

Developing a second language listening and speaking assessment instrument for Use in student-led seminars in a chinese middle school english language acquisition class

International Journal of Chinese Education
January-April 2024, 1–16
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2212585X241234335
journals.sagepub.com/home/cne



John Stewart Clark¹  and Matthew Terrett²

Abstract

L2 speaking and listening practice typically involves inauthentic highly scripted learning experiences that are not reflected in real-world contexts. Seminars represent an unscripted authentic learning experience preparing students for future college courses and public discourse. Using an action research design, we developed a listening and speaking assessment instrument for L2 students to be implemented in student-led seminar discussions. The assessment instrument also served as a method of feedforward. Our results indicate an increase in student engagement across the seminar sessions. This was demonstrated by corresponding increases in the number of contributions, number of times students initiated dialogue, and positive listening indicators. Limitations in our assessment instrument were identified but the positive outcomes in increasing student participation and engagement justify further development of this instrument.

Keywords

L2 speaking and listening, speaking and listening assessment instrument, authentic speaking and listening assessment, feed-forward, student-led seminar

¹Curriculum and Instruction, American College of Education, USA

²Education and Assessment, The Open University, UK

Corresponding Author:

John Stewart Clark, Curriculum and Instruction, American College of Education, 101 W Ohio St Suite 1200, Indianapolis, IN 46204-1982, USA.

Email: stew.clark@hotmail.com



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without

further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Received 23 August 2023; accepted 4 February 2024

Introduction

Learning through peer discussion in tutorials or seminars is one of the common learning experiences at university level, but students from English as a Second Language (ESL) backgrounds are ill-prepared for such experiences (Alexander et al., 2018). Reasons for this include not knowing the conventions surrounding academic discussion, reliance on their lecturers in previous learning environments, not valuing the contributions of other students, inadequate preparation, and not understanding the benefits of class discussions. University instructors express concern about ESL students being unwilling or unable to interact with their peers in classroom discussion (Robinson et al., 2001) and speaking in academic contexts is cited by international students as challenging (Alexander et al., 2018). Learners need to possess sufficient confidence in listening and speaking skills in authentic discussion contexts.

English for academic purposes courses include peer discussion to better support students' preparation for university. Researchers recommend authentic preparatory listening input with an explicit focus on the discourse functions and strategies employed by expert speakers (Alexander et al., 2018). This means ensuring the students practice strategies necessary for target performance with appropriate scaffolding to avoid cognitive overload (Alexander et al., 2018). Within the Chinese context, Xu and Kou (2017) also stress the importance of teaching, modeling and practicing group interaction strategies to improve Chinese university students' oral English performance. Noman et al. (2023) documented Chinese university students reporting difficulties transitioning from their high school teacher-centric instruction to university student-centered seminar discussion. There seems to be no reason why high school or even middle school students could not also engage in such learning, so long as they have sufficient English language ability, appropriate scaffolding, and the self-management skills.

Listening is a passive skill that is the most frequently used form of communication (Feyten, 1991). The majority of second language listening instruction is practiced and assessed through inauthentic scripted texts (Wagner, 2014). Large-scale high-stakes testing negatively influences authentic assessment and listening materials available to teachers, which are often aligned with these tests due to washback (Wagner, 2013; Wagner & Wagner, 2016). English examinations for listening proficiency focus on accuracy with the goal being to limit the number of overall mistakes. Alexander et al. (2018) identify the English language classroom as a space where assessment can focus on more authentic performance and interaction rather than relying on materials targeting such high stakes exams, allowing classroom assessment to “focus on validity, authenticity and impact” (p. 321).

A review of 87 listening research studies demonstrated for the past 20 years researchers are still calling for an increase in assessment authenticity emphasizing real-world contexts (He & Jiang, 2020). In the Chinese context, university students have discussed how they were unprepared for their professors' different English-speaking accents and the need to adapt their listening skills (Noman et al., 2023). The overuse of inauthentic assessment and its use of standard English accents is arguably insufficient for second language learners. Real-world contexts do not necessarily require accuracy for minute details, but for general understanding and ability to respond appropriately (Ockey & Wagner, 2018). Speaking and listening are inseparable often with each influencing the other in a series of input and output. This dynamic relationship should be authentically modeled and practiced within the classroom for second language learners to be able to communicate effectively.

Researchers sought to increase student engagement through a student-centered inquiry-based seminar format in which students practiced listening, speaking, and writing skills using authentic formative assessments. This format and the instrument for assessment had been previously implemented by one researcher in an academic English program on an English medium of instruction undergraduate course in China and was adopted to this action research's context of three grade 9 International Baccalaureate (IB) Middle Years Programme (MYP) ESL classrooms.

Purpose

The purpose of this action research study was to develop an instrument to measure speaking and listening in classroom discussion settings while raising student engagement. Aligned with this purpose we hoped implementing this instrument would cause a shift towards a more student-centered classroom with students taking agency and ownership of their learning.

Literature Review

Listening

Calls for changes to listening assessment have occurred for the past two decades with little development in instructional and assessment practices (He & Jiang, 2020). There remains a gap between authentic language and textbook language (Gilmore, 2007). Researchers have hypothesized textbooks continue to publish highly scripted inauthentic practice dialogues for a variety reasons including lack of researcher understanding of classroom instruction, one-size-fits-all packaging models, fear of potential buyers' perceptions of "unprofessional" product due to purposeful mistakes added to make content authentic, level of difficulty to incorporate material into classroom instruction, and difficulty in creating meaningful tasks aligned to authentic unscripted texts post hoc (Wagner, 2014).

Listening as a skill includes not only the audio input for cognition, but the visual, contextual, and nonverbal cues (Wagner, 2014). Textbooks and assessments have improved upon listening content by adding videos as content for listening, but these videos are still highly scripted and do not represent authentic real-world interactions.

Scripted and unscripted texts are fundamentally different with researchers defining three areas of distinction between the two: organization/discourse, lexico-grammatical, and phonological characteristics (Wagner & Toth, 2014; Wagner & Wagner, 2016). Scripted texts are over pronounced with a reliance on British or American speakers. Texts do not vary in speech rate, do not include extended pauses or errors, slang, digressions, or listener response features (Wagner & Wagner, 2016). The use of scripted spoken texts for instruction and assessment for listening calls into question the validity of these texts when compared to real-world contexts. Studies comparing scores between scripted and unscripted listening assessments have documented higher scores for scripted text assessments (Wagner et al., 2021; Wagner & Toth, 2014). This could mean students are receiving inflated scores compared to actual listening ability. The continued use of scripted texts in instruction and assessment poses a threat to students' ability to communicate successfully in real world contexts.

Speaking

Speaking instructional content and assessments restrict students to isolated speech acts and situations which are often limited in pragmatic quality (Limberg, 2016; Nguyen, 2011; Ren & Han, 2016). Lacking pragmatic contexts threatens the real-world validity for these activities. Most

activities are highly structured pair work in the form of listen and repeat (Diepenbroek & Derwing, 2013). These instructional activities do not teach actual conversation strategies, but focus on bottom-up scripted speaking points which are then overly adapted to a broad range of situations without regard to socio-cultural implications (Han, 1992).

There are an unlimited number of situations with contextual variations in which a non-native speaker may find they need to orally communicate making the task of teaching all situations impossible with the classroom (Bardovi-Harlig et al., 1991). Teachers should use instruction time to support students in developing transferable skills providing competence when faced with unfamiliar non-explicitly taught conversation situations (Bardovi-Harlig & Griffin, 2005; Koester, 2002). In classroom discussion, these skills would include awareness of appropriate turn taking conventions as well as paralinguistic signals such as eye contact and gesture (Robinson et al., 2001). In the context of the IB these skills are defined as approaches to learning including communication skills, specifically exchanging thoughts, messages, and information effectively through interaction by negotiating ideas and knowledge with peers and teachers (IBO, 2022).

Student Participation and Engagement

Academic engagement is a predictor of academic achievement (Brophy & Good, 1986; Lee, 2013; Skinner et al., 1990). Active student participation leads to higher levels of learning when compared to students who do not participate (Weaver & Qi, 2005). Common et al. (2020) observed teachers who increased opportunities to respond (OTR), times in which students can vocally answer, led to an increase in student engagement. However, the OTR methods described were highly teacher centric. Teaching practices, such as OTR or increasing student verbal responses, have a direct effect on student achievement and an indirect effect on student engagement (Tomaszewski, et al., 2022). Chinese students have reported enjoying classroom group discussion because it is interactive and authentic (Zhang et al., 2022). The interactive and authentic nature of student-seminars increase effective student learning time which could foster student engagement and increase academic achievement.

Student-centered

Finkel (1999) appeals to teachers to decrease the amount of teacher talk and reconceptualize quality teaching as an environment fostering and facilitating student participation. Neumann (2013) suggested a framework for student-centered learning including student choice in content, authentic learning experiences with defined outcomes, and the reciprocal learning partnership between teacher and students. Open-ended seminars task students with bringing in prepared material outside of class to support classroom inquiry and discussion (Finkel, 1999). Students take ownership of learning with teacher authority incrementally decreased with the end goal of the teacher becoming a participant observer instead of director.

Researchers have highlighted issues with student-centered learning in the Chinese context stating the lack thereof even with the government's mandate for holistic *sushi* or holistic student development and assessment (Lou & Chan, 2023). Lou and Chan (2023) discuss the validity of these assessments given by teachers, questioning their authenticity and specifically pointing to issues regarding the absence of experiential learning in the classroom. They recommend integrating experiential learning within the curriculum and tying this authentic learning to assessment. This is directly aligned with the purpose of this study and the intention of developing the instrument.

Feedforward not Feedback

Recently, positive psychologists have researched the relationship between positive emotions and learning outcomes suggesting positive emotions can enhance student engagement in language acquisition (Dewaele & MacIntyre, 2014; Jiang & Dewaele, 2019; Shao et al., 2020; Wang et al., 2021; Weihua et al., 2022). Positive psychology research on ESL students linked foreign language enjoyment to proportion of time spent speaking in class, the degree to which the foreign language was spoken in class, and the teacher (Dewaele et al., 2018). Thus, strategies aiming to increase the amount of time students are speaking should also raise the positivity of the classroom experience and thereby further enhance engagement in a kind of virtuous cycle.

Our study looked to increase student speaking time in the target language through student-centered seminars. Regarding the teacher, we adopted a growth mindset approach to feedback. Xiaoyu et al. (2022) demonstrated an association between growth mindset and English language performance in Chinese ESL students with positive emotions being a mediating factor. Growth mindset is aligned with feedforward, a form of feedback which can be acted upon by the student in future work (Careless, 2007; Ormond et al., 2011). Students apply feedback in subsequent assessments demonstrating mastery (Koen et al., 2012; Rodriguez-Gomez & Ibarra-Saiz, 2015). Feedforward can not only be used to close the achievement gap, but also to engender dynamic skill development targeting problems which may spontaneously arise. Researchers exploring feedforward in the IB curriculum suggested teachers provide opportunities for students to reflect on feedback and improve self-regulation (Dulfer & Koopaei, 2021), which the researchers in this study undertook in the form of the speaking and listening assessment instrument. Monitoring students in class to produce a clear record of their progress is an effective use of short-term achievement assessment and, if successfully implemented, can help motivate students (Nation & Macalister, 2010).

Methodology

Classroom action research was chosen because the design allows for teachers to collaborate exploring possible solutions to classroom problems (Cohen et al., 2018). We sought to collaboratively address an immediate classroom problem, the lack of student participation in class discussion. Action research allowed us to foster our own professional development through reflection and evaluation of our teaching practices, scrutinizing listening and speaking language instruction for authentic real-world contexts. Simultaneously, we attempted to foster student engagement and student-agency in the belief these factors can support student learning creating a student-centered classroom. This aligns with the IB's programme development plan (PDP). The emphasis for both action research (Cohen et al., 2018) and the IB PDP is on the process with results being seen as ongoing, not static, but cyclical.

Reflection plays an important role within action research (Cohen et al., 2018). This study utilizes reflection in several formats, including teacher reflection on the research, student reflection on the skills of listening and speaking, and student reflection on engagement. Researchers not only attempted to improve student learning, but deliberately sought to improve their own practices not only from the results of the study but through their involvement in the process.

Instrument

The instrument was intended to measure student speaking and listening in an authentic classroom discussion setting while also increasing student engagement. This is the second cycle of this action research. Action research is a cyclical process and to increase the rigor of action research studies,

experts suggest running the research in more than one cycle (Melrose, 2001). The first cycle was in an undergraduate setting in which teachers first explored the development of the instrument due to classroom issues associated with Covid-19 and virtual instruction. Teachers discovered students were using translation software and reading directly from scripted speeches for online presentation assessments, producing inflated scores compared to actual unscripted speaking ability. Teachers reflected and developed the instrument (Figure 1) to facilitate student speaking based on a more authentic real-world context: seminars.

The instrument was originally developed recording students' *number of spoken turns* and *quality of contributions*. *Number of spoken turns* was selected to enable even the lowest proficiency students to make a score for speaking, whilst also being relatively straightforward for raters to count even in online conditions. As Rivers (1987) noted, students who are given more opportunities to interact in the language and who take the opportunity to speak more will improve their speaking skills. *Quality of contributions* was chosen to balance these spoken turns to reward students who built on the contributions of others, thus developing the discussion, instead of single or short utterances or simply reading from a prepared script. This is aligned with modern English proficiency assessment criteria for speaking which include "topic development" (Cambridge English, 2008, p. 1) and "interactive communication" which "maintains and develops interaction" (UCLES, 2008, p. 2).

These communicative interactions have been shown to act in tandem by Tsou (2005) who reported an increase in motivation and English academic scores for students who were assigned to an experimental group tasked with speaking more in classroom discussion compared to their control-group peers. Increased oral practice was suggested to significantly improve student speaking proficiency scores. Students in the experimental group self-reported higher levels of confidence believing they not only participated more in class when compared to previous classroom experiences, but also believed they improved in their speaking proficiency.

A 'traffic light' system was used to make the instrument relatively easy to use for teachers assessing the seminars online. The students were awarded zero (red), 1 (yellow) or 2 (green) depending on the number of times they exhibited each criterion, with two representing two or more. The maximum a student could receive on each criteria would be a 2. The score for *number of spoken turns* represents a simple count of the instances that occurred. Similarly, *quality of contributions* was scored as a count of the instances teacher-researcher's considered a student's turn reached or surpassed the threshold of a sufficient contribution as defined earlier. A score of 2 under *quality of contributions* does not indicate any better content quality than a score of 1, but only indicates the number of times the researcher felt this standard was attained. Whilst the decision of whether or not a given contribution met the requirement of developing the discussion was down to the rating teacher-researcher's opinion, the two teacher-researchers were working closely on the course and previous discussion lent a form of limited standardization between them. The color-coded system provided a convenient way for teachers to visualize students' performance. The instrument was tested in 12 classes over the first half of the semester after which

Students	Number of spoken turns	Quality of spoken contributions*
Student 1	1	1
Student 2	1	0
Student 3	2	1

Figure 1. Seminar instrument first edition.

changes were made based on student and teacher reflection. Two more categories were added based on this reflection: *initiative* and *evidence for preparation* (Figure 2).

Students called on each other forcing class members to answer before the addition of these categories. Teachers and students felt recognizing students who took the initiative to raise their hand and who had prepared materials on the discussion topic beforehand was important. These suggestions by students are similar to Tsou's (2005) research who also framed her study in Eastern classroom cultural values commenting on the passiveness of students and the potential detriment to language practice. The term and category *Initiative* was chosen by students and added to the rubric, but within the literature more accurate terminology would be willingness to communicate (WTC) (MacIntyre et al., 1998). *Initiative* was chosen to distinguish student voluntary WTC instead of being called on by student seminar leaders. High levels of classroom WTC have been correlated with higher second language proficiency scores (Baghaei et al., 2012). Willingness to communicate has also been shown to correlate with higher levels of integration by second language learners within authentic language contexts leading to greater learning outcomes (Yashima et al., 2004). Initiative was identified as students who raised their hand or students who added to class discussion without being called upon. Students were scored in the same simple count format as *number of spoken turns* with a maximum of 2.

Evidence of preparation was reported by students as important to classroom discussion with students stating by prepping information on the seminar topic they were better equipped to participate. Similarly, Gross et al. (2015) found flipped classrooms encouraged student engagement with students increasing preparation with course material leading to significant improvement in student performance when compared to students in a non-flipped classroom. *Evidence of preparation* was determined by the teacher to the degree to which students researched resources outside of the classroom and brought these ideas to class. Students who did not provide evidence were marked 0. Students who only provided surface (shallow) information or minimally contributed their resources during discussion were given a 1, and students who demonstrated in-depth research were marked 2.

The updated instrument was used for the following half semester with researchers and students both commenting positively on the outcomes, but a formal analysis was not conducted. In the current study, definitions and scoring methodology from previous instruments (Figures 1 and 2) were reviewed and accepted for the newly developed instrument (Figure 3). We looked to further develop the instrument by adding the *listening* portion for our study (Figure 3).

Listening is a cognitive process that is observed through behaviors (Witkin, 1990). Viewing listening through the lens of a communicative approach, we decided to rely on observable behaviors as a potential objective measure of listening, but also understood the difficulty of measuring listening due to the complex processes involved (Morrow, 1979). *Listening* was scored as a zero or 2 at 5-min intervals, meaning students either were or were not presenting observable listening behaviors. We defined positive observable listening behaviors as eye contact with the speaker, note-taking, and physical gestures of agreement or disagreement. Negative behaviors were inappropriate body posture such as head on desk, attention not focused on the presenters,

Students	Number of spoken turns	Quality of spoken contributions*	Initiative	Evidence of preparation
Student 1	1	1	1	1
Student 2				
Student 3	2	1	2	

Figure 2. Seminar instrument second edition.

Students	Speaking (0, 1 or 2 with 2 representing 2 or more)				Listening/engagement (off task=0; on task=2)			
	Number of spoken turns	Quality of spoken contributions*	Initiative	Evidence of preparation	at 5mins	at 10mins	at 15mins	at 20mins
Student 1	1	1	1	1	2	2	2	2
Student 2	1						2	2
Student 3	2	1	2			2	2	2

Figure 3. Discussion skills assessment.

talking to desk-mates on unrelated topics, and doing work unrelated to the topic presented. These lists are not exhaustive but provide a basis for what we perceived as positive and negative indicators of listening. Scores were assessed by teachers within the classroom during the seminars and discussion.

The overall *total speaking scores* were calculated as the sum of *number of spoken turns*, *quality of spoken contributions*, *initiative*, and *evidence of preparation*. Researchers chose to sum these individual categories for *total speaking scores* based on the literature presented earlier as well as the understanding that authentic speaking activities, such as seminar discussion, require multiple processes to work in tandem. We view these processes through a Gestalt lens believing the sum of the parts to be greater than the whole. Researchers agreed to run post-hoc tests as well, ensuring all criteria would also be assessed individually for differences.

The overall *total listening scores* were calculated as the total scores for each five-minute segment. As mentioned previously, listening and speaking are difficult to separate in authentic settings. We chose to focus only on observable listening behaviors because this was the first time trialing the instrument with the listening section. Future development may look to measure the interactions and influences between listening and speaking, but was outside the scope of this study.

Classroom discussion is a dynamic process in which learners collaborate to define and redefine understandings through a constant process of self-reflection and responding to classmates' ideas. Teacher-researchers also wished to align the instrument with IB practices of student-centered pedagogy by increasing student agency to develop student engagement and foster a positive classroom environment. We adopted a growth-mindset for student feedback seeing feedback as a mechanism of feedforward which we aligned our instrument with and communicated to students. Second language learners who reflect on self-level language performance may benefit from feedforward to plan for future improvement (Huang, 2016). Feedforward is further aligned with action research because both emphasize the improvement in the individual or group based on the cyclical research process and not necessarily end results (Piggot-Irvine, 2015). The 'traffic light' visual from the original instrument was repurposed as a means of emphasizing student reflection and feedforward. Scores were posted and provided a visual for students to assess personal performance.

Planning

We discussed the difficulties in authentic listening assessment as well as raising student engagement and student agency. Several methods were discussed to address each of these issues. Researchers settled on a seminar format delivery of classroom content. The collaborative development of this planning emulates Ferrance (2000) description for how action research supports teacher collaboration on identifying classroom problems which enhances professional development evidenced by this study.

This study resulted from the discussion between both researchers reflecting on the previous unit and difficulties eliciting student participation as well as planning ahead for the second unit. The IB

assesses four criteria for language acquisition: reading, listening, speaking, and writing. The topic for the second unit was people can have different points of view about creative scientific innovations and their purposes. The assessment criteria included listening and writing. The IB also requires teachers to set approaches to learning skills (ATL) for each unit. These skills fall into the categories of communication skills, social skills, self-management skills, research skills, and thinking skills. The focus for the unit was communication skills in which students negotiate ideas and knowledge with peers and teachers.

The emphasis for these student seminars would be placed on student communication with their peers to cooperatively build a discussion. A group of students was assigned each week to introduce and lead the seminar. The interaction between students and the discussion elicited from the introductory presentations was to serve as authentic practice of listening and speaking skills which would be useful in the DP program as well as undergraduate. The seminar format also lends itself to the chosen ATL skill making it an appropriate method for instruction because students and peers discussed topics creating new collective knowledge and understandings through dialogue. For assessment, the previous instrument referred to in [Figure 2](#) was edited to include the *listening* portion ([Figure 3](#)).

The class met three times a week for 40-min periods. Teachers planned lessons with the first lesson of each week as a student-led seminar on the advantages or disadvantages of a given topic about future innovations. Student groups chose from a list of preselected topics: AI/robots, space colonization, cloning, VR/metaverse, and body modification. The second lesson was hosted by the teacher giving the opposite opinion of the student seminar leaders. Students wrote an argumentative essay in the third class using information gained from the seminars. Permission for this action research was provided by the heads of school with the understanding this research aligned with the previously stated IB required PDP.

Implementing

Three classes were involved in this study with two being taught by teacher A and one class taught by teacher B. Teacher A's classes were comprised of 16 and 11 students respectively. Teacher B's class had 18 students. Teachers modeled the seminar format explaining the instrument and how students would be assessed during the first week. Students reflected with teachers on what were the positives and areas of improvement for each seminar. This feedforward growth mindset was instilled early in the course and modeled by teachers. We especially encouraged students to consider what parts of the teachers' presentations and which activities led to more discussion. In the second week, students chose topics and began to research content while in class. Students were given agency in choosing topics and group members. Researchers believed this agency may increase student engagement. Students formed groups of two to four members in each class. Student presentations began in the third week and lasted for 5 weeks with the sixth week used for the summative assessment.

Observing (Results)

Instrument

The sample size of the study was 43 participants which affected the normality of distribution. Shapiro-Wilk tests were calculated for the following variables: *total speaking scores*, *total listening scores*, *number of spoken turns*, *quality of spoken contributions*, *initiative*, and *evidence of preparation*. Shapiro-Wilk tests indicated a violation for the assumption of normality. There was also a violation for the assumption of homogeneity of variances. We chose not to remove outliers

and adopted non-parametric procedures in the form of Mann-Whitney U tests with the scores of the first and last seminar being compared.

We believed through progression and exposure of weekly seminars students' scores on the instrument would improve given time. Therefore, we chose to compare the instrument criteria scores of the first and last seminar to assess growth or change. The level of difficulty between the first and last seminar was assumed to be the same because the task remained the same throughout the course. Since the students chose their seminar topics, and the nature of the instrument itself, they likely chose topics they were sufficiently familiar with and comfortable enough to present to their peers, meaning the degree of difficulty for topic content would remain consistent.

The instrument only had two criteria, *quality of spoken contributions* and *evidence of preparation*, which could possibly be conceptually aligned with difficulty of seminar content. However, for the students leading the seminar, *evidence of preparation* was measured by PowerPoint content, specifically if there was or was not material presented by the group. Each group presented material so *evidence of preparation* was consistent across presenters with no difference in difficulty. *Quality of spoken contributions* was also assumed to be similar in degree of difficulty across the different weekly topics because students should have been familiar with seminar content prior to the discussion taking place because they had the chance to prepare. Seminar leaders were given ample opportunity to demonstrate some knowledge of the topic if they wished during their opening of the session and throughout the seminar. For students who were not seminar leaders, but were a part of the seminar discussion, it was assumed these students had enough conversational skills, justified by student APTIS scores, to participate in the conversation even if they did not have a prior in-depth understanding of the topic being discussed.

A Mann-Whitney U test was run to determine if there were differences between the first and last seminar for overall *total speaking scores* and *total listening scores*. *Total speaking scores* were not significantly different, but *total listening scores* were statistically significant. Mean *total listening scores* were statistically significantly higher in the last seminar (6.93) than in the first seminar (5.95), $U = 621$, $z = -2.064$ $p = .009$.

	Statistic	df	<u>p</u>			
Mann-Whitney U	621		0.009			
	Group	N	Mean	Median	SD	SE
Total Listening Score	TLS1	43	5.95	6.00	2.25	0.343
	TLS2	41	6.93	8.00	2.10	0.328

Post-hoc tests were run for individual criteria. The variables *quality of spoken contributions* and *evidence of preparation* did not demonstrate a significant difference when comparing the first and last seminar. A Mann-Whitney U test was run to determine if there were differences for *number of spoken turns* and *initiative* between the first seminar and the last seminar. Mean *number of spoken turns* were statistically significantly higher in the last seminar (1.66) than in the first seminar (1.07), $U = 589$, $z = -3.498$ $p = .003$.

	Statistic	df	<u>p</u>			
Mann-Whitney U	589		0.003			
	Group	N	Mean	Median	SD	SE
Spoken Turns Score	SPT1	43	1.07	1.00	0.936	0.143
	SPT2	41	1.66	2.00	0.575	0.089

Mean *initiative scores* were statistically significantly higher in the last seminar (1.44) than in the first seminar (0.930), $U = 625$, $z = -2.676$ $p = .012$.

	Statistic	df	p			
Mann-Whitney U	625		0.012			
	Group	N	Mean	Median	SD	SE
Initiative Score	IN1	43	0.930	1.00	0.936	0.143
	IN2	41	1.44	2.00	0.808	0.126

Reflecting (Discussion)

Results indicate students did significantly increase in observable listening behaviors. A significant difference for the variable of *listening* between the first and last seminar could be interpreted as indicative of increased student engagement, thus supporting the purpose of our study. Aligned with this aim, we hoped this instrument would serve as a form of feedforward fostering student engagement through student agency and student-centered learning practices.

Initiative and *number of turns* spoken were also significant. Both variables indicated a positive increase from the first to last seminar. Students increased the number of opportunities they voluntarily added to the discussion. We believe this increase in engagement is related to the student agency we embedded in the seminar learning activity mentioned earlier. There may have been some influence from feed-forward as well, but with the variables *quality of contributions* and *evidence of preparation* not indicating a positive increase, the degree of the feedforward influence of the instrument on our students cannot be determined.

While students increased their classroom engagement and contributed more to class discussions, the quality of these contributions did not improve over time with the averages remaining low overall. *Quality of contributions* was defined as introducing a new point, extending or developing an idea, or challenging the speaker. Our observations demonstrate students have the most difficulty in this area of the seminar format. Teacher A observed scores on this criterion often came from students posing questions of the group leading the seminar rather than developing or challenging ideas. This may be related to evidence of preparation. Without preparing for the seminars, students were unable to make quality contributions supporting the discussion. However, a correlation for *quality of contributions* and *evidence of preparation* was not significant. Preparation did not influence the quality of student contributions to class discussion, which seems to indicate better preparation does not mean a student will necessarily have a better quality contribution. This is counterintuitive to what teacher-researchers and students expected and has caused us to reflect on how to better improve quality of contributions, how to more accurately measure these criteria, and to further explore the relationship between these two variables.

The relationship between motivation and performance requires one to be motivated and engaged before improving academic achievement (Ito & Umemoto, 2022). Our results indicate the same progression, with students first increasing in engagement and motivation, signified by increases in students voluntarily adding to class discussion, before showing significant progress in quality of contributions. We interpret these results as positive meaning the seminar format as well as instrument supported student engagement and to some degree learning. We believe a third round of action research is required to better scaffold *quality of contributions* and *evidence of preparation* for students. We noted that the student seminar leaders typically scored higher in *evidence of preparation* than the other students. This is expected because the student seminar leaders usually came with a prepared PowerPoint presentation. Student seminar leaders

necessarily had to introduce the topic and ask questions to encourage discussion so also scored well on *initiative* and *quality of contributions*.

Limitations of the seminar assessment instrument were identified through this study. The *listening* score was based on our observation of students' behavior at specific times during the seminar. Thus, this is an indicator of students' appearance of engagement rather than their listening, especially given the lack of progress in *quality of contributions* and the requirement to build on or challenge other students' ideas. While absence of distraction is a prerequisite for successful listening, and conversational management and appropriate paralinguistic skills undoubtedly contribute to successful discussion (Robinson et al., 2001), a more direct measure of successful listening comprehension would help further validate the instrument (Nation & Macalister, 2010). We will further amend the assessment instrument considering additional methods of directly assessing listening skills. Some students objected to the cap of two points for each criterion on the assessment instrument, expressing that a raw score with no cap would more fairly reflect their performance. We see this critique of the instrument as a positive indicator of student buy-in to the study with students beginning to take ownership of their learning and assessment giving us the confidence to further involve them in future development of the seminar format and assessment instrument.

Re-planning (Future Research)

We look to further develop upon the seminar format and our instrument in a third round of action research in our next unit. Student engagement is established, demonstrated through the increase in *number of times spoken* and *initiative*. We wish to now focus on criteria for speaking assessment and expand upon the listening to assess more than observable behaviors. For speaking we will develop requirements for vocabulary, grammar, fluency, coherence, and pronunciation, which are currently left implicit in the instrument. Listening will be evidenced through spoken output as well as written graphic organizers helping students to take notes of the seminar discussion. We also look to further support students' in developing quality contributions by subdividing this criterion, with the requirements to build on or challenge others' ideas contributing to the listening score as well as the speaking score.

In this way, the speaking output also functions as an indicator of listening comprehension which could further validate the instrument in assessing listening. Evidence of preparation does not seem to have an effect on *quality of contributions* and we are reflecting on whether this section should be removed or somehow supported through the use of the planned graphic organizer. Students could use their notes on the seminar to create connections to previous knowledge in their preparatory work. These mental maps may also serve as a scaffold to support students' development of critical thinking skills and possibly provide another measurable item to assess student listening.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

John Stewart Clark  <https://orcid.org/0000-0003-4475-3832>

References

- Alexander, O., Argent, S., & Spencer, J. (2018). *EAP essentials* (2nd ed.). Garnet Publishing Ltd.
- Baghaei, P., Dourakhshan, A., & Salavati, O. (2012). The relationship between willingness to communicate and success in learning English as a foreign language. *MJAL*, 4(2), 53–67.
- Bardovi-Harlig, K., & Griffin, R. (2005). 12 pragmatic awareness: Evidence from the Esl classroom. *System*, 33(3), 401–415. <https://doi.org/10.1016/j.system.2005.06.004>
- Bardovi-Harling, K., Hartford, B. A. S., Mahan-Taylor, R., Morgan, M. J., & Reynolds, D. W. (1991). Developing pragmatic awareness: Closing the conversation. *ELT Journal*, 45(1), 4–15. <https://doi.org/10.1093/elt/45.1.4>
- Brophy, J., & Good, T. L. (1986). *Handbook of research on teaching* (3rd ed.). MacMillan.
- Cambridge English. (2008). *IELTS speaking band descriptors*. <https://assets.cambridgeenglish.org/webinars/ielts-speaking-band-descriptors.pdf>
- Carless, D. (2007). Learning oriented assessment: Conceptual bases and practical implications. *Innovations in Education & Teaching International*, 44(1), 57–66. <https://doi.org/10.1080/14703290601081332>
- Cohen, L., Manion, L., & Morrison, L. (2018). *Research methods in education* (8th ed.). Taylor & Francis Group.
- Common, E. A., Lane, K. L., Cantwell, E. D., Brunsting, N. C., Oakes, W. P., Germer, K. A., & Bross, L. A. (2020). Teacher-delivered strategies to increase students' opportunities to respond: A systematic methodological review. *Behavioral Disorders*, 45(2), 67–84. <https://doi.org/10.1177/0198742919828310>
- Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching*, 4(2), 237–274. <https://doi.org/10.14746/ssl.2014.4.2.5>
- Dewaele, J.-M., Witney, J., Saito, K., & Dewaele, L. (2018). Foreign language enjoyment and anxiety: The effect of teacher and learner variables. *Language Teaching Research*, 22(6), 676–697. <https://doi.org/10.1177/1362168817692161>
- Diepenbroek, L. G., & Derwing, T. M. (2013). To what extent do popular ESL textbooks incorporate oral fluency and pragmatic development? *TESL Canada Journal*, 30(7), 1–20. <https://doi.org/10.18806/tesl.v30i7.1149>
- Dulfer, N., & Koopaei, F. A. (2021). Moving from feedback to feedforward in IBDP classrooms. *Journal of Research in International Education*, 20(2), 101–113. <https://doi.org/10.1177/14752409211032528>
- Ferrance, E. (2000). *Action research*. Providence, Rhode Island: Northeast and Islands Regional. Educational Laboratory at Brown University.
- Feyten, C. M. (1991). The power of listening ability: An overlooked dimension in language acquisition. *The Modern Language Journal*, 75(2), 173–180. <https://doi.org/10.1111/j.1540-4781.1991.tb05348.x>
- Finkel, D. L. (1999). *Teaching with your mouth shut*. Heinemann.
- Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40(2), 97–118. <https://doi.org/10.1017/S0261444807004144>
- Gross, D., Pietri, E. S., Anderson, G., Moyano-Camihort, K., & Graham, M. J. (2015). Increased preclass preparation underlies student outcome improvement in the flipped classroom. *CBE-Life Sciences Education*, 14(4), ar36. <https://doi.org/10.1187/cbe.15-02-0040>
- Han, C. (1992). A comparative study of compliment responses: Korean females in Korean interactions and in English interactions. *Working Papers in Educational Linguistics*, 8(2), 17–31. <https://files.eric.ed.gov/fulltext/ED354726.pdf>
- He, L. Z., & Jiang, Z. Y. (2020). Assessing second language listening over the past twenty years: A review within the socio-cognitive framework. *Frontiers in Psychology*, 11, 2123. <https://doi.org/10.3389/fpsyg.2020.02123>

- Huang, S. C. (2016). Understanding learners' self-assessment and self-feedback on their foreign language speaking performance. *Assessment & Evaluation in Higher Education*, 41(6), 803–820. <https://doi.org/10.1080/02602938.2015.1042426>
- International Baccalaureate Organization (IBO). (2022). *Curriculum*. <https://www.ibo.org/programmes/middle-years-programme/curriculum/>
- Ito, T., & Umemoto, T. (2022). Examining the causal relationships between interpersonal motivation, engagement, and academic performance among university students. *PLoS One*, 17(9), 02742299. <https://doi.org/10.1371/journal.pone.0274229>
- Jiang, Y., & Dewaele, J.-M. (2019). How unique is the foreign language classroom enjoyment and anxiety of Chinese EFL learners? *System*, 82, 13–25. <https://doi.org/10.1016/j.system.2019.02.017>
- Koen, M., Bitzer, E. M., & Beets, P. A. D. (2012). Feedback or feedforward? A case study in one higher education classroom. *Journal of Social Sciences*, 32(2), 231–242. <https://doi.org/10.1080/09718923.2012.11893068>
- Koester, A. J. (2002). The performance of speech acts in workplace conversations and the teaching of communicative functions. *System*, 30(2), 167–184. [https://doi.org/10.1016/S0346-251X\(02\)00003-9](https://doi.org/10.1016/S0346-251X(02)00003-9)
- Lee, J.-S. (2013). The relationship between student engagement and academic performance: Is it a myth or reality? *The Journal of Educational Research*, 107(3), 177–185. <https://doi.org/10.1080/00220671.2013.807491>
- Limberg, H. (2016). Teaching how to apologize: EFL textbooks and pragmatic input. *Language Teaching Research*, 20(6), 700–718. <https://doi.org/10.1177/1362168815590695>
- Luo, J., & Chan, C. K. (2023). Twenty years of assessment policies in China: A focus on assessing students' holistic development. *International Journal of Chinese Education*, 12(2), 2212585X2311731. <https://doi.org/10.1177/2212585X231173135>
- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545–562. <https://doi.org/10.1111/j.1540-4781.1998.tb05543.x>
- Melrose, M. (2001). Maximising the rigor of action research: Why would you want to? How could you? *Field Methods*, 13(2), 160–180. <https://doi.org/10.1177/1525822X0101300203>
- Morrow, K. (1979). Communicative language testing: Revolution or evolution? In C. J. Brumfit & K. Johnson (Eds.), *The communicative approach to language teaching* (pp. 143–159). Oxford, UK: Oxford University Press.
- Nation, I. S. P., & Macalister, J. (2010). *Language curriculum design*. Taylor & Francis.
- Neumann, J. W. (2013). Developing a new framework for conceptualizing “student-centered learning”. *The Educational Forum*, 77(2), 161–175. <https://doi.org/10.1080/00131725.2012.761313>
- Nguyen, M. (2011). Learning to communicate in a globalized world: To what extent do school textbooks facilitate the development of intercultural pragmatic competence? *RELC Journal*, 42(1), 17–30. <https://doi.org/10.1177/0033688210390265>
- Noman, M., Kaur, A., Mullick, J., & Ran, L. (2023). Navigating new terrain: First-year Chinese students' transitionary experiences in a Sino-US joint venture university in China. *International Journal of Chinese Education*, 12(2), 2212585X2311751. <https://doi.org/10.1177/2212585X231175167>
- Ockey, G., & Wagner, E. (2018). *Assessing L2 listening ability: Moving towards authenticity*. John Benjamins.
- Orsmond, P., Maw, S. J., Park, J. R., Gomez, S., & Crook, A. C. (2011). Moving feedback forward: Theory to practice. *Assessment & Evaluation in Higher Education*, 38(2), 240–252. <https://doi.org/10.1080/02602938.2011.625472>
- Piggot-Irvine, E. (2015). Reflecting on evidence: Leaders use action research to improve their teacher performance reviews. *The Canadian Journal of Action Research*. 16(3). 3-26. <https://doi.org/10.33524/cjar.v16i3.225>

- Ren, W., & Han, Z. (2016). The representation of pragmatic knowledge in recent ELT textbooks. *ELT Journal*, 70(4), 424–434. <https://doi.org/10.1093/elt/ccw010>
- Rivers, W. M. (Ed.). (1987). *Interactive language teaching*. Cambridge University Press.
- Robinson, P., Strong, G., Whittle, J., & Nobe, S. (2001). *Research perspectives on English for academic purposes*. Cambridge University Press.
- Rodríguez-Gómez, G., & Ibarra-Sáiz, M. S. (2015). Assessment as learning and empowerment: Towards sustainable learning in higher education. In M. Peris-Ortiz & J. Merigó Lindahl (Eds.), *Sustainable learning in higher education. Innovation, technology, and knowledge management*. Springer. https://doi.org/10.1007/978-3-319-10804-9_1
- Shao, K., Pekrun, R., Marsh, H. W., & Loderer, K. (2020). Control-value appraisals, achievement emotions, and foreign language performance: A latent interaction analysis. *Learning and Instruction*, 69, 101356–101412. <https://doi.org/10.1016/j.learninstruc.2020.101356>
- Skinner, E. A., Wellborn, J. G., & Connell, J. P. (1990). What it takes to do well in school and whether I've got it: A process model of perceived control and children's engagement and achievement in school. *Journal of Educational Psychology*, 82(1), 22–32. <https://doi.org/10.1037/0022-0663.82.1.22>
- Tomaszewski, W., Xiang, N., Huang, Y., Western, M., McCourt, B., & McCarthy, I. (2022). The impact of effective teaching practices on academic achievement when mediated by student engagement: Evidence from Australian high schools. *Education Sciences*, 12(5), 358–414. <https://doi.org/10.3390/educsci12050358>
- Tsou, W. (2005). Improving speaking skills through instruction in oral classroom participation. *Foreign Language Annals*, 38(1), 46–55. <https://doi.org/10.1111/j.1944-9720.2005.tb02452.x>
- University of Cambridge Local Examinations Syndicate (UCLES). (2008). *Assessing speaking performance level B2*. <https://www.cambridgeenglish.org/images/168619-assessing-speaking-performance-at-level-b2.pdf>
- Wagner, E. (2013). *Assessing listening. The companion to language assessment* (Vol. 1, pp. 47–63). John Wiley & Sons. <https://doi.org/10.1002/9781118411360.wbcla094>
- Wagner, E. (2014). Using unscripted spoken texts to prepare L2 learners for real world listening. *TESOL Journal*, 5(2), 288–331. <https://doi.org/10.1002/tesj.120>
- Wagner, E., Liao, Y.-F., & Wagner, S. (2021). Authenticated spoken texts for L2 listening tests. *Language Assessment Quarterly*, 18(3), 205–227. <https://doi.org/10.1080/15434303.2020.1860057>
- Wagner, E., & Toth, P. (2014). Teaching and testing L2 Spanish listening using scripted vs. unscripted texts. *Foreign Language Annals*, 47(3), 404–422. <https://doi.org/10.1111/flan.12091>
- Wagner, E., & Wagner, S. (2016). Scripted and unscripted spoken texts used in listening tasks on high stakes tests in China, Japan, and Taiwan. In V. Aryadoust & J. Fox (Eds.), *Trends in language assessment research and practice: The view from the Middle East and the Pacific Rim* (pp. 103–123). Cambridge Scholars.
- Wang, Y. L., Derakhshan, A., & Zhang, L. J. (2021). Researching and practicing positive psychology in second/foreign language learning and teaching: The past, current status and future directions. *Frontiers in Psychology*, 12, 1–10. <https://doi.org/10.3389/fpsyg.2021.731721>
- Weaver, R. R., & Qi, J. (2005). Classroom organization and participation: College students' perceptions. *The Journal of Higher Education*, 76(5), 570–601. <https://doi.org/10.1353/jhe.2005.0038>
- Witkin, B. R. (1990). Listening theory and research: The state of the art. *International Listening Association. Journal*, 4(1), 7–32. https://doi.org/10.1207/s1932586xijl0401_3
- Hu, X., Sidhu, G. K., & Lu, X. (2022). Relationship between growth mindset and English language performance among Chinese EFL university students: The mediating roles of grit and foreign language enjoyment. *Frontiers in Psychology*, 13, 1–14. <https://doi.org/10.3389/fpsyg.2022.935506>
- Xu, J., & Kou, J. (2017). Group interaction strategies and students' oral Performance in Chinese EFL classrooms. *Tesol Quarterly*, 52(1), 198–209. <https://doi.org/10.1002/tesq.398>

- Yashima, T., Zenuk-Nishide, L., & Shimizu, K. (2004). The influence of attitudes and affect on willingness to communicate and second language communication. *Language Learning*, 54(1), 119–152. <https://doi.org/10.1111/j.1467-9922.2004.00250.x>
- Yu, W., Wu, H., & Zhao, W. (2022). A qualitative analysis of control-value appraisals, positive achievement emotions, and EFL performance in a Chinese senior high school context. *Frontiers in Psychology*, 13, 1–10. <https://doi.org/10.3389/fpsyg.2022.986684>
- Zhang, X., Huang, Y., & Liu, Y. (2022). Enhancing language task engagement in the instructed language classroom: Voices from Chinese English as a foreign language students and teachers. *International Journal of Chinese Education*, 11(2), 2212585X2210977. <https://doi.org/10.1177/2212585X221097718>