Using Peer-Led Evaluation in Productive and Receptive L2 Coursework

Jeffrey Martin
Momoyama Gakuin University
Sachi Oshima
Chuo Gakuin University

Reference Data:

Martin, J., & Oshima, S. (2024). Using peer-led evaluation in productive and receptive L2 coursework. In B. Lacy, R. P. Lege, & P. Ferguson (Eds.), *Growth Mindset in Language Education*. JALT. https://doi.org/10.37546/JALTPCP2023-07

Aligning student learning with second language course objectives can include considering the question of who evaluates whom. For instance, allowing students to lead evaluation or create evaluation materials provides them with opportunities to directly engage with learning goals on their own terms. Working within a supportive framework for this purpose, set up by the teacher, students follow a peer-exchange process that uniquely draws their attention to language performance through the act of evaluation. As learners compare, contrast, think, and communicate with peers, from a position of agency, they can also develop self-efficacy and metacognitive skills. The purpose of this paper is to outline how ***peer-led evaluation*** can be implemented for both productive and receptive language-focused coursework, elaborate on potential benefits for the students, and conclude with a list of practical tips that can help teachers create such coursework according to the circumstances and aims of their teaching.

第二言語学習において、学習目標に沿った学生の学びには、「誰が誰を評価するか」という考慮も必要である。例えば、学生が他の学生を評価したり、教材を作成したりすることで、学生は自ら学習目標に直接関わる機会を得ることができる。教師が設定する支援的な枠組みのなかで、学生は互いの学習成果を交換し、評価という行為を通じて互いの言語活動に注目することになる。主体的な立場から、仲間と比較・対照し、考え、コミュニケーションを取ることにより、学習者は自己効力感やメタ認知スキルを高めることもできる。本論文では、受講者主導の相互評価（peer-led evaluation）を、生産的または受動的言語スキルを対象とした授業でどのように実施できるかを概説するとともに、受講者主導の相互評価が第二言語学習者にもたらす効果を論ずる。最後に、各教師が自身の環境や授業目的に応じて、受講者主導の相互評価を授業に取り入れる際の実践的なヒントについても紹介する。

Students of second language (L2) courses are familiar with their teachers assessing their presentations, scoring their quizzes, tallying their journal entries, and so on. Naturally, teachers can provide helpful feedback and add other formative steps to such assignments through a variety of valuable pedagogical approaches which need no introduction here. Rather, the contribution of this paper is to present a practice-oriented resource on how evaluation of L2 performance can be in part led by the L2 learners themselves. Peers evaluating each other's presentations and peers making quizzes to be exchanged, for example, are ideas for language courses that can allow students to assume mutual responsibility for learning, as though being allowed to "drive the car" as a course progresses over a school term. L2 learners can assume such responsibility within a supportive classroom procedure and engage in learning objectives with agency and from different perspectives.

In this paper, we first consider the influence of testing and peer-exchange on students' perceptions of their coursework. We also distinguish the two peer-exchange roles of evaluator and evaluation materials creator and the modality of these L2 language foci—the language skills used when students exchange samples of their L2 output (e.g., essay writing) compared to when they exchange self-made evaluation materials for eliciting L2 input (e.g., reading quizzes). Then, we introduce the term *peer-led evaluation* to tie these forms of peer-exchange together within a broad scope that applies to both productive (writing, speaking) and receptive (reading, listening) language-focused coursework. From this perspective, we provide one sample project each to showcase the two roles, illustrate procedural differences, and summarize the benefits of peer-led evaluation according to theory and empirical research. Setting up an evaluation project for students to lead, however, requires multi-step planning by L2 teachers to meet pedagogical aims, and it can be challenging for teachers to adequately foresee what design features can optimize a

project's success. To address this, we conclude with a list of five practical tips. These can be flexibly applied as needed to implement peer-led evaluation projects in alignment with a teacher's chosen L2 course objectives.

The Influence of Testing on Language Learning

The testing conducted in language classrooms tends to shape students' perspectives on L2 study and ultimately the focus of their learning. With testing being decided by teachers and administrators, it is natural for students to adjust accordingly and work to receive their desired grades. Recognizing this impact of language testing, educators and researchers have held a longstanding interest in adequately matching evaluation and assessment practices in formal language education to principled course objectives. Standardized language proficiency tests and university entrance exams are also prominent topics. Key interests in the educational measurement literature include formative and summative testing (Bachman, 1990), transferability of tested skills (Carpenter, 2012), consequential validity (Messick, 1998), washback effects (Cheng et al., 2015), and test impact in specific contexts such as in Japan (Allen & Tahara, 2021). This body of research has advanced the development of language testing in terms of *what* productive and receptive language abilities can and should be tested. A parallel consideration has been the exploration of *who* can take on testing roles.

Students Learning With and From Peers

In education broadly, students in a course form a resource for each other as peers and their perspectives are a valuable influence on learning. Insights about this influence and the impact of testing have driven development of peer-related pedagogy in the wider field of educational psychology (e.g., Falchikov, 2001; Liu & Carless, 2006; Topping, 1998). In university-level education, academics such as Boud et al. (1999) have discussed the impact of testing in that "assessment is the single most powerful influence on learning in formal courses" (p. 413), and that *peer learning* is a principled approach for aligning assessment efforts with formative learning objectives. Emphasis is on students encountering opportunities to learn with and from each other without teachers immediately intervening. Introducing testing from this perspective can help students focus on collaboration and self-awareness, communicate more in the subject matter, and develop transferable learning skills.

Related pedagogies are *peer feedback* and *peer assessment*, with peer feedback being "primarily about rich detailed comments but without formal grades, whilst peer assessment denotes grading irrespective of whether comments are also included" (Liu & Carless, 2006, p. 280). Topping (1998) defined peer assessment as an arrangement for peers of equal status to consider the level, value, or worth of each other's products or outcomes of learning. Topping and Ehly (1998) placed these feedback and assessment approaches under a paradigm of *peer-assisted learning*, with its mechanism for learning accounted for by *cognitive conflict* (Piaget, 1971). It explains how students' cognitive development is stimulated through the processes of adaptation and resolution as they grapple with differences between prior and new experiences with peers of relatively equal ability but with varied competencies.

Topping (1998) further describes this pedagogy as *learning by assessing*, where students learn from the process of using a standard for evaluation (i.e., a rubric) to assess peers' work. In doing so, students can glean as much or more from this process than from received guidance. Students formulate questions to ask peers and themselves, which drives engagement and reasoning (Graesser et al., 1996). The peer-exchange does not need to be face-to-face. For example, it could be an anonymous exchange of student writing. Depending on the learning project, teachers facilitate the process by providing scoring rubrics, check lists, and other coursework materials. In the end, students have increased their time thinking, comparing, contrasting, and communicating. Topping (1998) highlighted that these activities are intellectually challenging and have the potential to consolidate and reinforce learning.

Evaluating Both Productive and Receptive L2 Performance

Peer-exchange research in the field of teaching English to speakers of other languages (TESOL) has focused on the development of L2 production in written form (e.g., Farrokhi et al., 2012; Holster et al., 2013; Hyland & Hyland, 2006; Yu & Lee, 2016) and spoken form (e.g., Cheng & Chau, 2009; Rodríguez-González & Castañeda, 2018; Rotsaert et al., 2018; Zhang et al., 2020). The exchanges in these studies have made use of learners' first and second languages depending on L2 proficiency and project design. Input and output language processing are built into peer-exchange learning projects, but to our knowledge, qualities of receptive L2 language performance are not typically emphasized. The exchange of essays and presentations in peer-exchange research has been in service of having L2 learners engage with the language production of their peers.

In contrast, our conception of peer-exchange and evaluation of L2 performance includes both the productive and receptive language foci and it distinguishes their features. An example of the latter is having students create reading quizzes based

on a level-appropriate reading source, work within a small support group of peers, and exchange the quizzes within the larger group of peers. In contrast to essay and presentation projects, this receptive language-focused example involves students writing quiz items. However, the creation of such quizzes is in service of having L2 learners engage with the receptive language abilities of their peers.

Peer-led Evaluation

The purpose of this paper is to present a practice-oriented resource for L2 teachers who want to incorporate *peer-led evaluation* and map out the differences and commonalities between productive and receptive language-focused peer-exchange coursework. This approach can help facilitate a wide range of peer-exchange procedures that align L2 performance with specific course objectives. The approach is implemented with intact groups of students in L2 courses. Language programs can set a common syllabus, or teachers can design their own specific plans. In any case, adopting this approach entails a common sequence of stages: peers lead each other from awareness-raising and preparation stages to a group-wide exchange among classmates of the course. They then inspect a range of received information and conclude with a reflection on all that they experienced.

Certainly, *evaluation* and *assessment* have been interrelated in the literature, but the present use of evaluation is meant to denote a continuing learning process rather than the concluding outcome of an assessment. This distinction is reflected in the work of testing experts, such as Harding and Kremmel (2021), who argued that assessment is a broad action that "leads to decision-making" (p. 55) and comes from the evaluations of performances. Conceptual overlap remains, however, in that formative assessment (e.g., Bachman, 1990) does not imply summative testing but denotes continuing feedback as teachers evaluate student development.

Our main line of reasoning for adopting *evaluation* is that the word denotes cognitive development as described in psychology research. Evaluation is central to question-driven reasoning in distributed learning environments (Graesser et al., 1996) and the word describes a high level in Bloom's revised taxonomy of higher-order thinking skills (cognitive domain; Anderson & Krathwohl, 2001; Bloom, 1956). Evaluation of L2 performance can be facilitated for both L2 productive language focused exchanges (e.g., oral presentations or written essays) and L2 receptive language focused exchanges (e.g., comprehension quizzes using written or audio/visual input). For either focus, students in peer-led evaluation projects encounter continuing opportunities to develop their language skills. Building self-awareness of their learning (Boud et al., 1999), self-efficacy (Bandura, 1997), and higher-order thinking skills (Anderson & Krathwohl, 2001) are other potential benefits.

Stages of Peer-led Evaluation

Procedures will vary based on circumstances and course objectives, but the following outlines a common progression of stages applicable to either productive or receptive language-focused L2 coursework:

- *Awareness-raising Stage* – Students are made aware of the learning objectives
- *Preparation Stage* – Students prepare the work that they will exchange
- *Peer-exchange Stage* – Students exchange the work among peers
- *Patterns-in-data Stage* – Students inspect the information that they receive
- *Reflection Stage* – Students reflect on their experiences across all stages

Features of Peer-led Evaluation

In our conceptualization, three features are present when students of a course take leading roles in evaluating either productive or receptive language performances of each other.

- *Students engage with a standard of evaluation* – This standard is based on the learning objectives of the course. Teacher-provided rubrics and peer-made quizzes can respectively serve as the standards of evaluation for productive and receptive language-focused L2 coursework.
- *Students engage with L2 performances of peers* – Students engage with writing or speaking performances of peers at the Peer-exchange Stage in a productive language-focused project. Students engage with reading or listening performances of peers at the Pattern-in-data Stage in a receptive language-focused project.
- *Students engage with feedback from peers* – Students receive informal peer-feedback from support group members at the Preparation Stage. Other peer-feedback is received differently through peer-exchange according to stage, language skill focus, and the teacher's design choices (e.g., open or anonymous exchange).

Options for Key Language Skills

Peer-led evaluation can be used with a focus on any of the key language skills of speaking, writing, reading, and listening. According to circumstances and teaching aims,

teachers can design their projects in numerous ways. See Table 1 for a short list of ideas. The ideas for speaking (Sample 1) and reading (Sample 2) are featured in the following section.

Table 1.
Peer-led Evaluation Ideas for Four Key Language Skills

Skill	Description of Idea	Source
Speaking	Repeated presentation exchange, building self-efficacy (see Sample 1)	Oshima (2022)
Listening	Exchange of listening activities for chosen songs, tapping feet allowed	Oshima (2020)
Writing	Anonymized essay exchange, experience being the "unknown reader"	Martin (2020)
Reading	Anonymized reading quiz exchange, questioning is thinking (see Sample 2)	Martin (2023)

Reasons for Procedural Differences in How Peers Engage with L2 Performance

Important to note, there are procedural differences in peer-led evaluation by language focus due to the different natures of experiencing the productive or receptive L2 performances of others. Performance in peers' language production (speaking or writing) is witnessed at the Peer-exchange Stage. Performance in peers' language reception or comprehension (reading or listening) can be ascertained from the results of a testing instrument such as a comprehension quiz, revealed at the Patterns-in-data Stage. Figures 1a and 1b overleaf illustrate these differences by stage.

Two Sample Projects

Teachers can implement peer-led evaluation language projects and adjust their details according to circumstances and course objectives. To convey the differences and similarities in procedure, we present two sample projects in parallel while following the common sequence of stages.

Sample 1: Oral Presentation Exchange Project (Productive Language Focus)

This simplified sample project is based on Oshima (2022) and it is about exchanging video-recorded presentations. Oshima (2022) was set in a language course of 22 beginner-level students that met twice a week over a 15-week term. This video-recording method is equally applicable to online and face-to-face classroom settings. As an option following the Awareness-raising Stage, the teacher can have students do two or three presentations by repeating the other stages every few weeks. Three example topics:

- Topic #1 – Self-introductions with details such as hobbies and future ambitions.
- Topic #2 – Introducing interesting trivia about their university
- Topic #3 – Sharing future travel plans with details and reasons why

Sample 2: Reading Quiz Exchange Project (Receptive Language Focus)

The next simplified sample project is based on Martin (2023) and it is about making and exchanging multiple-choice reading comprehension quizzes. Martin (2023) was set in an L2 reading course of 21 students who met once per week over 15 weeks. For this procedure, students are in support groups of three classmates each, and the peer-exchange is anonymous. Supported by peers and the teacher, each student creates one reading quiz based on simplified news articles from a website for L2 learners. Optionally, the teacher could have each student make two or more quizzes. Another alternative is to have students use articles directly from newspapers.

Samples 1 and 2: Sequence of Stages Presented in Parallel

Awareness-raising Stage

Common – The students need to build their awareness of the project's objectives and the procedure (See Tip #1 below about making the students' procedure simple from their perspective). Assigning classmates to small support groups is advised (see Tip #2). Initial awareness-raising is important, but students further extend their understanding from stage to stage.

Sample 1: Oral Presentations – The presentation project's objectives are reflected in the categories of the rubric (and are the standard for evaluation). Examples of categories are organization of content, body language, and communication strategies (e.g., Canale, 1983), and so on, as set by the teacher. Each category should have informative descriptors and be made as easy to use as possible. Teachers should consider providing a bilingual

Figure 1a.
Common Progression of Project Stages, Procedural Features of L2 Productive Language Focus

Peer-led Evaluation – L2 output (i.e., peers' essays or oral presentations)			
Focus	L2 productive language skill building		
Standard of Evaluation	Categories from a rubric (provided by teacher)		
	Unique to peer-led evaluation, students engage with…		
	…a standard of evaluation		
		…L2 (productive) performances of peers	
Stage			…feedback from peers
Awareness-raising	Yes		
Preparation	Yes (but not unique to peer-led evaluation		Yes (informally, if in support groups)
Peer-exchange	Yes	Yes	
Patterns-in-data	Yes		Yes
Reflection	Yes (freely referenced)	Yes (recalled from Peer-exchange Stage)	Yes (received from Patterns-in-data Stage)

Figure 1b.
Common Progression of Project Stages, Procedural Features of L2 Receptive Language Focus

Peer-led Evaluation – L2 input (i.e., peers' reading or listening quizzes)			
Focus	L2 receptive language skill building		
Standard of Evaluation	Set of quizzes (made by peers)		
	Unique to peer-led evaluation, students engage with…		
	…a standard of evaluation		
		…L2 (receptive) performances of peers	
Stage			…feedback from peers
Awareness-raising	in development		
Preparation	in development		Yes (informally, likely from support group)
Peer-exchange	Yes (peer-made)	Students take these quizzes as in any reading quiz exercise	
Patterns-in-data	Yes (self-made)	Yes	
Reflection	Yes (recalled quizzes from peers while freely referencing own)	Yes (received from Patterns-in-data Stage)	Yes (recalled from Preparation Stage)

rubric and setting scales to six points (see Tip #3). Teachers should explicitly explain the content of the rubric so students can best use it.

Sample 2: Reading Quizzes – Question-making reflects the learning objectives for reading (and the quizzes become the standard for evaluation). To build this awareness, students work together and consider different kinds of thinking needed to answer questions from, for example, a critical reading textbook. See Tip #4 regarding how question type relates to reading strategies such as differentiating gist from details, making inferences about unfamiliar vocabulary, and making inferences about implied details.

Preparation Stage

Common – The students work with the teacher and within support groups. Informal feedback between groupmates is encouraged.

Sample 1: Oral Presentations – Students have one week (two sessions) to make their short two-minute video presentations about the given topic using reference materials. In preparing for the exchange, students submit their videos via a learning management system (LMS). The teacher posts them, for example, using a blog function on the LMS. The students are informed that they will use the LMS for viewing the videos, posting comments, and assigning scores using the rubric.

Sample 2: Reading Quizzes – The students make reading quizzes based on simplified news articles on Breaking News English (www.breakingnewsenglish.com). Teachers can guide the students to post their reading quizzes online using Google Forms, with each consisting of the article text, the URL to the article, and six multiple-choice questions. In Martin (2023), students prepared at least one question for each of the following four types: detail questions, vocabulary questions, inference questions, and *not* questions (e.g., "Which of the following was not a reason for…?"). For each question, students set correct answers and make three distractor options. Each student chooses the reading level of their article on the website, but keeps in mind the level needed to yield enough information to make questions.

In preparing for the exchange, students likely need class time over about five weeks to prepare their reading quizzes. Over two weeks, students draft their quizzes on paper and then make them online using a template Google Form. Students enable the teacher to edit their quizzes. The teacher organizes them and provides additional feedback. Over three more weeks, the students continue to revise their questions for grammar and rigor.

Peer-exchange Stage

Common – The teacher makes design choices about how many presentations, essays, quizzes, etc. are exchanged depending on the nature and workload of the project. See Tip #2 about peer-exchange allocation and Tip #5 about anonymity. Moreover, teacher-feedback and teacher-provided language examples can also be included anonymously. Even if anonymized, the teacher tracks the actions taken by the students during the exchange to ensure quality. Optionally, students can assess their own pre-recorded performances or take their own quizzes for self-reflection.

Sample 1: Oral Presentations – At this stage, students engage with the performances of peers (productive language focus). Students watch, rate, and write feedback for each peer-made video. For the 22-student class in Oshima (2022), watching two-minute videos was not an overly heavy workload, so each student completed the task for all 21 videos. The teacher collected the scoring sheets to ensure anonymity. Students wrote feedback comments openly on the blog to help ensure self accountability. Students wrote at least two sentences in English for each video (or more for higher-proficiency students), but not single word comments.

Sample 2: Reading Quizzes – On exchange day, in the sixth week, each student receives a list of links for eight peer-made quizzes. Students have 80 minutes of class time to complete those quizzes, and additional quizzes can be shared for students who finish early.

Patterns-in-data Stage

Common – Students receive an organized set of data: the teacher organizes ratings and feedback information, and quiz responses are collected automatically by the Google Forms. Students should work with their support group members and with the teacher.

Sample 1: Oral Presentations – The teacher organizes scoring data and peer-feedback and reports them to each student and, using this information, students seek to find patterns and contemplate reasons for them. For repeating presentation tasks, students obtain actionable information about how to improve for the next round. Optionally, the students can also take a self-efficacy questionnaire (as in Oshima, 2022).

Sample 2: Reading Quizzes – At this stage, students engage with the performances of peers (receptive language focus). Each student inspects the collected quiz responses and they tally and compare the information to find patterns in the data and contemplate

their reasons. Tasks include thinking about why some questions were easier or harder to answer correctly and hypothesizing why some incorrect answers might be common.

Reflection Stage

Common to both Sample 1 and Sample 2 – Students write their reflections about the project via a Google Form with questions such as, "What was your impression about this presentation (or quiz-making) exchange project?", "What interesting information did you find in the feedback and performances of peers?", "Overall, what do you think you have learned?", and "Was anything a surprise to you?" Later, students can share their reflections with classmates.

Summary of Benefits

We argue that the overarching benefit of peer-led evaluation to L2 learners is that they are given agency within the language development process. This allows students to "drive the car" for part of a language course instead of only being the recipients of instruction, scores, and comments from language teachers. This fundamental difference is one way, though not the only, to facilitate the following benefits for L2 learners.

Motivation

Encouraging students to create and exchange within a supportive framework helps to satisfy learners' inherent needs for learner autonomy, reward for competency, and interpersonal relatedness (Deci & Ryan, 1985). L2 learners respond to these three components to varying degrees. In the Japanese EFL context, Agawa and Takeuchi (2016) reported that learner motivation more strongly depends on a sense of relatedness and collective responsibility than the need for autonomy, while Agawa (2020) reported that some already motivated learners prioritize competence or autonomy needs most highly. Overall, Japanese college students' autonomy needs to be supported. As Tanaka and Hiromori (2007) emphasized, student satisfaction in peer group work comes from a sense of belonging, especially for students who might initially lack motivation. In sum, although some students might thrive more independently, most peers rely more heavily on group support as they progress through each stage.

Self-efficacy

Another benefit of peer-led evaluation is L2 learners' building their self-efficacy, defined generally as individuals' judgments of their capabilities in specific activities (Bandura, 1997). Research suggests that self-efficacy significantly influences language learning success by raising motivation and increasing effort in subsequent learning (Ma, 2022; Ruegg, 2018). Bandura (1997) proposed four sources of self-efficacy: enactive mastery experiences, vicarious experiences, verbal persuasion or social persuasion, and the physiological and affective state of an individual. *Enactive mastery experiences* are considered to be most effective because learners experience success personally. *Vicarious experiences* help learners believe that they can succeed by observing others similar to themselves. *Verbal persuasion* comes from feedback and comments by people they trust, providing that "the positive appraisal is within realistic bounds" (Bandura, 1997, p. 101). Lastly, one's *physiological and affective state* is important because stress and negative emotions reduce self-efficacy regardless of other factors. Within the supportive procedure, the teacher sets a reasonable standard to achieve, gives verbal persuasion, and monitors the exchange of feedback and other information to prevent unneeded criticism. Moreover, mastery experiences for the self and vicarious experiences of others increase over the stages and all students complete the project as a group.

Metacognitive Skills

L2 learners can build metacognitive awareness of their language performance and other higher-order thinking skills through experiencing the three features of peer-led evaluation: students engaging with a standard of evaluation, L2 performances of peers, and feedback from peers. These three features occur differently depending on if coursework is focused on productive language or receptive language (see Figures 1a and 1b above), but each is influential in promoting students' higher-order thinking skills. Bloom's taxonomy (Bloom, 1956), later revised (Anderson & Krathwohl, 2001), is a widely used taxonomy for these actions of thought. They include *remembering*, *understanding*, *applying*, *analyzing*, *evaluating*, and *creating*, and the objects of these actions are forms of knowledge categorized as *factual*, *conceptual*, *procedural*, and *metacognitive*. A peer-led evaluation project reveals to L2 learners, even if implicitly, the benefits of engaging in this range of thinking.

Topping (1998) similarly described these cognitively demanding activities when learning by assessing, including "reviewing, summarizing, clarifying, giving feedback, diagnosing misconceived knowledge, identifying missing knowledge, and considering deviations from the ideal" (p. 254). Importantly, it is likely that not all peer-judgements

will be reliable, nor will students perfectly understand all aspects of rubrics, feedback, or quizzes (e.g., Farrokhi et al., 2012; Holster et al., 2013). However, as students think about their thinking within the theme of the project, there is ample opportunity for each peer to encounter a combination of information and increase their familiarity with it as they develop their metacognitive skills. As Carpenter (2012) noted, these are transferable skills that are usable in future contexts.

List of Practical Tips

This section starts with two initial pieces of advice. First, put students in support groups at the Awareness-raising Stage. Having three classmates in each support group is ideal for positive peer-to-peer dynamics (Liu & Hansen, 2002). Second, have them practice mini peer-exchange activities to get acquainted with the procedures coming in later stages. The following five tips are provided in detail.

Tip #1: Simplify the Sharing of Information

Streamlining the sharing of information helps to ensure a successful procedure. Give students a single place to reference information and submit work. Students might be familiar with aspects of the procedure (e.g., using Google Forms), but an organized central point of reference is essential. An example is to have a dedicated Google Document shared with students via the course LMS. Students learn to use the appropriate Google account to gain access and see information provided by the teacher, such as instructions, explanations, weblinks, and due dates.

Tip #2: Group Membership and Peer-exchange Allocation According to a Plan

It is important to assign support group membership and exchange allocation class-wide among the students in a systematic way in order to avoid mistakes or other confusion during the procedure. Figure 2 illustrates an exchange plan (in the form of a grid) that can help teachers with this otherwise complicated planning step. Teachers adjust their plans to accommodate different class sizes. The two axes represent the two roles taken by each student. First, the horizontal axis represents the students (i.e., *s1*, *s2*, and so on) as peer raters, givers of feedback, quiz takers, etc. Second, the vertical axis represents peers' work (i.e., *w1*, *w2*, and so on), such as essays to evaluate or quizzes to take, etc. The order of the students and their work must match exactly (i.e., *s1* matches with *w1*, *s2* matches with *w2*, and so on).

Figure 2

A Sample Demonstration of an Exchange Plan

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15	s16	s17	s18	s19	s20	s21	s22	s23	T
w1					○	○					○			○										☆
w2						○	○					○			○									☆
w3							○	○					○			○								☆
w4								○	○					○			○							☆
w5									○	○					○			○						☆
w6										○	○					○			○					☆
w7											○	○					○			○				☆
w8												○	○					○			○			☆
w9													○	○					○			○		☆
w10														○	○					○			○	☆
w11	○														○	○					○			☆
w12		○														○	○					○		☆
w13			○														○	○					○	☆
w14	○			○														○	○					☆
w15		○			○														○	○				☆
w16			○			○														○	○			☆
w17				○			○														○	○		☆
w18					○			○														○	○	☆
w19	○					○			○														○	☆
w20	○	○					○			○														☆
w21		○	○					○			○													☆
w22			○	○					○			○												☆
w23				○	○					○			○											☆
Ex	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆

An exchange plan is useful for managing the workload of the intact class of students. For either productive-language exchanges (e.g., L2 essays) or receptive-language exchanges (e.g., L2 reading quizzes), students often cannot be asked to exchange with the work of all of their peers, as the workload would be overbearing. Instead, a workload of evaluating, for example, five essays could be sufficient. Likewise for a quiz exchange project, taking eight peer-made quizzes could be sufficient. Workloads will vary by task, class size, and learners' proficiency level. The teacher will save much trouble by setting allocations according to a plan.

In the example illustrated in Figure 2, the class size is 23 students and there are seven support groups of three or four classmates each (represented by gray blocks connected

diagonally from corner to corner). In the case of 21 students, there would be seven support groups of three classmates each and the gray blocks would be the same size. The teacher can set the order of the students along the horizontal axis when setting the support groups (e.g., to ensure a balance of traits in each group). There is no utility in the students knowing their number on the grid. If the teacher chooses to make the peer-exchange open (not anonymous), the students' names can be left on their submissions.

Looking at Figure 2 in detail, the exchange plan facilitates rater allocation that is broad and evenly distributed. Teachers can allocate the exchange by placing diagonally aligned circles on the grid that do not overlap the support groups shown in gray. In the case shown in Figure 2, each student is tasked with evaluating four essays from their peers. *Student 7* receives *Work 2*, *Work 3*, *Work 17*, and *Work 20*, which were written by *Students 2*, *3*, *17* and *20*. *Student 7* does not receive essays from support group members (*Students 8* and *9*). Selecting the position of the diagonal lines of circles makes the judging sufficiently randomized for this purpose. Figure 2 also illustrates the option of anonymously adding the teacher as a rater (shown as *T*) and one common example essay to be evaluated (shown as *Ex*), and these exchanges are represented as stars. In such a case, each student receives five essays to evaluate (four circles and one star) and each student's essay is evaluated by four classmates and the teacher.

Tip #3: Make Rubrics as Easy to Use as Possible

First, prepare the text of a rubric using as familiar L2 vocabulary as possible and provide this text bilingually when needed (Nemoto & Beglar, 2014). Considerations vary for student groups that are not homogeneous and teachers address this issue accordingly. Second, alleviate the cognitive challenge of the task. Nemoto and Beglar (2014) advised using scales of six points for rubric categories to reduce cognitive strain on users and help ensure the reliability of the scoring data. The even number of six points also prevents neutral scoring. The better students can use the rubric, the better they can attend to the learning objectives represented by its categories. Even though students are not expected to evaluate peers as reliably as teachers, the main purpose of its use is for students to enhance their understanding and learn from the act of evaluation.

Tip #4: Vary Question Type and Use the Multiple-Choice Format on Quizzes

Automatic scoring of multiple-choice questions will make comprehension quiz exchanges manageable. Importantly, this format requires them to provide one correct answer and three that are plausible but incorrect. Question-making challenges students to engage their higher-order thinking skills. Teachers should try different question types themselves to identify with what the students experience at this quiz-making stage. Teachers should also provide support materials to enhance the students' range of questioning. The rigor of question-making applies to Bloom's taxonomy. This hierarchy of inquiry is clearly illustrated in an easily searchable online document called *A Model of Learning Objectives* (Iowa State University, n.d.). Additionally, Brown and Abeywickrama (2019) detail a system for identifying types of questions that is not organized hierarchically like Bloom's taxonomy. Examples include questions for main idea or topic, inference (implied detail), pronoun antecedents or other referents, vocabulary in context (guessing meaning), and identifying facts that are not stated. Prepare simple resources for students about how to systematically generate questions for their reading or listening content.

Tip #5: Consider Anonymity in Quiz Making and L2 Writing Exchanges

Social considerations can complicate students' perceptions of peer-feedback. Arguably, this is also true of quiz exchanges. Xu and Liu (2010) found that anonymity raises the worth of feedback. They anonymously shared peer and teacher feedback together and found these subsets of feedback to be valued by students at similar rates. Anonymity can potentially enhance the quality of peer exchange, as found by Rotsaert et al. (2018), who concluded that familiarity with writers' identities can hinder genuine responses. Taking the opposite view, it is also feasible that open exchange can apply positive pressure on students to take on evaluation tasks forthrightly. Consider these insights about how to handle peer exchange.

Conclusion

This practice-oriented paper introduces the concept of peer-led evaluation, and we hope it serves as a resource for teachers aiming to enhance L2 learners' agency and language development. The approach involves students taking on the roles of peer evaluator and creator of evaluation materials as they align with course objectives. Uniquely, peer-led evaluation is applicable to both productive and receptive language-focused coursework. The paper outlines a common progression of stages, including awareness-raising, preparation, peer-exchange, patterns-in-data, and reflection. Each stage brings different experiences with language, which promotes deeper awareness and learning from different perspectives. Two sample projects, one focusing on oral presentations and the other on reading quizzes, illustrate the procedural differences

based on the nature of experiencing productive or receptive L2 performances. Potential benefits for L2 learners include developing motivation, self-efficacy, and metacognitive skills. The list of practical tips addresses issues such as information sharing, systematic exchange allocation, rubric simplicity, anonymity considerations, and question variety in quizzes. Guided by theory and empirical research, teachers can experiment as they implement peer-led evaluation and develop valuable coursework for their L2 learners.

Acknowledgments

We would like to thank the Testing and Evaluation (TEVAL) special interest group (SIG) of JALT for featuring our work at their forum session at JALT2023. The feedback from the online and in-person audience and later from members of the TEVAL community were insightful. We are also thankful to the reviewers for their helpful suggestions.

Bio Data

Jeffrey Martin has taught EFL learners in a variety of settings within secondary and tertiary education. He has an M.S.Ed. from Temple University, Japan Campus. Some of his research interests include materials design, motivation, language assessment, vocabulary learning, and the study of L2 listening. <jeffmjp@gmail.com>

Sachi Oshima is a lecturer at the Faculty of Law at Chuo Gakuin University, Chiba. She is also a research fellow at the Research Institute for Policy Studies, Tsuda University, Tokyo. After working for the Japan Foundation as a chief officer, she has been an English teacher at several universities in the Kanto area. Her research interests are related to TESOL and intercultural exchange projects. <oshima.s@mc.cgu.ac.jp>

References

Agawa, T. (2020). A qualitative examination of relatedness needs in Japanese EFL classrooms and task motivation. *Journal for the Psychology of Language Learning*, *2*(1), 6–25. http://www.jpll.org/index.php/journal/article/view/agawa

Agawa, T., & Takeuchi, O. (2016). A new questionnaire to assess Japanese EFL learners' motivation: Development and validation. *Annual Review of English Language Education in Japan*, *27*, 1–16. https://doi.org/10.20581/arele.27.0_1

Allen, D., & Tahara, T. (2021). A review of washback research in Japan. *JLTA Journal*, *24*(0), 3–22. https://doi.org/10.20622/jltajournal.24.0_3

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals, by a committee of college and university examiners. Handbook 1: Cognitive domain*. Longman.

Boud, D., Cohen, R., & Sampson, J. (1999). Peer Learning and Assessment. *Assessment & Evaluation in Higher Education*, *24*(4), 413–426. https://doi.org/10.1080/0260293990240405

Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (Third edition). Pearson Education.

Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 2–27). Longman.

Carpenter, S. K. (2012). Testing Enhances the Transfer of Learning. *Current Directions in Psychological Science*, *21*(5), 279–283. https://doi.org/10.1177/0963721412452728

Cheng, G., & Chau, J. (2009). Digital video for fostering self-reflection in an ePortfolio environment. *Learning, Media and Technology*, *34*(4), 337–350. https://doi.org/10.1080/17439880903338614

Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*, *48*(4), 436–470. https://doi.org/10.1017/S0261444815000233

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press.

Falchikov, N. (2001). *Learning together: Peer tutoring in higher education*. Routledge.

Farrokhi, F., Esfandiari, R., & Schaefer, E. (2012). A many-facet Rasch measurement of differential rater severity/leniency in three types of assessment. *JALT Journal*, *34*(1), 79. https://doi.org/10.37546/JALTJJ34.1-3

Graesser, A. C., Baggett, W., & Williams, K. (1996). Question-driven Explanatory Reasoning. *Applied Cognitive Psychology*, *10*, 17–31.

Harding, L., & Kremmel, B. (2021). SLA researcher assessment literacy. In P. M. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing*. Routledge, Taylor & Francis Group.

Holster, T. A., Pellowe, W. R., Lake, J., & Hahn, A. (2013). Learning by assessing in an EFL writing class. In Q. Zhang & H. Yang (Eds.), *Pacific Rim Objective Measurement Symposium (PROMS) 2012 Conference Proceeding* (pp. 93–108). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37592-7_7

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

Iowa State University. (n.d.). *A Model of Learning Objectives*. Center for Excellence in Learning and Teaching.

Liu, J., & Hansen, J. G. (2002). *Peer response in second language writing classrooms*. University of Michigan Press.

Liu, N. F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, *11*(3), 279–290. https://doi.org/10.1080/13562510600680582

Ma, Y. (2022). The triarchy of L2 learners' emotion, cognition, and language performance: Anxiety, self-efficacy, and speaking skill in lights [*sic*] of the emerging theories in SLA. *Frontiers in Psychology*, *13*, 1002492. https://doi.org/10.3389/fpsyg.2022.1002492

Martin, J. (2020). An investigation into the use of Rasch analysis to aid L2 writers in anonymized peer-assisted learning. *Shiken: JALT Testing and Evaluation SIG Newsletter*, *24*(2), 39–53. https://doi.org/10.37546/JALTSIG.TEVAL24.2-3

Martin, J. (2023). Evaluating the higher-order thinking skills used within a student-made reading quiz exchange. *OnCUE Journal Special Issue*, *4*, 45–61. https://jaltcue.org/ocjsi/vol4

Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, *45*, 35–44. https://doi.org/10.1023/A:1006964925094

Nemoto, T., & Beglar, D. (2014). Developing Likert-scale questionnaires. In N. Sonda & A. Krause (Eds.), *JALT2013 conference proceedings* (pp. 1–8). JALT. Available online: http://jalt-publications.org/files/pdf-article/jalt2013_001.pdf (accessed on 18 November 2020).

Oshima, S. (2020). 洋楽を用いたpresentation活動の導入—「聴く」・「唄う」・「穴埋めする」をこえた、洋楽の新たな活用法 [Possible use of English songs: Using presentation activities]. *The IRLT Journal*, *19*, 93–100. https://cir.nii.ac.jp/crid/1522825130856262400

Oshima, S. (2022, November 12). *Effectiveness of Learners' Self-Assessment by Using Video Presentations* [Conference session]. JALT 2022, Fukuoka, Japan. https://researchmap.jp/sachi_oshima/presentations/40470442

Piaget, J. (1971). *Biology and knowledge: An essay on the relations between organic regulations and cognitive processes*. University of Chicago Press.

Rodríguez-González, E., & Castañeda, M. E. (2018). The effects and perceptions of trained peer feedback in L2 speaking: Impact on revision and speaking quality. *Innovation in Language Learning and Teaching*, *12*(2), 120–136. https://doi.org/10.1080/17501229.2015.1108978

Rotsaert, T., Panadero, E., & Schellens, T. (2018). Anonymity as an instructional scaffold in peer assessment: Its effects on peer feedback quality and evolution in students' perceptions about peer assessment skills. *European Journal of Psychology of Education*, *33*(1), 75–99. https://doi.org/10.1007/s10212-017-0339-8

Ruegg, R. (2018). The effect of peer and teacher feedback on changes in EFL students' writing self-efficacy. *The Language Learning Journal*, *46*(2), 87–102. https://doi.org/10.1080/09571736.2014.958190

Tanaka, H., & Hiromori, T. (2007). 英語学習者の内発的動機づけを高める教育実践的介入とその効果の検証 [The effects of educational intervention that enhances intrinsic motivation of L2 students]. *JALT Journal*, *29*(1), 59–80. https://doi.org/10.37546/JALTJJ29.1-3

Topping, K., & Ehly, S. (1998). *Peer-assisted learning*. Routledge.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249–276. https://doi.org/10.3102/00346543068003249

Xu, Y., & Liu, J. (2010). An investigation into anonymous peer feedback. *Foreign Language Teaching and Practice*, *3*, 44–49.

Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching*, *49*(4), 461–493. https://doi.org/10.1017/S0261444816000161

Zhang, X., Ardasheva, Y., & Austin, B. W. (2020). Self-efficacy and English public speaking performance: A mixed method approach. *English for Specific Purposes*, *59*, 1–16. https://doi.org/10.1016/j.esp.2020.02.001