

ARTICLE



Teachers' assessment journeys in a virtual exchange

Hülya Tuncer, Çukurova University

Eric Hagley, Griffith University*

Andrew Johnson, Future University Hakodate

Matthew Cotter, Hokusei Gakuen University

Abstract

Virtual Exchange (VE) has been gaining popularity amongst Foreign Language (FL) teachers as it allows students the opportunity to use the FL they are studying in real-world situations with peers around the world. While VE allows for extensive language production and the development of intercultural competence, the assessment of these aspects within VE poses significant challenges for teachers. Through the use of a qualitative descriptive design, this paper examines the assessment path eight teachers from six countries took as their students participated in a VE. It outlines the reasoning behind the assessment designs they chose, how they implemented them, the challenges they faced and the manner in which they overcame them. The teachers were surveyed and interviewed before and after participating in an eight-week, mainly asynchronous VE. Interviews were recorded, transcribed and then thematically analyzed as were the open-ended survey questions. Results show objectives of developing communicative and intercultural competence were major considerations when designing and implementing assessment of student participation in the VE. Most participants favored formative approaches and incorporated both quantitative and qualitative assessment to encourage intercultural communication. Analysis resulted in recommendations to help practitioners enhance their assessment of students' participation in VE.

Keywords: assessment; virtual exchange; foreign language learning; intercultural competence

Language(s) Learned in This Study: English

APA Citation: Tuncer, H., Hagley, E., Johnson, A., & Cotter, M. (2026). Teachers' assessment journeys in a virtual exchange. *Language Learning & Technology*, 30(1), 1–25.

<https://doi.org/10.64152/10125/73677>

Introduction

Foreign Language (FL) learning improves cognitive abilities, mental health, employability, academic achievement, communicative and intercultural competence (Fox et al., 2019). Furthermore, a willingness to communicate in the FL through actual cross-cultural interactions is greater when students can interact with peers in other countries (Nguyen et al., 2024; Roarty & Hagley, 2021). Therefore, FL educators should ensure classrooms are places where use of the FL in real-world international settings can take place, but this can be difficult to implement when students are primarily in classrooms with others from the same or similar cultural and linguistic backgrounds.

Virtual Exchange (VE) has been implemented across diverse educational contexts, enabling cross-cultural interaction among students (O'Dowd, 2018). The Stevens Initiative's typology (2021) describes VE as "a method that uses technology to connect people for education and exchange" (p. 5) and the EVOLVE project states VE is "a form of computer-mediated learning whereby students from geographically remote classes work together online (in pairs or small groups) on learning tasks developed by teachers or educational facilitators" (n.d.-a, para. 2). While the EVOLVE project emphasizes the importance of "regular synchronous or near-synchronous meetings using high social presence media" (EVOLVE, n.d.-b,

* **Corresponding Author:** Eric Hagley, eric@iveproject.org

para. 3), the present study aligns more closely with Akayoglu et al. (2022), where students primarily interacted through asynchronous communication yet successfully met their pedagogical goals, thus illustrating that regular synchronous interaction is not a prerequisite for meaningful engagement in VE. In our context, VE refers to the structured facilitation of both asynchronous and synchronous online interactions between students from different countries, guided by trained educators, with the aim of fostering FL development and Intercultural Competence (IC).

The two main VE models are the lingua franca VE model, in which all participants communicate in a shared language, and the dual-language VE model, in which students use two languages and collaborate with partners who are typically target-language speakers of the languages being studied (Hagley, 2020). The VE used in this study had English as the lingua franca among students whose first language was not English.

A large body of research on VE attests to its growing acceptance, as seen in the *Journal of Virtual Exchange*, bibliographies such as the Asia Pacific Virtual Exchange Association's (APVEA, n.d.), and the efforts of global VE organizations such as UNICollaboration and the Stevens Initiative. Benefits of VE for FL development include oral communication improvements (Canals, 2020; Roarty et al., 2023), enhanced critical literacy and writing (Hauck, 2019; Hilliker & Yol, 2022), and increased intercultural understanding and competencies (Machwate et al., 2021; Zilberberg Oviedo & Krimphove, 2022). Nevertheless, VE is still a relatively new practice (Barbosa & Ferreira-Lopes, 2023) and assessment has been a challenge (Wicking, 2022). Ideally, the assessment methods and criteria that teachers use in VE are shaped by their instructional goals and intended learning outcomes (e.g. IC and FL development), but many teachers must also consider practical restraints such as institutional requirements, workload and available resources. Understanding why and how assessment is carried out, and the outcomes it produces in terms of washback (Hughes, 2002), are essential but still not well understood.

While Czura and Dooly (2022) explored aspects of FL assessment in VE, assessment still remains an understudied area within the VE field. Moreover, no research to date has examined the VE assessment practices employed by teachers from diverse cultural backgrounds across all phases (from initial planning to post-reflection). This study therefore aims to explore factors that inform teachers' assessment design and implementation, how they assess their students according to their own intended learning outcomes, as well as the challenges they encounter and the ways in which these challenges are addressed. For this reason, the Research Questions (RQs) of this qualitative descriptive study are broad. They are:

1. Why do teachers use the assessment methods they choose?
2. How do teachers assess their students' participation in a VE?
3. What challenges do VE teachers encounter on their assessment journey and how are they overcome?

Literature Review

Assessment is defined as "appraising or estimating the level or magnitude of some attribute of a person" (Mousavi, 2009, p. 35). The two commonly-used methods are formative and summative assessment (Dixson & Worrell, 2016). While formative assessment focuses on collecting information across the period of instruction with the aim of enhancing the learning experience, summative assessment measures learning outcomes after instruction, typically in the form of a grade (Anderson & Krathwohl, 2001). Despite the differences, they should not be treated as a dichotomy but rather a continuum (Czura, 2022). There has also been growing interest in dynamic assessment which involves the interaction between a learner and an examiner aiming to assess the learner's modifiability and explore ways to enhance and sustain cognitive functioning and improvement (Lidz, 1987). The teacher in dynamic assessment takes on the mediator role and "engages cooperatively with learners and intervenes when difficulties arise" (Poehner et al., 2017, p. 244).

IC and FL development are integral components in assessment of students' participation in VE. Fantini (2012a, p. 269) notes that “[g]iven their common and overlapping areas, language education and intercultural communication ... are inextricably linked” which can influence the objectives of teachers who use VE (Belz, 2002; Machwate et al., 2021). This is evidenced through them being the most widely taught and assessed, as noted by Lewis and O’Dowd (2016) who found, of 54 VE studies, 23 focused on FL skills development, 23 on Intercultural Communicative Competence (ICC), and one on both. Another bibliometric study of 254 articles (Barbosa & Ferreira-Lopes, 2023) noted “research streams that comprise a larger number of studies are: language development, intercultural communication and teacher education” (p. 570). FL learning and intercultural communication are, therefore, the two most widely studied and assessed areas of VE.

ICC has become a common component of FL assessment but the “lack of clarity about the nature of ICC makes its assessment especially challenging” (Fantini, 2012b, p. 391). VE practitioners have drawn on a range of assessment approaches to overcome these difficulties, such as reflective portfolios (Izmaylova, 2022), Zoom-based discussion analyses (Machwate et al., 2021), and learner-created texts, videos, or audio recordings (Cavalari & Aranha, 2022). Regarding FL assessment, Lee and Sauro (2021) categorize VE-related language assessment practices into three approaches: evaluating changes in learners' language use over time, using pre- and post-tests to measure learning outcomes, and relying on students' self-reporting of learning. Czura (2022) outlines the myriad complicated factors involved in assessing learners engaged in VE within FL contexts, while building a succinct case for its importance. She notes that teachers' assessment should be based on carefully-designed, appropriate content aligned with course objectives. These objectives may include fostering IC, enhancing content-related knowledge, improving digital literacies, and developing “21st century skills such as collaboration, tolerance, critical thinking, problem-solving, leadership, and flexibility” (Czura, 2022, p. 32). Furthermore, assessment must have clearly defined constructs, demonstrate validity and reliability, and provide evidence to support ongoing course adjustments and future instructional planning.

It is evident from the literature that teachers' assessment of students' participation in VE is still nascent. Furthermore, as VEs bring together teachers from various cultures and countries, there is a need to explore the reality of assessment within those different contexts. By documenting eight teachers' assessment design and implementation journeys, the present study seeks a deeper understanding of how assessment of students' participation in a large-scale VE is carried out in different educational contexts.

Methods

This study adopted a qualitative descriptive design which is defined as “a comprehensive summarization, in everyday terms, of specific events experienced by individuals or groups of individuals” (Lambert & Lambert, 2012, p. 255). Thus, the assessment journeys of eight teachers were comprehensively documented through pre- and post-surveys and interviews noting key characteristics shaping their assessment designs and implementations, choice of assessment methods and criteria, tasks assigned to their students, and challenges they faced throughout their VE participation.

Study Context: The International Virtual Exchange Project

The International Virtual Exchange Project (IVEProject, n.d.) is a large-scale VE that aims to develop students' IC, FL and communication skills, as well as digital literacy, through authentic intercultural communication (Table 1). Two eight-week exchanges are held each year. During these, students are given autonomy to interact using English as a lingua franca, primarily through online forums centered on topics of personal interest. The eight teachers, who were the research subjects of this study, had their students participate in the October-November 2022 iteration of the IVEProject.

Table 1*Overview of the IVEProject*

Initial launch (two-country exchange)	2004
Since participation expanded with Japanese Kaken grant assistance	2015~2025 (greater than 50,000 students, 700 teachers, 33 countries)
Number of annual exchanges	Two
Exchange duration	Eight weeks
Participants' first language	Not English
Language of communication	English
Primary form of interaction	Asynchronous forums
Students' fields of study	Diverse (mostly non-English majors)
Additional activities	Side quests, the student-generated survey, etc.

As shown in [Table 2](#), the October–November 2022 IVEProject iteration on which this study is based, included 3,415 students from 16 countries, guided by 128 teachers. Students in the IVEProject were from diverse academic backgrounds: English as a compulsory part of their degree programs (e.g., engineering, medicine, tourism); English or intercultural studies; whilst some were pre-service teachers. A common feature among IVEProject participants is that all are engaged in learning English as a FL. Of interest to this study were the teachers guiding these students.

Table 2*Participant Data for the October–November 2022 IVEProject*

Number of active student participants	3415
Number of participating countries	16
Number of participating teachers	128

Researchers' Status

The authors, who all have experience as teachers in the IVEProject, were not the research subjects. We aimed to contribute to a broader understanding of assessment practices in VE contexts, particularly within the IVEProject. Viewing this as beneficial to both our own teaching and the wider educational community, we conducted a non-interventionist study documenting and analyzing the assessment methods used by the research subjects, eight participating teachers.

Research Subjects

To represent the different countries and cultures in the IVEProject and to provide a rich diversity of data regarding participating teachers, convenience and purposive sampling was used in selecting the research subjects. As Andrade (2021) acknowledges “research based on convenience and purposive samples can be important and necessary, such as when sociocultural and other factors are expected to influence outcomes. Through convenience and especially purposive sampling, the findings relevant for subpopulations can be identified” (p. 88). Thus, among the 128 teachers in the IVEProject (convenience), the researchers contacted a subset who represented different countries/cultures and had a varying degree of experience teaching in the IVEProject (purposive). Eight teachers based in six different countries agreed to participate. [Table 3](#) outlines detailed profiles.

Table 3*Subject Profiles*

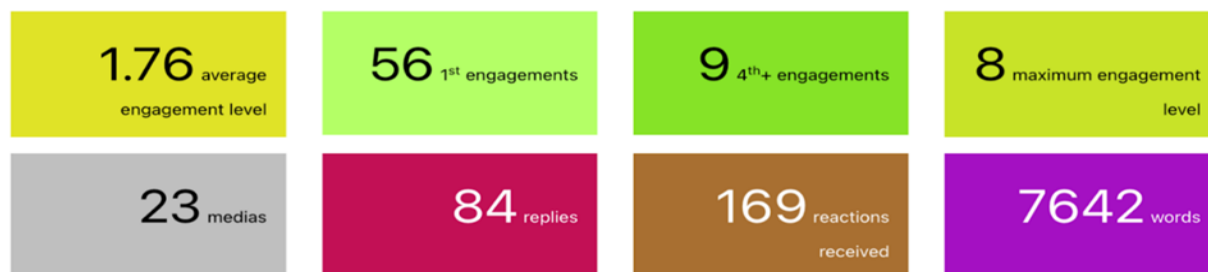
Teacher	Based in/ from	Course Name Average Class Size Weighting ¹	Self-assessed technology- expertise level	Times in the IVEProject	Participation of their students	Teaching experience (yrs)
T1	Japan/ US	Comparative Culture 40 30%	Advanced	5	Obligatory	30
T2	Japan/ US	Oral Communication 30 15%	Advanced	2	Obligatory	25
T3	Japan/ Canada	English Communication 26 10%	Intermediate	4	Obligatory	9
T4	Brazil	English Language 22 10%	Intermediate	1	Obligatory	17
T5	Chile	Cultural Knowledge through English in a Globalized World 35 15%	Intermediate	3	Obligatory	34
T6	Colombia	English for ESP 20 “depends”	Advanced	20	Voluntary	26
T7	Indonesia	World Englishes in TESOL 30 20%	Intermediate	1	Voluntary	5
T8	Türkiye	Teaching Language and Literature 50 15%	Intermediate	1	Voluntary	30

Note. ¹ Percentage of the overall course grade allocated to the IVEProject.

Participants' informed consent was obtained, and ethical approval obtained from the Ethics Committee of Çukurova University (E-74009925-604.01-1214569). Furthermore, as the IVEProject is GDPR compliant, previous years' exchange data is not kept unless it has consent (GDPR Art. 6.1(a)) and/or legitimate interests (GDPR Art. 6.1(f)).

Assessment Tools

Within the IVEProject, teachers retain autonomy over how they assess their own students. To support this process, the platform offers three key tools: the Forum Dashboard, the Complete Report, and the Forum Report. The Forum Dashboard (Figure 1) is designed to provide individual students access to their progress data. Upon each login, it displays quantitative participation metrics, such as engagement level, number of replies, and total word count, offering students a real-time overview of their participation.

Figure 1*Student Dashboard View*

The term “engagement level” in [Figure 1](#) needs further description as it is often a crucial data point used by teachers in their assessment. [Figures 2](#) and [3](#) (based on 2024 IVEProject data) illustrate how engagement level metrics in [Figure 1](#) are determined, with anonymized data consent obtained. [Figure 2](#) shows a Japanese and a Colombian student engaging in frequent exchanges over three weeks (shown), which was followed by continued discussion over six more weeks. We refer to this sustained, reciprocal interaction, characterized by threaded, indented replies, as “deep engagement”. Such interactions improve students’ engagement levels. In contrast, [Figure 3](#) depicts “shallow engagement”: five students from four countries replied to a Japanese student’s initial post, but the original poster did not respond, ending the interaction. These examples come from the 2024 October-November IVEProject exchange, as 2022 data was deleted to ensure GDPR compliance. Similar patterns were observed in 2022.

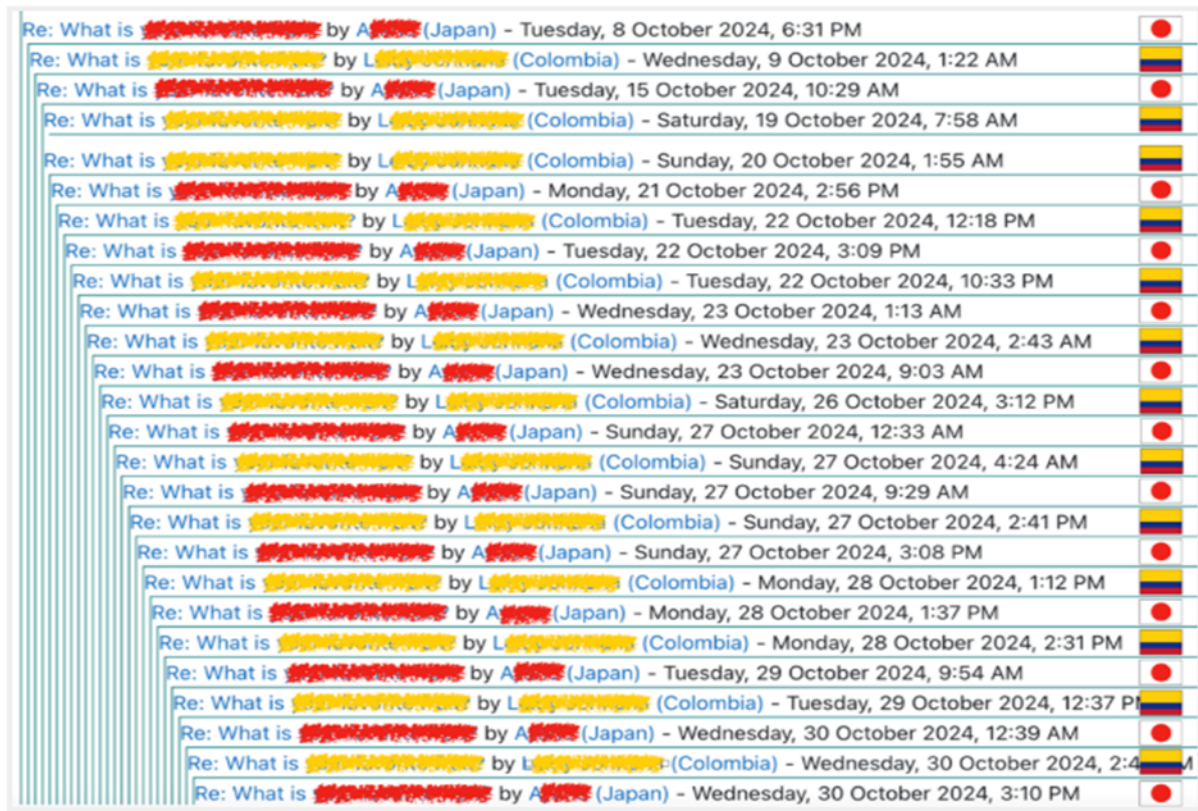
The second and third tools, the Complete Report and the Forum Report, respectively, are intended for teachers, and provide complete records and detailed analytics of students’ participation. The Complete Report enables teachers to capture an individual student’s total contribution while the Forum Report ([Figure 4](#)) presents quantitative data for all students in a class, including metrics such as number of posts and replies, total word count, days active, multimedia usage, and engagement level. It is important to note that while higher average engagement values may suggest deeper interaction, the true quality of engagement can only be evaluated by examining the content of students’ contributions. However, low engagement levels serve as an early indicator, helping teachers identify students who may benefit from additional guidance or support. Teachers can filter and export this data by specific time frames, allowing for targeted analysis. [Figure 4](#) shows a partial example of how quantitative data from the Forum Report appears when a class’s data is exported. Each participating IVEProject teacher has access only to their own students’ data, allowing them to efficiently assess the engagement of their class.

Instruments and Data Collection

Data from the eight participating teachers was obtained through pre/post-surveys and semi-structured pre/post interviews developed by the authors ([Figure 5](#)). Both surveys and interviews included open-ended and closed-ended questions ([Appendix A](#)) designed to explore all phases of assessment. Participants were explicitly informed that participation in both the surveys and the interviews was voluntary and that all responses would be anonymized. Each data collection step informed the preparation of the next phase. Surveys were submitted online via Google Forms and the IVEProject platform’s questionnaire module. Interviews were conducted via Zoom. All pre- and post-interviews (16 in total with a duration of 7 hours 16 minutes) were recorded. The interviews were conducted for two main purposes: to expand on the survey findings so that participants could clarify and elaborate on their written responses, and to pursue a deeper understanding of changes between the pre- and post-project stages by eliciting whether teachers’ individual expectations regarding assessment design and implementation were realized, evolved or unmet and why. Thus, the post-interviews incorporated questions concerning the pre-survey, pre-interview, and post-survey feedback, and utilized tailored interview protocols based on earlier data collection phases.

Figure 2

Example of Deep Engagement in a Forum Discussion



Continued to 88 iterations

Figure 3

Example of Shallow Engagement in a Forum Discussion

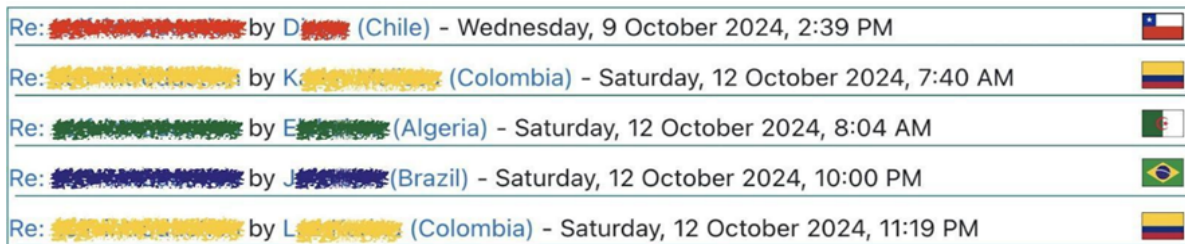


Figure 4

Example of Data the Forum Report Provides

First na	Last na	Posts	Replies	days active	Views	Word count	Multimedia	AvEnLevel	MaxEngLevel	React	Rec
Stu1	Dent1	8	106	19	34	5832	19	3.36	14	27	
Stu2	Dent2	1	5	2	3	606	0	1	1	1	
Stu3	Dent3	1	47	15	197	7874	33	2.38	7	51	
Stu4	Dent4	1	3	2	5	414	5	1	1	0	
Stu5	Dent5	2	43	14	63	4337	21	1.54	5	38	
Stu6	Dent6	11	99	30	327	10012	54	1.45	6	44	
Stu7	Dent7	2	3	5	1	313	1	1	1	1	

Figure 5

Instruments and Data Collection Phases



Data Analysis

Open-ended responses from surveys and interviews were analyzed thematically using an inductive approach and the free, open-source qualitative data analysis software Taguette (Rampin & Rampin, 2021). Interview recordings were initially transcribed using the Office365 transcription function. Transcripts were then checked against the recordings to confirm accuracy and corrected where necessary. Open-ended survey responses and finalized interview transcripts were subsequently imported into Taguette for analysis.

The analysis followed Braun and Clarke's (2006) six-step framework for thematic analysis to identify recurring codes and themes (see [Appendix B](#) for an example). Taguette supported collaborative familiarization with the dataset and enabled the researchers to highlight and code relevant words, phrases, and sentences. Codes were then organized into themes and sub-themes.

To promote consistency, all four researchers independently analyzed the same two participants' pre-interviews to generate preliminary codes and themes. The researchers then met to compare interpretations and collaboratively agree on theme and sub-theme labels. Themes were iteratively refined through ongoing comparison of the data until the team judged them to be well-supported. After consensus was reached, the remaining six pre-interviews were divided between two teams of two researchers for analysis. Several online meetings were held at the end of coding to finalize codes and themes. During these meetings, the researchers achieved over 90% agreement on themes and codes, with remaining discrepancies resolved through discussion. The same procedure was applied to the open-ended survey responses and post-interviews to support inter-rater reliability.

Results

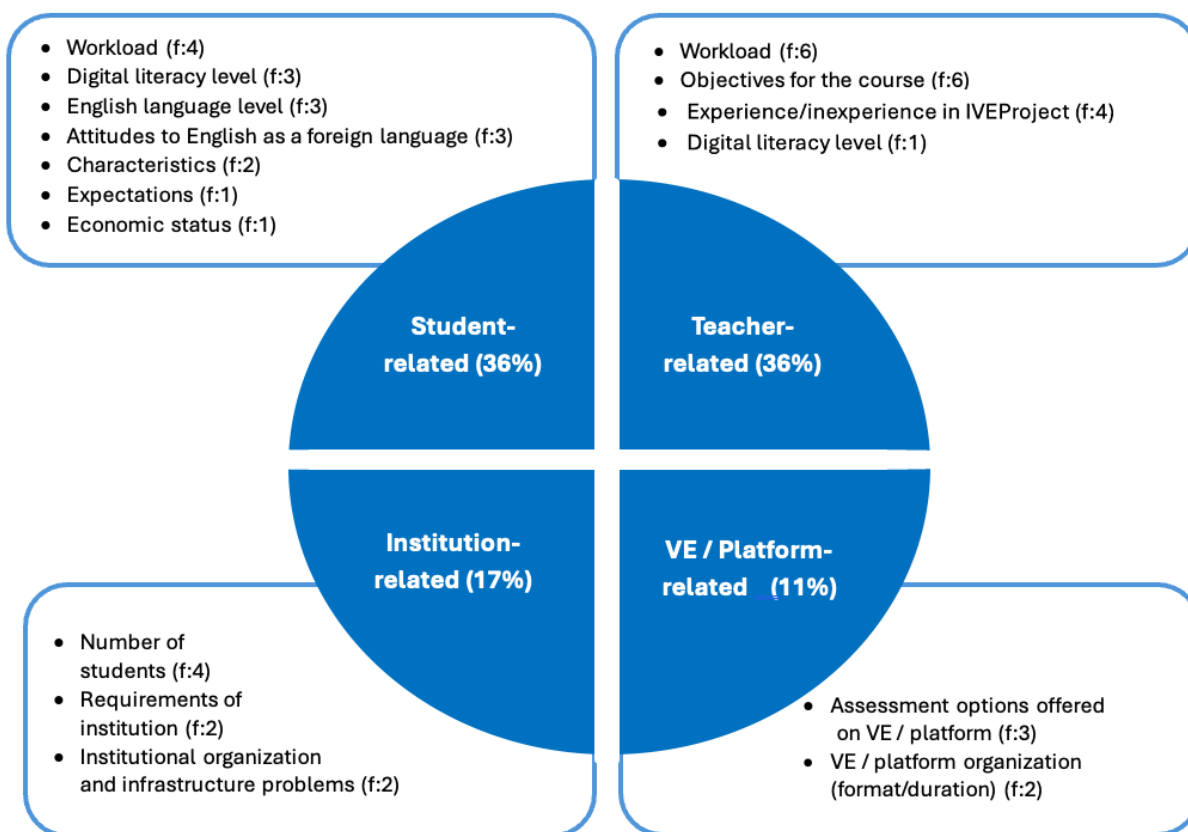
Reasoning Behind Teachers' Chosen Assessment Design and Implementation

To gain insights into teachers' decision-making processes, we asked them to describe the factors that affected their assessment design and implementation (RQ1). Thematic analysis revealed four overarching themes influencing assessment choices: student-related, teacher-related, institution-related, and VE/platform-related factors ([Figure 6](#)). These were further broken down into 16 sub-themes, with frequencies (f) indicating how many teachers mentioned each factor. Among the four themes, student-

related and teacher-related factors emerged as the most prominent, each accounting for 36% of all references. The student-related category included seven sub-themes, the highest among all categories, with “student workload” (f:4) being the most frequently mentioned. Other student-related considerations included “digital literacy level” (f:3), “English language level” (f:3) and “student characteristics” (f:2) (e.g. age and personality). In the teacher-related category, “teacher workload” and “objectives for the course” (e.g. language or culture) were most cited (f:6). Experience (or lack thereof) was also mentioned by half (f:4). These findings show that teachers’ assessment choices were primarily shaped by their own teaching conditions and their students’ needs. Thus, why they incorporated VE into their syllabi depended on each individual teacher’s context which, in turn, shaped the assessment tools they utilized.

Figure 6

Factors Influencing Teachers’ Assessment Design and Implementation



How Teachers Assessed

Building on the factors influencing assessment design and implementation identified in RQ1, the authors next examined what and how teachers assessed their students in the IVEProject (RQ2). Table 4 (A to F) details the assessment focus and methods used within the VE. Specifically, Table 4(A) reveals that teachers aimed to assess both IC (e.g. intercultural knowledge: T2, T4; intercultural communicative ability: T3, T7, T8) as well as FL development (e.g., T1, T5, T6). Given that it provides students with frequent opportunities for authentic intercultural communication in the target language, English, teachers perceived the IVEProject as an effective means to support course learning outcomes. In addition to institutional goals, some teachers were guided by personal motivations. For example, T8 viewed their participation as a form of professional development, while T1 sought to embed the IVEProject into a broader university-wide curriculum. T4 emphasized fostering student autonomy in language learning,

alongside a focus on developing linguistic-discursive skills through a genre-based pedagogical approach.

Table 4

Assessment in Virtual Exchange: Design and Implementation Details

Teacher		1	2	3	4	5	6	7	8	Total
(A) Objectives	Intercultural competence		+	+	+			+	+	5
	Foreign language development	+				+	+			3
(B) Assessment method	Formative				+	+			+	3
	Formative & summative	+	+				+			3
	Formative & dynamic							+		1
	Summative			+						1
(C) Assessment criteria	Quantitative & qualitative	+	+		+		+	+	+	6
	Quantitative			+						1
	Qualitative					+				1
(D) Assessment technique	Teacher assessment	+		+					+	3
	Teacher & self-assessment		+		+	+				3
	Teacher, self & peer assessment						+	+		2
(E) Presence of a rubric	Yes	+	+	+		+		+		5
	No				+		+		+	3
(F) Teacher-perceived difficulty level of their assessment of students	Easy	+		+		+		+		4
	Neither easy nor difficult		+		+				+	3
	Difficult						+			1

Building on the rationale behind teachers' assessment choices, this section addresses RQ2 by investigating the methods used to evaluate student participation, which typically involved assessing forum

interactions alongside reflective artifacts. As shown in [Table 4\(B\)](#), teachers reported employing various combinations of formative, summative, and dynamic assessment strategies for assessing IC and/or FL development. Formative assessment was the most commonly used (f:7) appearing in three of the four assessment methods used. Four teachers used summative assessment, with three combining it with formative assessment and one using it exclusively. One teacher stated they used dynamic assessment, alongside formative assessment.

[Table 4\(C\)](#) indicates that most teachers (f:6) used a combination of quantitative and qualitative assessment criteria, while only one relied exclusively on quantitative assessment and another solely on qualitative. It should be noted that three teachers altered their plan regarding the use of quantitative and qualitative criteria. Two went from solely quantitative assessment to a combination of quantitative and qualitative, while one went from “both” to “only qualitative”. On the relationship between quantitative and qualitative assessment criteria, T8 stated:

Well, that quantitative part always facilitates the qualitative part and opens up room for creating some questions to ask in the qualitative data collection part.

[Table 4\(D\)](#) reveals variation in who conducted the assessment. While three teachers relied solely on teacher-led assessment, others combined teacher assessment with self-assessment, or used a mix of teacher, self-, and peer-assessment (f:2). [Table 4\(E\)](#) shows that rubrics emerged as the most commonly used (f:5) assessment tool (see examples in [Appendix C](#)). Of the three who did not use a rubric, two were new IVEProject participants and indicated plans to incorporate rubrics in future iterations. The remaining teacher cited institutional constraints for not employing one.

[Table 4\(F\)](#) provides information about the perceived assessment difficulty level. Only one teacher perceived the assessment criteria they used to be “difficult” while half (f:4) deemed them “easy to achieve”. Furthermore, teachers reported intentionally designing assessment criteria that were achievable in order to promote increased intercultural interaction. Only T6 believed that achieving high grades in his assessment was “difficult,” although he attributed this to factors beyond his control, such as differences in students’ English proficiency and technological illiteracy. As a result, he added that even interacting with other students asynchronously via forums can be difficult for his students.

[Table 5](#) presents the assessment criteria teachers used, as identified in the pre- and post-surveys and post-interviews. During the pre-survey, the eight teachers chose from a closed-ended selection of quantitative and qualitative assessment criteria, with T7 adding three additional qualitative criteria. Among the quantitative assessment criteria, “engagement level” (pre-f:7/post-f:6) and “number of posts” (f:5/f:4) were the most frequently selected. For qualitative assessment, the most frequently cited criterion was “inquiry into other cultures” (f:5/f:6). This was followed by “inclusion of information about students’ own culture” (f:4/f:5) although how these were evaluated was not explicitly stated.

Student Tasks and Artifacts

In addition to analyzing students’ IVEProject forum contributions, regardless of whether formative, summative, or dynamic assessment was used, most teachers (f:5) assessed supplementary student work such as reflective journals, video presentations, and in-class presentations. These artifacts also afforded students opportunities to deepen and reflect on their learning. For example, T7, who stated she employed formative and dynamic assessment, required students to submit weekly reflective journal entries based on their IVEProject experience. In terms of formative assessment, she monitored her students’ learning through these journals. For dynamic assessment, she focused on what her students could achieve with her assistance and mediation through an interactionist dynamic assessment approach, meeting weekly with students, using their reflective journals to identify teaching moments and offer individualized support.

She prioritized this mediation process through one-on-one dialogic interaction to better observe students’ learning potential, stating that this allowed her to diagnose and address students’ issues (see [Appendix D](#) for example quotations). T4 used formative assessment, incorporating both teacher- and self-assessment.

She stated key words from the IVEProject, such as “engagement” were used in open-ended questions such as “What do you think about your level of engagement? What was your contribution to the forums?”

Table 5

Quantitative and Qualitative Criteria Used

Criterion		f (n:8)	
		Pre-survey	Post-survey/ post-interview
Quantitative criteria	Engagement level	7	6
	Number of posts	5	4
	Number of words	4	3
	Number of audio-visuals	4	3
	Number of forum views	3	3
	Interaction with participants from other countries	2	2
	Number of different discussions participated in	1	1
Total		26 (55%)	22 (46%)
Qualitative criteria	Inclusion of inquiry about other cultures	5	6
	Inclusion of information about own culture	4	5
	Respectfulness towards others	4	4
	Sustainment of discussions by using questions	3	5
	Quality of grammar, vocabulary, and other language used	3	3
	Care of opening and closing of a post	1	1
	Analyzing others' discussions and reflecting on issues discussed (T7)	1	-
	Consideration of others' ideas and opinions (T7)	-	1
	Offering of solutions and suggestions to issues raised (T7)	-	1
Total		21 (45%)	26 (54%)

After students answered, she also answered with her evaluation of each student, and assigned a score. Then, they held reflective sessions comparing evaluations with focused questions: “Do you agree or disagree with this evaluation? Do you think that my perception of your participation is what actually happened?”

In addition to video-recorded reflections, T5 used students' posts in the IVEProject as a means of formative assessment. Since she focused on developing the writing skills of her students via error analysis, every one or two weeks she chose one post from each student with the highest number of words and added that to a folder. She later used it in one-on-one meetings, giving feedback to her students. The only teacher who solely used summative assessment was T3, whose students needed to attain the total "number of words, number of posts, and number of audiovisuals" indicated in his "rubric".

Challenges to Overcome

Table 6, based on pre-interview responses, lists challenges either teachers or their students might encounter that could influence assessment practices (RQ3). While some of these challenges relate to factors identified in RQ1, the focus here is on how these challenges shaped their assessment design and implementation, including any modifications made to address them. Notably, the majority of challenges mentioned were those that teachers had anticipated (11, 69%).

Table 6

Challenges Foreseen Before Participating

	Challenges foreseen	Teachers
Perceived teacher challenges	Time constraints	T1, T5
	Integrating IVEProject into course objectives	T1, T4
	Limited technological experience	T4, T7
	Assessment of qualitative criteria	T4, T5
	Overextended criteria	T7
	Too many students	T1
	Possibility of students gaming the system	T3
	Total	11 (69%)
Perceived student challenges	Inclusion of inquiry about other cultures	T8
	Maintenance of interactions via appropriate questioning	T6
	Active participation	T6
	Special needs students	T4
	Total	4 (25%)
No challenges	Ability to foresee problems and plan accordingly	T2
	Total	1 (6%)

While all teachers indicated that their teaching schedules were demanding, T1 and T5 explicitly identified "time required for assessment" as a particular challenge. Integration of the IVEProject into course syllabi was a possible challenge for T1 and T4. T4 and T7 expressed concerns about their ability to navigate the system from a technological viewpoint whilst T4 and T5 worried about their capacity to objectively assess qualitative criteria.

The challenges of assessing qualitative criteria are demonstrated clearly by the following quotations:

T1: I don't have time to look at them [qualitative criteria] that carefully. But for example, like

“how respectful” but some of them are tough ... how respectful during the interactions. I mean, that’s a very hard thing to put a grade on. Inclusion about the other culture and these are very kind of subtle.

T2: Are they more connected to culture? Are they asking deeper questions you could do that but that’s just a lot of data to get into and try to wrangle and deal with in order to and it’s such a difficult thing to really assess. So yeah, going beyond the quantitative part is ... difficult.

An example of how to overcome these was offered by T2, who asked his students to record two-minute spoken reflections as a mid-term project, addressing the quality of their participation, what they learned about other cultures, what they shared about Japanese culture, and whether they enjoyed participating. He explained how he assessed their reflections:

T2: And even in the qualitative part in the speaking part ... I was assessing their English ability - their speaking fluency, their grammar, their vocabulary and the content was basically did they cover the topics that were listed in the assignment? ... So, it was pretty easy to assess and I think easy for them too.

Regarding possible student challenges, T8 was worried as to whether her students would be able to inquire deeply enough about peers’ cultures whilst T4 had a special needs student that was of concern. T2 stated that he was not worried as he had planned the assessment well in advance.

Post-IVEProject, teachers were asked whether they had altered their assessment practices during the VE. Three teachers reported making changes. T3 discovered his students had actively participated more than he had envisaged, achieving their goals quickly. As a result, he increased the required word- and post-count required to attain a full grade. T5, as she had anticipated, decided not to use quantitative criteria as the qualitative aspect was more relevant to her course. During the IVEProject, T7 had added 14 qualitative criteria in addition to the 13 provided in the pre-survey (27 criteria), but, as she had foreseen, she could not assess that many so decreased the number to 15.

The remaining teachers mentioned other challenges. For example, T1 noted that while his students posted multimedia, time constraints precluded integrating it into assessment. T2 did not encounter any major problems but observed that some of his students struggled to develop interactions, occasionally leaving their interlocutors “hanging”. He therefore encouraged his students to engage more deeply through more effective questioning techniques. T4 did not mention the problems she had envisaged but felt some students had not participated as actively due to perceived shame regarding their English level.

Teachers’ Reflections on Their Assessment Practices

To understand their challenges and how they were overcome (RQ3), teachers were asked in the post-survey and post-interview whether they were satisfied with their assessment methods. Four teachers reported complete satisfaction and would continue using their chosen methods. Two (T3, T7) stated they believed their methods aptly appraised their students’ level of participation and achievement of the intended learning goals. T4 did not specifically state a reason whilst T2 indicated “ease of assessment” as the reason. Of the teachers who expressed dissatisfaction, T8, a first-time participant, noted she had aimed for a fully objective rubric but was unable to develop one. T5 found her qualitative assessment criteria overly time-consuming, particularly when assessing “inclusion of information about their own culture.” As a result, she expressed uncertainty about how she would manage future assessment. T1 was dissatisfied due to time constraints analyzing video presentations. Lastly, T6 expressed a desire to study assessment practices more thoroughly and felt unable to assess effectively due to a heavy workload and frequent schedule changes.

Irrespective of their satisfaction with their assessment methods, teachers highlighted the idea that students were learning more than what was formally assessed during the IVEProject. The following quotations illustrate this point:

T5: The autonomous learner, ... ones that are more critical about or conscious of or use a lot more metacognition than the others. They benefit from it [VE].

T3: One student wrote how she hated English throughout all of junior and senior high school, and now she's happy to come to class. Now she's happy to speak English in my class because of the IVEProject. And you see, I'm not asking so much of them. And yet they're having this great interaction, this great experience.... From their feedback, ... a lot of them self-report improving their writing and improving their vocabulary, which wasn't actually tested.

Another teacher noted the reason for his assessment style as:

T2: Basically, I want them to be active in the project and my understanding and my belief is that through that activity, ... they're still going to learn a lot. So, in order for me to get them to be active in the project, I based my assessment [on their activity].

Discussion

Reasoning Behind Teachers' Chosen Assessment Design and Implementation

Assessment plays a crucial role in teaching. The reasons *why* the chosen assessment designs are used often determine *how* they were implemented, and this in turn affects the way students interact with the assessment system and what they gain from it. The data gathered shows that this is true for the teachers using VE in this study. As Mahoney (2022, p. 117), a former IVEProject teacher, emphasized “teachers are obligated to offer students pedagogical guidance, yet ... they should also consider taking a step back to give students the freedom to allow their language skills to grow organically in the VE setting, without too much pressure of evaluation.” The authors believe this concept extends beyond language skills, with teachers in this study wanting to create assessment methods that would organically promote IC and other benefits commonly cited in VE literature, including digital literacies (EVOLVE Project, 2020), learner agency (Kohn & Hoffstaedter, 2017), and collaborative learning (The EVALUATE Group, 2019). This was done by applying assessment methods and criteria that promoted students' engagement, leveraging both formative and summative assessment to do so, lending further credence to Czura (2022), who questioned the dichotomy between the two assessment types.

Analysis found four main factors (student-related, teacher-related, institution-related, and VE/platform-related) and 16 sub-factors that shaped teachers' VE assessment design. This aligns with Kurek's (2015) description of VE as a dynamic interplay of various agents: teachers, students, collaborative formats, institutions and varying linguistic and cultural backgrounds. Similar to Lewis and O'Dowd (2016) and Barbosa and Ferreira-Lopes (2023), our findings show the eight research subjects prioritized IC and FL development in their use of VE and these guided their assessment approach. This follows, as previous studies have demonstrated the positive impact of VE on both IC (Roarty & Hagley, 2021; Zilberberg Oviedo & Krimphove, 2022) and FL development (Canals, 2020; Machwate et al., 2021). All teachers in this study aimed to incorporate assessment criteria that would foster both, reflecting an awareness of the close connection between IC and FL development in VE highlighted in prior research (Belz, 2002; Cavalari & Aranha, 2022; Fantini, 2012a).

How Teachers Assessed

Assessing students' learning outcomes in VE requires a multifaceted approach, as teachers often seek to evaluate both FL development and IC. Of the three VE language assessment approaches Lee and Sauro (2021) proposed, two were evident in this study. One approach, demonstrated by T5, involved evaluating learners' language development over time by analyzing errors made in the forums and compiling an inventory to support future FL development. T4 emphasized student self-reporting through a genre-based approach to asynchronous forum writing, while others employed more holistic strategies.

As Fantini (2012b) notes, assessing various dimensions of IC remains a complex challenge, and this difficulty arises “not only because educators hesitate to assess learner subjectivity and the realm of values and dispositions, but also because of the difficulty of bringing together, in some way, the diverse facets of intercultural practices and capabilities” (Scarino, 2017, p. 23). Given that students in VEs are at different stages in their lifelong IC development, many teachers in this study were reluctant to assess it based solely on forum participation. Instead, some incorporated reflective tasks designed to foster self-awareness. This strategy also avoided privileging students who enter with higher IC levels, enabling more individualized and equitable assessment. Additionally, reflective activities (Czura & Dooly, 2022; Izmaylova, 2022) were tailored to align with specific course objectives, providing practical and targeted alternatives to evaluating forum posts for evidence of learning outcomes.

To carry out assessment of language learning and IC, most teachers in this study applied a combination of qualitative and quantitative assessment criteria (f:6) rather than qualitative-only (f:1) or quantitative-only (f:1). Teachers used qualitative criteria (54%) slightly more than quantitative (46%) demonstrating the importance of both in VE assessment. While recognizing the value of qualitative assessment, teachers are busy and have limited time to assess student-created content (Cavalari & Aranha, 2022; Izmaylova, 2022). The use of the Forum Report in the IVEProject, particularly the engagement levels, by seven of the eight teachers, highlights that system-generated quantitative assessment reports are effective and valuable time-saving tools. Three teachers noted the numerical output visible on the Student Dashboard was specific, objective and fair which helped gain student “buy-in”. This provided timely feedback for students and helped them monitor their progress toward meeting their instructors’ assessment criteria.

Teachers reported using the quantitative data from the Forum Report in assessment rubrics to enhance consistency and objectivity. Brookhart (2013, p. 4) states “a rubric is a coherent set of criteria for students’ work that includes descriptions of levels of performance quality on the criteria.” As the use of quantitative data in the teacher-created rubrics did not directly measure “performance quality,” this did not align neatly with the traditional use of rubrics. However, teachers were able to use the Forum Report metrics to serve as meaningful proxies, such as using engagement level metrics to indicate student participation in an online forum within a rubric framework.

Sole reliance on word counts or engagement metrics risks encouraging students to “game” the system rather than engage in meaningful communication. To mitigate this, most teachers complemented quantitative criteria with qualitative criteria to assess their students’ interactions with international peers. Furthermore, non-standard language and multimedia use were assessed based on their ability to convey meaning effectively, rather than adherence to formal norms. This evaluative focus indicates that communicative effectiveness was prioritized over linguistic accuracy which aligns with Hymes’s (1972) perspective that communicative success lies in the ability to use language purposefully, as well as with Savignon’s (2018) support for elevating communicative competence as a goal of 21st-century foreign language evaluation.

Challenges

Teachers face multiple challenges when assessing their students’ participation in VE, particularly given their already demanding classroom responsibilities. Czura (2022) proposed a series of extensive and potentially effective assessment practices in settings where a single teacher is responsible for a small number of students. However, in many tertiary-level FL classrooms globally, student-to-teacher ratios often exceed 20:1, and instructors frequently manage eight or more classes per week alongside additional institutional responsibilities. While the authors believe Czura’s VE assessment practices are educationally desirable, the unfortunate reality is they are not achievable in most cases, particularly given that many FL classes today are taught by part-time instructors who have less time to devote to their classes. Class profiles also change constantly, and the teachers may not be fully aware of one class’s capabilities compared to another. This was the case with T3 in our study who required modifications to his assessment criteria, which was consistent with Izmaylova (2022).

This study found that teachers perceive VE assessment as an evolving process. Challenges, both anticipated and encountered, such as the difficulty of assessing qualitative criteria—often prompt the development of tailored assessment methods and implementation strategies to address these complexities. Another challenge was institutional requirements that “tend to be designed for accountability purposes more than for educational ones” (Scarino, 2017, p. 27). This situation was mentioned by two teachers who articulated their institutions’ expectation that quantitative assessment be used, highlighting the importance of VE platforms providing teachers with such data.

Recommendations for Assessment in VE

Teachers were asked about their expectations of the IVEProject as well as their recommendations for future participating teachers. Generally, teachers reported being appreciative and cognizant of the IVEProject support available to them via the pre-exchange workshops, the Teachers’ Forum, the Student Dashboard, the Complete Report and the Forum Report. Nevertheless, two teachers mentioned they would like more explanation of the tools available. Another two teachers hoped for an AI detection tool, whilst another suggested establishing a working group within the teachers’ forum to share assessment-related knowledge. Finally, participants were asked what they would recommend to other teachers regarding assessment of students’ participation in VE. The following were proposed:

- obtain as much information as possible about the VE in which you will participate;
- review VE studies, particularly those conducted by teachers participating in that specific VE;
- have a clear understanding of your objectives and align them to your VE assessment design and implementation;
- utilize the assessment tools provided by the VE (if any), and supplement if necessary;
- avoid using excessive criteria; instead, prioritize a select few;
- allocate class time to showcase exemplary interactions, especially at the beginning of the exchange; and
- exchange ideas with experienced teachers of the VE.

Conclusion

This research explored the assessment journeys of VE teachers, providing insights into VE assessment design and implementation. Although this study had a small sample size and examined only one VE, it yielded several valuable conclusions. First, it found that teachers’ assessment designs and implementations were shaped by factors within four main themes: student-related, teacher-related, institution-related, and VE/platform-related. While all four themes are regarded as important, student- and teacher-related factors were prioritized by teachers, with “teachers’ workload” and “objectives for the course” identified as the most salient sub-themes. Second, students’ IC and FL development emerged as the two main objectives for VE participation, with the majority of teachers using formative assessment, often in conjunction with summative assessment. Third, the majority of teachers used a mixture of both quantitative and qualitative assessment criteria. However, teachers tended to rely slightly more on qualitative criteria despite the difficulty of assessing them. The most frequently selected quantitative and qualitative criteria reflected an emphasis on sustained forum engagement, with teachers valuing measurements such as “engagement level of the student” and “number of posts,” alongside evidence of “inquiry about the other culture” and “information about their own culture” to show achievement. Fourth, all teachers emphasized the use of rubrics in their assessment, citing reasons such as ensuring objectivity and meeting accountability requirements. Many also favored designing assessments that they considered easily achievable for their students that would promote increased interaction. Fifth, teachers faced challenges during assessment, and accordingly half of the teachers were not satisfied with their assessment design and implementation. However, only a small number of teachers needed to modify their assessment design, and it was observed that teachers felt that these challenges provided points for reflection for future exchanges. To enhance assessment practices, teachers emphasized utilizing the knowledge of fellow practitioners through sharing knowledge on the VE, exploring suitable assessment

tools, and learning from prior best practices. Finally, independent of the different assessment designs carried out, teachers concluded this assessment journey with the idea that whether assessed or not, the benefits in VE will be greater than what is assessed, so that participation in VE can be considered as a goal unto itself.

Acknowledgements

We would like to express sincere gratitude to the eight participating teachers for their valuable time and contributions. We also thank the reviewers and the editors for their comments and constructive feedback, which helped us to improve our manuscript. We are also grateful to Adam Jenkins, an organizing member of the IVEProject, for the technical support in the virtual exchange.

References

- Akayoglu, S., Üzüm, B., & Yazan, B. (2022). Supporting teachers' engagement in pedagogies of social justice (STEPS): Collaborative project between five universities in Turkey and the USA. *Focus on ELT Journal*, 4(1), 7–27. <https://doi.org/10.14744/felt.2022.4.1.2>
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Addison Wesley Longman.
- Andrade, C. (2021). The inconvenient truth about convenience and purposive samples. *Indian Journal of Psychological Medicine*, 43(1), 86–88. <https://doi.org/10.1177/0253717620977000>
- APVEA. (n.d.). *Asia Pacific Virtual Exchange Association*. Retrieved January 3, 2025, from <https://apvea.org/>
- Barbosa, M. W., & Ferreira-Lopes, L. (2023). Emerging trends in telecollaboration and virtual exchange: A bibliometric study. *Educational Review*, 75(3), 558–586. <https://doi.org/10.1080/00131911.2021.1907314>
- Belz, J. A. (2002). Social dimensions of telecollaborative foreign language study. *Language Learning & Technology*, 6(1), 60–81. <http://dx.doi.org/10125/25143>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Association for Supervision & Curriculum Development.
- Canals, L. (2020). The effects of virtual exchanges on oral skills and motivation. *Language Learning & Technology*, 24(3), 103–119. <https://doi.org/10.64152/10125/44742>
- Cavalari, S. M. S., & Aranha, S. (2022). Learners' diaries as a tool for teachers' assessment in teletandem. In A. Czura & M. Dooly (Eds.), *Assessing virtual exchange in foreign language courses at tertiary level* (pp. 65–78). <https://doi.org/10.14705/rpnet.2022.59.9782383720102>
- Czura, A. (2022). Virtual exchange: Issues in assessment design. In A. Czura & M. Dooly (Eds.), *Assessing virtual exchange in foreign language courses at tertiary level* (pp. 29–45). Research-publishing.net. <https://doi.org/10.14705/rpnet.2022.59.1408>
- Czura, A., & Dooly, M. (2022). Assessment in virtual exchange: A summary of the ASSESSnet project. In A. Czura & M. Dooly (Eds.), *Assessing virtual exchange in foreign language courses at tertiary level* (pp. 47–61). Research-publishing.net. <https://doi.org/10.14705/rpnet.2022.59.1409>
- Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory Into Practice*, 55(2), 153–159. <https://doi.org/10.1080/00405841.2016.1148989>
- EVOLVE Project. (2020). The impact of virtual exchange on student learning in higher education. <http://hdl.handle.net/11370/d69d9923-8a9c-4b37-91c6-326ebbd14f17>
- EVOLVE. (n.d.-a). *About EVOLVE*. Retrieved January 15, 2025, from <https://evolve-erasmus.eu/about-evolve/>

- EVOLVE. (n.d.-b). *What is virtual exchange?* Retrieved January 15, 2025, from <https://evolve-erasmus.eu/about-evolve/what-is-virtual-exchange/>
- Fantini, A. E. (2012a). Language: An essential component of intercultural communicative competence. In J. Jackson (Ed.), *The Routledge handbook of language and intercultural communication* (pp. 263–278). Routledge.
- Fantini, A. E. (2012b). Multiple strategies for assessing intercultural communicative competence. In J. Jackson (Ed.), *The Routledge handbook of language and intercultural communication* (pp. 390–405). Routledge.
- Fox, R., Corretjer, O., & Webb, K. (2019). Benefits of foreign language learning and bilingualism: An analysis of published empirical research 2012-2019. *Foreign Language Annals*, 52, 699–726. <https://doi.org/10.1111/flan.12424>
- Hagley, E. (2020). Musings on virtual exchange in the Asia-Pacific and beyond. In E. Hagley & Y. Wang (Eds.), *Virtual exchange in the Asia-Pacific: Research and practice* (pp. 231–240). Research-publishing.net. <https://doi.org/10.14705/rpnet.2020.47.1154>
- Hauck, M. (2019). Virtual exchange for (critical) digital literacy skills development. *European Journal of Language Policy*, 11(2), 187–210. <http://www.liverpooluniversitypress.co.uk/journals/article/20990>
- Hilliker, S., & Yol, Ö. (2022). Virtual exchange in teacher education: Focus on L2 writing. *ELT Journal*, 76(1), 1–19. <https://doi.org/10.1093/elt/ccab071>
- Hughes, A. (2002). *Testing for language teachers* (2nd ed.). Cambridge University Press.
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). Penguin Books.
- IVEProject. (n.d.). *International Virtual Exchange Project*. Retrieved January 3, 2025, from <https://iveproject.org/>
- Izmaylova, A. (2022). Assessing intercultural learning in virtual exchange. In A. Czura & M. Dooly (Eds.), *Assessing virtual exchange in foreign language courses at tertiary level* (pp. 135–146). Research-publishing.net. <https://doi.org/10.14705/rpnet.2022.59.1415>
- Kohn, K., & Hoffstaedter, P. (2017). Learner agency and non-native speaker identity in pedagogical lingua franca conversations: Insights from intercultural telecollaboration in foreign language education. *Computer Assisted Language Learning*, 30(5), 351–367. <https://doi.org/10.1080/09588221.2017.1304966>
- Kurek, M. (2015). Designing tasks for complex virtual learning environments. *Bellaterra Journal of Teaching & Learning Language & Literature*, 8(2), 13–32. <https://doi.org/10.5565/rev/jtl3.633>
- Lambert, V. A., & Lambert, C. E. (2012). Qualitative descriptive research: An acceptable design. *Pacific Rim International Journal of Nursing Research*, 16(4), 255–256. <https://he02.tci-thaijo.org/index.php/PRIJNR/article/view/5805>
- Lee, J., & Sauro, S. (2021). Assessing language learning in virtual exchange: Suggestions from the field of language assessment. *Journal of Virtual Exchange*, 4, 33–49. <https://doi.org/10.21827/jve.4.36087>
- Lewis, T., & O’Dowd, R. (2016). Online intercultural exchange and foreign language learning: A systematic review. In R. O’Dowd & T. Lewis (Eds.), *Online intercultural exchange: Policy, pedagogy, practice* (pp. 21–66). Routledge.
- Lidz, C. S. (1987). Historical perspectives. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 3–34). Guilford Press.
- Machwate, S., Bendaoud, R., Henze, J., Berrada, K., & Burgos, D. (2021). Virtual exchange to develop cultural, language, and digital competencies. *Sustainability*, 13(11), Article 5926. <https://doi.org/10.3390/su13115926>

- Mahoney, J. (2022). Opening the door: Using the International Virtual Exchange Project to enrich discussion classes. *Journal of Multilingual Pedagogy and Practice*, 2, 110–118. [https://fler.rikkyo.ac.jp/journal/jc0e3e0000000a9-att/JMPP\(2022_vol.2\)_12_Mahoney.pdf](https://fler.rikkyo.ac.jp/journal/jc0e3e0000000a9-att/JMPP(2022_vol.2)_12_Mahoney.pdf)
- Mousavi, S. A. (2009). *An encyclopedic dictionary of language testing* (4th ed.). Rahnama Publications.
- Nguyen, N. D-N., Leung, A. H-C., Hien, H-T., & Thiet, N-T. (2024). The impact of virtual exchanges using English as a lingua franca on students' linguistic and intercultural competence. *Teaching English as a Second Language Electronic Journal (TESL-EJ)*, 28(3), 1–29. <https://doi.org/10.55593/ej.28111a1>
- O'Dowd, R. (2018). From telecollaboration to virtual exchange: State-of-the-art and the role of UNICollaboration in moving forward. *Journal of Virtual Exchange*, 1, 1–23. Research-publishing.net. <https://doi.org/10.14705/rpnet.2018.jve.1>
- Poehner, M. H., Davin, K. J., & Lantolf, J. P. (2017). Dynamic assessment. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (3rd ed., pp. 243–256). Springer.
- Rampin, R. & Rampin, V., (2021). Taguette: Open-source qualitative data analysis. *Journal of Open Source Software*, 6(68), Article 3522. <https://doi.org/10.21105/joss.03522>
- Roarty, A., & Hagley, E. (2021). Analysis of using virtual exchange to develop intercultural understanding in EFL students. *Teaching English as a Second Language Electronic Journal (TESL-EJ)*, 25(3). <https://tesl-ej.org/pdf/ej99/a14.pdf>
- Roarty, A., Tuncer, H., & Tang, L. (2023). Adding synchronous sessions to asynchronous virtual exchange: Insights from the IVEProject. *Intercultural Communication Education*, 6(1), 19–32. <https://doi.org/10.29140/ice.v6n1.23>
- Savignon, S. J. (2018). Communicative competence. In J. I. Lontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1–7). Wiley. <https://doi.org/10.1002/9781118784235.eelt0047>
- Scarino, A. (2017). Culture and language assessment. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (3rd ed., pp. 15–31). Springer. https://doi.org/10.1007/978-3-319-02261-1_3
- Stevens Initiative. (2021). Virtual exchange typology. Retrieved from https://www.stevensinitiative.org/wp-content/uploads/2021/09/Stevens-Initiative-Virtual-Exchange-Typology_090121_singlepages.pdf
- The EVALUATE Group. (2019). *Evaluating the impact of virtual exchange on initial teacher education: A European policy experiment*. Research-publishing.net. <https://doi.org/10.14705/rpnet.2019.29.9782490057337>
- Wicking, P. (2022). Learning-oriented assessment in an international virtual exchange. In S. M. Hilliker (Ed.), *Second language teaching and learning through virtual exchange* (pp. 9–33). De Gruyter. <https://doi.org/10.1515/9783110727364-002>
- Zilberberg Oviedo, L. E., & Krimphove, J. (2022). Virtual exchange contributions to the development of intercultural competence: A Brazilian higher education institutions' perspective. *Actualidades Investigativas en Educación*, 22(1), 312–336. <https://doi.org/10.15517/aie.v22i1.47481>

Appendix A. Surveys and Interviews

Pre-survey

1. Which course(s) do you teach this term? How is the IVEProject related to objectives of that course(s)?
2. What is your objective in joining the IVEProject?
3. Participation of your students in the IVEProject is
 - a. Obligatory

b. Voluntary

4. How many times have you participated in the IVEProject?

5. Below are some criteria you can use for the assessment of your students' performance in the IVEProject. Please choose those included in your own assessment model. (First seven criteria are provided by the IVEProject Forum Report)

- a. Number of words
- b. Number of posts
- c. Number of forum views
- d. Number of different discussions participated in
- e. Interaction with participants from other countries
- f. Engagement level of the student
- g. Amount of multi-media
- h. Quality of grammar, vocabulary, and other language used
- i. Inclusion of inquiry about other cultures
- j. Inclusion of inquiry about the other culture
- k. Respectfulness towards others during interactions
- l. Sustainment of discussions by using questions
- m. Care of opening and closing of a post
- n. Other:

6. Please explain your assessment model in detail (What percentage is given for IVEProject participation? How do you calculate that percentage? Which criterion obtains the highest/lowest percentage and why? etc.)

7. In the Teachers' Forum of the IVEProject, assessment models were discussed. Do you think such discussion threads are helpful? Please explain.

8. The IVEProject provides calculations such as "engagement level, number of words, etc." to make assessing easier for teachers. Recently, two exportable models of assessment were prepared too and shared with teachers. What suggestions do you have for additional IVEProject tools/systems to assist with assessment? Please share with us.

9. Further comments/suggestions?

Pre-interview Example

(Individualized interview for T5/conducted by screensharing the pre-survey answers of the participant)

Questions based on T5's pre-survey answers:

1. For the 6th question, do you have a rubric? Is it calculated weekly or for the whole exchange? Can you give us the specifics?

2. Can you elaborate on what you wrote for "further comments"? What is your specific expectation from the IVEProject?

Pre-interview questions asked to all participants:

3. This is your (number) participation in the IVE, how did you decide on your assessment method? Can you please tell us about that process? Was it difficult or not?, etc.

4. From your students' perspective, which criterion might be the hardest /easiest for them? Why?
5. From a teacher's perspective, do you envisage any possible problems in assessing the success of your students?
6. Please look at the criteria (in the pre-survey) again, and choose one that you didn't include and explain the reason why you didn't include that.

Post-survey

1. What is your top priority in assessing your students in the IVEProject?
 - a. Development of intercultural knowledge
 - b. Development of intercultural communication
 - c. English language development
 - d. Other:
2. Which of the following options did your assessment method involve? (more than one can be chosen)
 - a. Teacher assessment
 - b. Peer-assessment
 - c. Students' self-assessment
 - d. Other:
3. My assessment type is primarily:
 - a. Formative
 - b. Summative
 - c. Both formative and summative
 - d. Other:
4. Why do you choose to use more qualitative or quantitative assessment methods? Or do you use an equal "measure" of both? Please give some reasons why you choose the assessment style you use.
5. Which of the following factors affect your assessment method? (more than one can be chosen)
 - a. Course objectives
 - b. The extent of IVEProject integration into my course
 - c. Number of students
 - d. My digital literacy level
 - e. My students' digital literacy level
 - f. My workload (for that term)
 - g. Students' workload
 - h. English language level of my students
 - i. Institutional requirements
 - j. Other:
6. Did you modify your assessment plan during the exchange?
 - a. Yes
 - b. No
7. If yes, why? and how did you modify it?
8. Do you think your students achieved what you hoped they would by participating in the IVEProject?

9. Do you have any recommendations for improving the assessment of your students in the IVEProject?

Post-interview Example

(Individualized interview for T3/conducted by screensharing the pre- or post-survey answers of the participant where necessary)

Based on T3's pre-survey answers

1. At the beginning, you chose three criteria “number of words, number of posts and number of audio visuals”, did you change those or add another criterion? If so, why?

Based on T3's pre-interview

2. During the pre-interview, you stated the “word count” criterion might be difficult. Was that so?

Based on T3's post-survey

3. You stated “Students were participating more than expected. I increased the quantity of words and posts needed to get 100%. Only the quantity was changed”. Can you please explain this? Did your students know about this change? How did they react? For future exchanges, will you keep this increased word and post quantity?

Post-interview questions asked to all participants:

4. Are you satisfied with your assessment method? Did you have difficulty in assessing your students? Why or why not?

5. What do you think about your assessment system in terms of difficulty? (easy - neither easy nor difficult - difficult)? Why is that?

6. Regarding assessment, what more could the IVEProject do to help you? Any suggestions?

7. Regarding assessment, what do you want to recommend to future teachers who will participate in the IVEProject?

8. What is it that you think you are assessing? i.e. Is it English Language proficiency / competency?

9. Which language skills are most important for you to assess and which one(s) do you think the IVEProject lends itself best to for doing that?

Appendix B. An Example of Theme Generation

Example of initial codes, sub-theme and theme generation from the pre-interview in response to “How did you decide on your assessment method? Can you please tell us about that process?”

	Quotation	Initial code	Sub-theme	Theme
T1	... it's a little bit higher percentage than I've usually done [in previous iterations of the IVEProject]. I've usually done 20% this semester. I'm doing 30%. And so that means both for me	Change in percentage and word	Teachers' experience/ inexperience	Assessment preparation factor/

	and for them we need to. We need to kind of make it a big part of every week ... Whereas 2-3 semesters ago when I was counting words. It wasn't like that. It was more of ah, oh geez, ... he wants us to. Post more words type of thing you know, but it's ... changed. And since so, since I think really a big change has been me understanding this, it kind of started last semester.	count	in the IVEProject	Teacher-related
T2	In the spring [previous iteration of the IVEProject]. Uh, it was basically looking at just getting them to participate. Uh, I hadn't really made it detailed as to what I was going to look at, and I realized when I was doing my grading that it really wasn't enough. For me as ah, to do the assessment without an actual rubric to look at, and it was all very subjective. So I mean basically all my students who did some participation got maximum points. And I knew that wasn't, I mean, that wasn't fair to everyone.	Improving on the previous IVE experience and assessment	Teachers' experience/inexperience in the IVEProject	Assessment preparation factor/Teacher-related

Note: Although quotations may relate to several sub-themes, this example highlights "Teachers' experience/inexperience in the IVEProject". Data from Pre-interview Question 3 (Appendix A) informed the design of Post-survey Question 5 (Appendix A). Figure 6 presents a combined summary of both sets of responses.

Appendix C. Rubric Examples for Quantitative and Qualitative Criteria

1- Rubric example for quantitative criterion (used by T2)

Criterion: *Number of words written (I encourage ~80 words / post or reply)*

1800+ words: 5%

1500-1799 words: 4%

1000-1499 words: 3%

500-999 words: 2%

1-499 words: 1%

2- Rubric example for qualitative criterion (used by T7)

Criterion: *The quality of grammar & vocabulary used in the interactions*

Vocabulary selection is very appropriate; vocabulary writing is very precise & grammar selection is very proper (score 5+): 5%

Vocabulary selection is appropriate; vocabulary writing is precise & grammar selection is proper (score 4): 4%

Vocabulary selection is quite appropriate; vocabulary writing is quite precise & grammar selection is quite proper (score 3): 3%

Vocabulary selection is less appropriate; vocabulary writing is less precise & grammar selection is less proper (score 2): 2%

Vocabulary selection is inappropriate; vocabulary writing is not precise & grammar selection is improper (score 1): 1%

Appendix D. An Example of How T7 Applied Formative and Dynamic Assessment

The example can be reached [here](#).

About the Authors

Hülya Tuncer (Ph.D.) is a teacher educator in the English Language Teaching Department at Çukurova University, Türkiye. She teaches undergraduate and postgraduate courses in teacher education, literature, and intercultural communication. Her research interests include English language teacher education, virtual exchange, reflective practice, speech and thought presentation, and interculturality.

E-mail: hulyatncr@gmail.com

ORCID: <https://orcid.org/0000-0001-8536-6471>

Eric Hagley taught for 24 years in Japan, as associate professor at Muroran Institute of Technology and research fellow at Hosei University. He is now researching doing his PhD at Griffith University, Australia. The data collected in this study is not associated with his doctoral dissertation. He is chair of APVEA, was vice-president of PacCALL and has written and presented on VE extensively.

E-mail: eric@iveproject.org

ORCID: <https://orcid.org/0000-0002-4795-8043>

Andrew Johnson is an Associate Professor at Future University Hakodate and has taught communication and ESL-related courses in Japan since 2002. His interests include intercultural communication, virtual exchange, and the application of generative AI in language learning contexts.

E-mail: andrewclarkjohnson@gmail.com

ORCID: <https://orcid.org/0009-0004-6965-5897>

Matthew Cotter is a New Zealand-trained educator who now teaches at Hokusei Gakuen University in Sapporo, Japan. He holds degrees in Māori Studies, Psychology, Teaching, and an M.Ed. specializing in TESOL. His research interests include computer-assisted language learning, indigenous education, and of course, virtual exchange.

E-mail: m-cotter@hokusei.ac.jp

ORCID: <https://orcid.org/0000-0003-0906-1400>