;

connected-speech simplification; fluency; vocal delivery and paralinguistic features; and discourse-pragmatic interaction. Each criterion contained four ordered descriptors, from poor (1) to excellent (4). Individual judgments were condensed into a criterion-level descriptive synthesis reporting one category per dimension. The available summary does not specify the aggregation rule; its category values are therefore treated as ordinal summary labels, not group means, achievement frequencies, or student-level scores.

Perceptions were collected using a ten-statement scale with five response options, ranging from strongly disagree to strongly agree. Items concerned perceived support for speaking, grammar, vocabulary, collaboration, creative language use, confidence, awareness of strengths and limitations, vocal resources, argumentation, and real-life communication. We examined the stacked response chart qualitatively, identifying predominant categories and visible neutral or unfavorable segments. The chart does not provide a verified denominator or the original item frequencies; no exact percentages were reconstructed.

To preserve the distinction between reporting units, we treated the rubric synthesis as one dimension-level display and the valuation chart as a separate item-level display. We did not calculate a total proficiency score from the ordinal summary values. Nor did we estimate a cohort proportion for any assessment category or infer participant-level convergence from the two displays. This reporting strategy follows directly from the information retained in the summaries, rather than assuming that a chart permits the reconstruction of its underlying records.

We compared the two summary displays at the criterion and item levels and interpreted their conceptual convergence or tension through published learning and assessment scholarship. This was interpretive triangulation, not a statistical correlation or a separate qualitative coding analysis. No pretest–posttest, interrater reliability coefficients, validation statistics, speech recordings, or individual-level instrument linkage formed part of the dataset analyzed.

Before collecting data, we informed students about the study's educational purpose and the voluntary nature of participation and obtained verbal informed consent. We excluded personally identifying information from the analysis and report, and we do not identify the teaching institution. We report only anonymized descriptive summaries.

Results

Analytic rubric synthesis

The rubric synthesis distinguished achievement categories across ten oral-performance dimensions (Table 1). Organization and coherence received category 4 (excellent). Mechanics, segmental production, vocal delivery, and discourse pragmatics received category 3 (good); grammar, vocabulary, and fluency received category 2 (acceptable); and prosody and connected-speech simplification received category 1 (poor). Thus, the summary contained one category 4, four category 3, three category 2, and two category 1 judgments. The codes add arithmetically to 24, but this sum is not an individual score, cohort mean, or standardized proficiency index.

[Table 1 near here]

The contrast between segmental production (category 3) and prosody and connected speech (category 1) suggests dimensions requiring differentiated diagnostic attention. The synthesis does not indicate how many participants attained each level or whether those categories co-occurred in particular students.

Valuation scale

The rubric labels also differentiate delivery from phonological organization. Vocal delivery and paralinguistic behavior were summarized at a higher level than prosodic performance, which encompasses aspects such as stress and intonation. These categories therefore should not be merged into a single pronunciation outcome. Because the displayed values are a descriptive synthesis, they identify distinctions requiring further investigation rather than a quantified distribution of error types across participants.

[Table 2 near here]

The valuation chart indicated predominantly favorable appraisals across its ten statements (Table 2). Strong agreement was prominent for speaking development, confidence, vocal resources, and real-life communication. For the real-life communication item, only the two agreement categories were visible. Grammar, vocabulary, collaboration, and awareness of strengths and limitations also attracted favorable responses. The argumentation and counterargumentation item displayed the largest neutral segment; small disagreement segments appeared

elsewhere. These patterns reflect perceived task usefulness rather than independently measured proficiency gains.

Discussion

The two instruments offer complementary rather than interchangeable evidence. The rubric synthesis orders oral-performance dimensions, whereas the valuation scale describes perceived opportunities afforded by the task. Their juxtaposition is consistent with the working proposition that students may value a dialogue while the assessment summary highlights uneven features of speaking. It cannot establish correspondence within individual learners because scores and responses are not linked.

Constructivist and sociocultural perspectives support contextualized interaction, modeling, feedback, and guided participation.[3,4,2,6] The appointment-request task afforded a recognizable purpose and turn-taking, but neither the task format nor the positive valuation responses demonstrate transferable language development. In particular, lower synthesized categories for prosody and connected speech should not be collapsed into a general pronunciation judgment when segmental production has a higher category. The distinction provides a rationale for future model-based practice and explicit feedback on stress, rhythm, intonation, reductions, and linking. Such activities are recommendations, not interventions evaluated here.

Cognitive theories allow a cautious interpretation of why oral production may involve simultaneous demands on attention, lexical access, sentence construction, and real-time response.[1,2] Yet the data contain no recordings, timed measurements, or process indicators with which to test a processing-load explanation. It would therefore be inappropriate to infer that grammatical selection caused prosodic weakness, or that the observed summary categories represent a developmental sequence. A follow-up study would need speech samples, clearly defined phonological measures, and repeated observations to address those possibilities.

The rubric also includes mechanics and supporting materials. These may support task delivery, yet they are not equivalent to linguistic competence. Moreover, the category codes do not establish comparable measurement intervals or justified equal weighting. Criterion-specific descriptors and opportunities for formative self- and peer-assessment could improve transparency and participation.[9,10] Trust, emotional support, and learners' cultural contexts are relevant to implementation, but the valuation scale is not a validated measure of anxiety, confidence change, or inclusion.[8,11,17]

The neutral segment for argumentation warrants attention to task–instrument alignment. Requesting an appointment is primarily transactional and need not elicit sustained argument or counterargument; the pattern cannot be interpreted as evidence of argumentative inability. Future tasks should distinguish transactional speech, persuasive discourse, pronunciation, and fluency. Networked examples and audio-supported rehearsal may extend instructional opportunities, provided that access and structured guidance are considered.[13,14,15,16,18]

The study has important limitations. The summary rubric contains no individual score distributions or documented aggregation formula; the questionnaire chart lacks raw counts and a verified response denominator. There are no paired records, independent ratings, recordings, or repeated performances for validation. These constraints preclude estimates of difficulty prevalence, group mean performance, associations, causal effects, and demonstrated gains. They do not prevent descriptive interpretation, provided the unit of reporting remains explicit.

Conclusion

A descriptive synthesis of individual oral assessments placed organization above several linguistic and phonological dimensions, particularly prosody and connected speech. The valuation chart, meanwhile, showed generally favorable perceptions of dialogic practice. The findings address the study's evaluative objective and remain consistent with its provisional proposition, but do not establish individual relationships between ratings and perceptions. Their educational contribution lies in separating performance evidence from perceived usefulness and identifying criteria for more focused feedback.

The study also illustrates a reporting principle with wider application to classroom assessment. A criterion-level summary is informative when it is clearly labeled and interpreted as a summary, but its meaning changes if readers assume that its numbers are group averages or individual scores. We therefore retained the ordinal

achievement labels and avoided reconstructing unreported statistics. The positive valuation results are similarly interpreted as appraisals of a learning opportunity rather than measurements of attained proficiency.

Future research should document the rule used to synthesize scores, retain anonymized individual ratings and response counts, and evaluate agreement among independent assessors. Repeated oral performances and instrument validation would permit stronger claims about development. Until then, contextualized role-play is best viewed as one setting for eliciting and assessing oral communication rather than proof of improved proficiency.

Acknowledgments

We acknowledge the collaborative work in this investigation which belongs to the Chilean study group on Literacy and Oracy; the aforementioned authors in this research article are members of this group, born in the English Pedagogy program from USACH around 10 years ago.

References

1. Arias Gallegos WL. Antecedentes, desarrollo y consolidación de la psicología cognitiva: un análisis histórico. Tesis Psicol. 2021;16(2):172-198. doi:10.37511/tesis.v16n2a9.

2. Manjarrés Leal EP. Fundamentos del cognoscitivismo y sus aportes en la praxis educativa. REDHECS. 2021;29(19):93-106.

3. Ortiz Granja D. El constructivismo como teoría y método de enseñanza. Sophia, Colección de Filosofía de la Educación. 2015;(19):93-110. doi:10.17163/soph.n19.2015.04.

4. Ronquillo Murrieta GV, De Mora Litardo E, Bohórquez Morante AM, Padilla Plaza JL. Modelo constructivista y su aplicación en el proceso de aprendizaje de los estudiantes. J Sci Res. 2023;8(Supl 2):256-273. doi:10.5281/zenodo.10420471.

5. Gortaire Díaz D, Beltrán Moreno M, Mora Herrera E, Reasco Garzón B, Rodríguez Torres M. Constructivismo y conectivismo como métodos de enseñanza y aprendizaje en la educación universitaria actual. Cienc Lat Rev Cient Multidiscip. 2022;6(6):14046-14058. doi:10.37811/cl_rcm.v7i1.4672.

6. Villagómez-Cabezas AV, Bonilla-Andrango LJ, Bonilla-González GP, Torres-García TD. El aprendizaje social de Albert Bandura como estrategia de enseñanza de educación para la ciudadanía. Polo Conoc. 2023;8(5):1286-1307. doi:10.23857/pc.v8i5.5644.

7. Ojeda-Martínez RI, Becerill Tello MN, Vargas LA. La importancia del aprendizaje social y su papel en la evolución de la cultura. Rev Argent Antropol Biol. 2018;20(2):1-13. doi:10.17139/raab.2018.0020.02.02.

8. Valenzuela L, Miño M. Aprendizaje social y emocional: un camino a construir en la educación superior. Cienc Lat Rev Cient Multidiscip. 2021;5(3):2952-2963. doi:10.37811/cl_rcm.v5i3.499.

9. Hernández-López CA, Jiménez-Álvarez T, Araiza-Delgado IY, Vega-Cueto M. La escuela como una comunidad de aprendizaje. Ra Ximhai. 2015;11(4):15-30.

10. Escobar N. El aprendizaje en la escuela: ¡Ahí está el poder! Acción Pedagógica. 2012;21:64-75.

11. Eirín Nemiña R. Las comunidades de aprendizaje como estrategia de desarrollo profesional de docentes de educación física. Estud Pedagóg. 2018;44(1):259-278. doi:10.4067/S0718-07052018000100259.

12. Villanueva Chumpitaz AA, Riveros Cones LK, Camacho Porras CA. Comunidades de aprendizaje en el contexto educativo: un artículo de revisión sistemática. Rev InveCom. 2026;6(2):e602006. doi:10.5281/zenodo.15750887.

13. Siemens G. Conectivismo: una teoría de aprendizaje para la era digital [documento traducido]. Leal Fonseca DE, traductor. 2007. Obra original publicada en 2004.

14. Gutiérrez Campos L. Conectivismo como teoría de aprendizaje: conceptos, ideas y posibles limitaciones. Rev Educ Tecnol. 2012;(1):111-122.

15. Sobrino Morrás Á. Aportaciones del conectivismo como modelo pedagógico post-constructivista. Propuesta Educ. 2014;23(42):39-48.

16. Rama C. El nuevo paradigma de educación y el papel de las industrias culturales. Cienc Docenc Tecnol. 2004;15(28):13-23.

17. Rey G. Industrias creativas y educación: la escuela en medio de las transformaciones culturales. Educ Ciudad. 2018;(35):13-24.

18. Castro Chaparro MO. La educación como industria cultural: una aproximación conceptual. Rev Int Inst Pensam Lib. 2024;1(2):17-76. doi:10.51660/ripl.v1i2.35.
19. Bullón Gallego I. La neurociencia en el ámbito educativo. Rev Int Apoyo Inclus Logop Soc Multicult. 2017;3(1):118-135.
20. Freire Mora MA, Torres Merino JD, Navarro Barzola GL, Campoverde Delgado ME, Orellana Len VA. La neuroeducación y su impacto en las estrategias de enseñanza. Cienc Lat Rev Cient Multidiscip. 2025;9(3):5001-5021. doi:10.37811/cl_rcm.v9i3.18128.

Table 1. Descriptive categories in the synthesis of individual oral assessments

Assessment dimension	Category	Descriptor
Organization and coherence	4	Excellent
Mechanics and supporting materials	3	Good
Grammatical structures	2	Acceptable
Lexical range and appropriateness	2	Acceptable
Segmental production	3	Good
Prosodic performance	1	Poor
Connected-speech simplification	1	Poor
Fluency	2	Acceptable
Voice delivery and paralinguistic features	3	Good
Discourse and pragmatics	3	Good

Note. These are ordinal categories in a descriptive synthesis of individual assessments. The rule used to summarize individual ratings is not available. Categories are not individual scores, cohort means, or frequencies.

Table 2. Qualitative patterns displayed in the task valuation chart

Valuation statement	Largest visible response	Other visible segments
Developing speaking	Strongly agree	Agree; small strongly disagree segment
Supporting grammar use	Agree	Substantial strongly agree; small strongly disagree
Expanding specific vocabulary	Strongly agree	Agree; small disagree segment
Reinforcing collaboration	Strongly agree	Agree; small neutral segment
Using everyday language creatively	Strongly agree	Agree; neutral and small strongly disagree segments
Building performance confidence	Strongly agree	Agree; small strongly disagree segment
Recognizing strengths and limitations	Agree	Substantial strongly agree; small strongly disagree
Using vocal resources	Strongly agree	Agree; small disagree segment
Argument and counterargument	Agree	Strongly agree; largest neutral segment
Language use in real-life situations	Strongly agree	Agree only

Note. Categories describe relative visual prominence, not exact percentages or verified response counts. The chart does not supply a respondent denominator.