

Quality Assurance, Program Evaluation, and Curriculum Revision: A Case Study from Rikkyo University

Davey Young

ABSTRACT

Quality assurance (QA) in education is an important process for maintaining a high standard of lesson delivery and student learning outcomes. QA is closely tied to program evaluation and development, which has been a central aspect of the administration of English Discussion Class (EDC) at Rikkyo University in Tokyo, Japan since the course was piloted in 2009. Using Lynch's (1996) connoisseurship model of curriculum evaluation, this paper examines how a curriculum revision taking effect in the 2020 academic year will impact the viability of preexisting QA and program evaluation instruments in EDC. The analysis reveals that nearly all such instruments will be unviable. Furthermore, there will be few available resources for investigating the impact of the curriculum revision on student learning from 2020 onward. Suggestions for limiting potential adverse effects on overall program quality and maintaining some degree of quality assurance conclude the paper.

INTRODUCTION

Current Context

Quality assurance (QA) is an important aspect of educational administration across contexts, and is a normal feature of institutional accreditation. At the heart of QA is the need to ensure that practitioners are meeting predetermined standards in providing educational services to students. These standards can be dictated by the curriculum, the institution, policy, or any combination therein. In Japan, the Higher Education Bureau of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) has traditionally been charged with overseeing QA of institutes of higher education. MEXT's Higher Education Bureau uses a quality assurance framework that directs the application of the Standards for Establishing Universities, which was first drafted in 1956. These standards have not been revised since that time, and have instead been bent to fit a changing educational landscape through ongoing bureaucratization. As of 2004, these standards have been enforced through the Establishment-Approval System and the Quality Assurance and Accreditation System. The latter system was created in 2004 to address MEXT's concerns that universities may not have been prepared to assure their own quality autonomously, as MEXT was not performing any QA checks of their own (MEXT, 2009).

Currently, MEXT performs an accreditation review using a simplified version of the 1956 Standards for Establishing Universities once every seven years. A number of concerns understandably remain. Universities are still left to independently warrant the quality of their educational programming, and the stipulations for how this programming should be designed and delivered is scarcely present. There is no explicit mention of English language education in the Standards for Establishing Universities, for instance, and the only article pertaining to classroom practice is remarkably broad:

第二十三条（各授業科目の授業期間） Article 25 (Methods for Conducting Classes)
1 授業は、講義、演習、実験、実習若しくは実技のいずれかにより又はこれらの併用により行うものとする。(1) Classes shall be conducted using any one of the methods of lectures, exercises, experiments, practical training, or skills practice or a combination thereof (MEXT, 2009, p. 19).

As language teachers are typically accustomed to taking clearly defined approaches based on theories of language acquisition, principles of teaching, and current research, these guidelines cannot reasonably be applied to measure the quality of English language education in postsecondary contexts in Japan. English language programs are therefore left to conduct quality assurance on their own with virtually no guidance or accountability from a policy perspective. The general lack of such measures in Japanese postsecondary English language programs is well noted, however, and has been connected to poor course and faculty organization (Honisz-Green, 2013) and insufficient faculty development schemes (Iijime, Takahashi, Watanabe, & Watari, 2020).

Since 2010, all first-year students at Rikkyo University in Tokyo, Japan have been required to take four English language courses, a cornerstone of which has been English Discussion Class (EDC). With between 4,500 and 4,700 students matriculating each year and EDC class sizes kept between seven and nine students, the Center for English Discussion Class has operated separately from the English Language Program, which oversees the delivery of the three other mandatory English classes and a number of elective courses. As such, EDC has a separate teaching faculty and management structure headed by a Director and Deputy Director, two positions that are specially appointed from other university faculty by the university's president's office each year. In practice, a team of fulltime administrative staff and four program managers oversee the day-to-day delivery of the course. Program managers also teach four EDC classes each semester, a course load of roughly one third the fulltime instructors, though they do not teach any other courses. This faculty organization has allowed program managers to focus on managing and developing the academic aspects of EDC, including syllabus and materials design, assessment, and faculty development. Over the years, program managers have endeavored to engage instructors with respect to both course and faculty development to improve the quality of learning for EDC students and to establish and maintain a *quality culture* (Rose, 2007), one in which all participants are invested in overall program quality, within the Center for English Discussion Class.

Program Evaluation

Foregrounding all program managers' responsibilities has been the execution of regular and ongoing program evaluation projects, a duty made easier by the strongly unified nature of the EDC syllabus. Evaluation is vital to the successful administration and delivery of any language program. *Product-focused evaluation* positions the curriculum as a product to ensure both *program accountability* and *program development*. In product-focused evaluation, accountability "refers to the extent to which those involved in a program are answerable for the quality of their work" while development denotes "the quality of the program as it is being developed" (Richards, 2017, p. 282). In contrast with product-focused evaluation is *process-focused evaluation*, which can be either *descriptive* or *reflective*. The descriptive mode of process-focused evaluation is concerned with collecting information to gain insight about how a program is implemented, while the reflective mode investigates the teachers' pedagogy that undergirds that implementation (Richards, 2017). Product- and process-focused evaluation are not mutually exclusive. Effective and efficient evaluation mechanisms will contain some degree of both. Such hybridized evaluation is a key component of QA and helps determine to what extent a program serves its students and achieves its stated aims. Furthermore, program evaluation should be viewed as ongoing and cyclical; such a view helps establish an *evaluation culture* that can maintain standards and stay responsive to evolving student needs (Davis, Sinicrope, & Watanabe, 2009).

Previous program evaluation projects within EDC reflect the creation of such a culture. Over the years, these projects have focused on: instructors' views on EDC class size (Moroi, 2014), non-teaching tasks (Livingston & Moroi, 2015), and working within a unified syllabus (Brereton,

2019); faculty development (Brereton & Young, 2018; Lesley 2017, 2018; Schaefer & Lesley, 2019); students' willingness to communicate (Doe, 2014) and perceptions of native and non-native teachers (Kamada, 2014); assessment instruments (Doe, 2012; Schaefer & Young, 2019; Young & Nicklin, 2019); textbook development (Young, 2016, 2017, in press); syllabus design (Schaefer, 2018; Lesley & West, 2019); and support for students with disabilities (Young & Schaefer, 2019; Young, Schaefer, & Lesley, 2019).

In addition to the program evaluation projects just listed, an optional student survey administered at the end of every semester has been an essential tool for evaluating the course. This survey consists of 13 Likert-scale items asking students about discrete aspects of and their experience within the course, as well as provides a space for open comments. Brereton, Schaefer, Bordilovskaya, and Reid (2019) conducted a detailed analysis of the spring 2018 student-survey data and found that 2,398 of the 4,538 students that semester had a positive view of the course, with over 93% of respondents agreeing or strongly agreeing with all 13 items on the survey. These results are not inconsistent with those of previous years, nor with those reported in the 2019 academic year (Appendix A). A qualitative analysis of the responses to the spring 2018 survey's final, open-ended item was consistent with the quantitative data, and allowed the authors to make concrete recommendations to further refine course design and delivery.

The ongoing and iterative execution of product- and process-focused program evaluation has been vital to quality assurance within the Center for English Discussion Class. The need for such QA has always been especially acute given the scale and strongly unified nature of the EDC syllabus. Throughout the years, the presence of systematized and transparent QA instruments has ensured both program accountability for all stake-holders, as well as rolling program development to maximize learning outcomes for all students enrolled in the course. In this way, the quality culture and evaluation culture within the Center for English Discussion Class have been inexorably intertwined.

Inevitable Change

Rikkyo University is abolishing the Center for English Discussion Class and establishing a new center, the Center for Foreign Language Education and Research, in April of 2020. EDC as a course will still exist, though condensed to one semester and with class sizes increased from around eight students to ten. This increase runs counter to Moroi's (2014) finding that a class size of nine or more students negatively impacts EDC teachers' ability to accurately monitor students' speaking performance and offer meaningful, individualized feedback. This change may appear especially startling when viewed alongside the student-survey results. In the spring 2019 survey, for instance, 97.6% of respondents agreed or strongly agreed that EDC's small class size helped improve their discussion ability, while 96.9% agreed or strongly agreed that their teacher's feedback during lessons helped them understand what skills they needed to use more. While ten students may still be a small class, it remains to be seen how this change will impact teachers' perceived efficacy or students' perception of their learning.

Beyond the changes to EDC itself, a new debate class with around 20 students per class will replace the second semester of EDC. The faculties currently employed by the Center for English Discussion Class and the English Language Program will merge, with over 60 full-time teachers, as well as between 100 and 150 part-time teachers. All full-time teachers have been provided with new contracts, and their new duties and schedules severely limit the feasibility of faculty development systems as outlined by Lesley (2017, 2018) and Schaefer and Lesley (2019).

These broader changes have had a cascading effect on other aspects of the syllabus. The limits placed on faculty development, for instance, preclude the ability to conduct norming of assessment instruments, and so assessment within EDC has had to be entirely reimaged. This

change comes in spite of the high degree of interrater reliability in EDC's discussion tests (Doe, 2012), as well as 92.2% of the spring 2019 student survey respondents' agreement or strong agreement that the discussion tests were good ways to assess their discussion ability.

Program evaluation in the new center will be further complicated by the fact that no needs analysis was ever performed ahead of the 2020 curriculum revision. The present paper aims to catalogue current QA measures in EDC, anticipate potential negative impacts on student learning owing to the loss of these measures, and suggest policies or procedures to dampen these impacts for the new leadership and teachers within the Center for Foreign Language Education and Research to adopt as they see fit.

METHOD

This paper uses the *connoisseurship model* (Lynch, 1996) of curriculum evaluation. Application of this model begins by establishing *critical guideposts*, a set of shared values and concepts that form the foundation of the evaluation. The critical guideposts to analyze QA instruments in the present context were modified from Mercado (2012) and Nation and Macalister (2010). These are divided into two principle domains. The first domain, quality of instruction, consists of six critical guideposts: teaching staff, teaching methodology, quality of syllabus design, instructional materials, assessment, and support systems for students. The second domain, quality of learning, has two critical guideposts: perceived learning and actual learning. Previous research conducted by program managers on each of these critical guideposts is used to comment on their observed efficacy as QA measures in the current context, as well as to form the basis for predicting potential negative impacts on student learning and suggesting policies or procedures to offset them from 2020 onward.

The eight critical guideposts and their corresponding QA instruments present both before and after the 2020 curriculum revision are summarized in Table 1 and detailed with respect to EDC's syllabus and course administration in the sections that follow.

Table 1. Critical guideposts for curriculum evaluation and quality assurance instruments in Rikkyo University's English Discussion Class before and after the 2020 curriculum revision.

Critical Guideposts for Curriculum Evaluation	QA Instrument in the Center for English Discussion Class (2010-2019)	QA Instrument in the Center for Foreign Language Education and Research (2020)
<u>Quality of Instruction</u>		
1. Teaching staff	- Orientation for new instructors - Regular faculty development - Observations for all teachers through first five semesters - Instructor Handbook - Student survey results	- Instructor Handbook
2. Teaching methodology	- Observations for all teachers through first five semesters - Regular faculty development - Instructor Handbook and Textbook Guide - Student survey results	- Instructor Handbook and Textbook Guide
3. Quality of syllabus design	- Analysis of the syllabus - Evaluation of discrete course components - Student survey results	- TBD
4. Instructional materials	- Mandatory teacher survey on course-specific textbooks - Evaluation checklist and principles for textbook revision - Student survey results	- Optional teacher survey on course-specific textbooks
5. Assessment	- Regular rater-norming - Double-reporting of grades to minimize human error - Student survey results	- TBD
6. Support systems for students	- Student Handbook - Office hours for students to meet with teachers - Framework for supporting students with disabilities - Student survey results	- Office hours for students to meet with teachers - Framework for supporting students with disabilities
<u>Quality of Learning</u>		
1. Perceived learning	- Student survey results - Lesson observations	- TBD
2. Actual learning	- Specific performance using individuated criteria on rater-normed assessment instruments	- General performance using non-rater-normed assessment instruments

RESULTS

Quality of Instruction

Teaching Staff and Teaching Methodology

The first two critical guideposts for quality of instruction are teaching staff and teaching methodology. Due to the interconnectedness of the QA measures used to evaluate these guideposts, they will be discussed here in tandem.

EDC's communicative approach maximizes real-time, meaningful interaction through small class sizes and prioritizes student talk-time and repetition of content and target language, as a high amount of speech production in a controlled environment is necessary for developing communicative competence (Ellis, 2002) and sustaining use in a normal context (Nation, 2001). Measures to safeguard quality of instruction that conforms to this methodology begin in EDC as soon as new instructors are hired.

All newly hired EDC instructors have historically undergone four days of orientation onboarding. Sessions within these orientations can be divided into three types: overviews of course aims, methodology, and administration; teaching workshops; and rater-norming of assessment rubrics for regular and test lessons. At the end of each orientation, new instructors completed a questionnaire designed so that program managers could gain feedback and refine the orientation experience and outcomes for future cohorts of new instructors accordingly.

In their first semester teaching the course, EDC instructors have traditionally participated in a progression of teaching workshops to improve their ability to teach discrete lesson stages in accordance with the prescribed methodology. In their second semester, instructors participated in another series of workshops with respect to the same lesson stages, only the focus shifted from teaching workshops to discussion aimed at achieving a shared understanding of the language-teaching methodology, principles, or research on which these lesson stages are based. For example, while instructors might workshop different ways of conducting target-language practice in the first semester, in the second semester they read and discussed an article on principles of creative automatization (Gatbonton & Segalowitz, 1988). This faculty development scheme was intended to help teachers achieve a shared understanding of the course's methodology, as well as develop their ability, confidence, and comfort applying that methodology. Such faculty development also provided a degree of QA, as it follows that by improving teachers' ability to teach, students are better able to meet lesson and course aims to a unified standard.

EDC's communicative teaching methodology has been reinforced over the years through a sequence of principled classroom observations conducted by video recording lessons for later viewing and reflection. Video recording lessons for critical reflection is a well-established practice that has been shown to have a variety of positive impacts on teacher performance (Farrell, 2018; Mann & Walsh, 2017). The progression from orientation and faculty development sessions to classroom observations to reinforce a common methodology follows Mercado's (2012) guidelines for teacher training for quality assurance. Indeed, classroom observations are also "the cornerstone of quality assurance" (O'Leary, 2014, p. 11). Without firsthand knowledge of teachers' abilities and performance in the classroom, language program administrators have nothing more than self-reports, hearsay, and preexisting assumptions on which to base their evaluation of the teaching staff (Bailey, 2006). Supervisors cannot fairly, accurately, or meaningfully evaluate teachers on such a basis.

Observations can take a number of forms and serve multiple purposes. Two basic modes of teacher evaluation are *summative evaluation* and *formative evaluation* (Bailey, 2006). Summative evaluation seeks to simply evaluate a teacher's ability to perform their duties to a predetermined standard. Formative evaluation goes one step further, and is conducted with the

aim of providing feedback to the teacher for their own development and improvement. The aims of such improvement can be general or specific.

A blend of summative and formative observations has been used within the Center for English Discussion Class, though the organization and procedure of these observations evolved over the years based on instructor feedback. Before the 2020 curriculum revision was announced, EDC instructors completed three observations in their first year teaching the course: two in their first semester and one in the second. For each of these observations, a video camera was set up to record the lesson with students' consent, and instructors completed a reflective observation form designed to both provide a summative description of and a formative reflection on lesson delivery. Program managers would then observe the lesson and respond to the instructors' comments and questions on the observation form. Finally, the observing program manager would conference with the teacher for around 45 minutes to discuss both general and discrete aspects of the lesson that could be improved. These would then be used to articulate concrete goals for the teacher to meet in the subsequent observation.

In their second year and first half of their third year teaching the course, instructors would participate in peer observations using a more freely formatted reflective observation form. The purpose of these observations was primarily formative, though program managers would also view the video to conduct a summative check and follow up with individual instructors on an as-needed basis. From the second half of their third year teaching the course onwards, observations were no longer mandated, though all instructors were required to complete a self-directed reflective professional development task. The list of recommended tasks, which included self- and peer-observations as options, was borrowed from Farrell (2016). Each semester, instructors were given a deadline to report on their selected task type to a program manager, and were otherwise trusted to complete the task to their own standard as teachers of the course. This progression from blended summative and formative observations to autonomous professional development was positively perceived by instructors (Brereton & Young, 2018), demonstrating that QA measures can be embedded within a meaningful professional development scheme.

The rationale and procedure for progressing through this professional development scheme, as well as rationales for the wider EDC teaching methodology, have traditionally been detailed in an Instructor Handbook revised and distributed to instructors at the beginning of each academic year. Teachers can turn to this resource if they have questions related to any discrete aspect of their responsibilities or performance requirements for both teaching and non-teaching tasks. This document provides a valuable, tangible baseline that program managers can point to if any instructor is not performing desirably, and this has obvious advantages from a QA perspective. Similarly, teachers can turn to a Textbook Guide created by program managers to accompany the course-specific textbook.

The fourth and final instrument used for QA of teaching staff and methodology is the student survey administered each semester, though these have always been viewed in total rather than by connecting specific responses with individual instructor performance. In other words, aggregated responses have been routinely analyzed to shed light on the teaching staff in total, rather than isolating responses to inform the teaching ability of any one instructor. As such, the student survey has been used to evaluate teaching methodology far more so than it has been used to evaluate the teaching staff.

From 2020 onwards, there will be no mandated faculty development scheme to train teachers on mutually agreed upon principles and practices for optimal lesson delivery, nor will there be any video observations or infrastructure for related professional development. An Instructor Handbook will be used, though it has been significantly reduced in both scope and detail as a result of other changes discussed in greater detail in other sections of this paper. A Textbook

Guide will also be available for the new textbook to be used after the curriculum revision. It is hoped that these holdovers will serve as a small measure in safeguarding EDC's core teaching methodology and performance standards for the teaching staff, though there will be no mechanisms in place to ensure that these standards are being met.

Finally, there are no plans at present to administer the end-of-semester student survey in its current form. Rather, it is likely (but not guaranteed) that a separate survey providing only a single free comment section will be provided to students. While this allows some degree of qualitative feedback from students who opt to complete the survey, it will be impossible to evaluate how discrete course elements are perceived after the 2020 curriculum revision without the 13 Likert-scale items used in the past surveys. For instance, without asking students if they agree or disagree that EDC's small class size was beneficial to their learning, course evaluators will not be able to compare the 2020 students' perception of ten-student classes with previous years' students' perceptions of eight-student classes. The loss of the quantitative aspect of the student survey will have a far-ranging impact in various aspects of quality assurance of EDC beyond just the teaching staff and methodology.

Quality of Syllabus Design

EDC was originally created in accordance with guidelines set forth by the Japanese Ministry of Education, Science, and Culture (MEXT) in a 1992 directive titled *The Course Study for Senior High School* (Hurling, 2012). This directive aimed to produce "cosmopolitan Japanese" with positive attitudes towards communicating in English (Gorsuch, 2000, as cited in Hurling, 2012, p. 1-2). MEXT's *New Course of Study 2009 for Foreign Languages: English* explicitly calls for a communicative language teaching (CLT) approach, and aims to "develop students' communicative abilities such as accurately understanding and appropriately conveying information, ideas, etc., deepening their understanding of language and culture, and fostering a positive attitude towards communication" (MEXT as cited in Underwood, 2012, p. 118). The EDC syllabus' initial creation fits Richards' (2013) description of backward design, one that begins "with a specification of learning out-puts and [uses] these as the basis for developing instructional processes and input" (p. 20).

EDC adopts a communicative competence approach (Savignon, 1983), which "attempts to hasten the attainment of fluency by focusing directly on the learning of functional utterances rather than the rules" (Gatbonton & Segalowitz, 1988, p. 478). Instruction in EDC is therefore organized within a *functional-notional* syllabus, which highlights "categories of language use rather than categories of language form" (Krahnke, 1987, p. 35). Widdowson (1989) notes that a course built around a functional-notional approach should include *lexical clusters* that can be "suitably adjusted to the communicative purpose required of the context" (p. 136), and indeed the EDC syllabus includes explicit instruction of several such clusters, or "fixed and prefabricated chunks of language that speakers use heavily in everyday interactions," (Celce-Murcia, 2007, p. 47). These are grouped into *discussion skills* and *communication skills* that target specific competencies within Celce-Murcia's (2007) model of communicative competence. Such instruction builds overall communicative competence and achieves the 26 cognitive objectives outlined by the creators of the course (Hurling, 2012). Additional affective and practical objectives for the course are derived from Bloom (1956) and Nation (2001) respectively; a more comprehensive discussion of these can be found in Hurling's (2012) introduction to the course.

In addition to the twelve discussion skills in the original EDC syllabus, three communication skills were included to help maintain the flow of discussion (Hurling, 2012). In the 2017 academic year, program managers analyzed the existing EDC syllabus and revised these skills to focus more specifically on strategic competence; these changes were incorporated into

the syllabus in the 2018 academic year (Schaefer, 2018). How each lesson's target language is taught and assessed conforms to Auerbach's (1986) description of Competency Based Language Teaching, as EDC students receive performance-oriented and modularized instruction geared towards explicit outcomes; students must also demonstrate a mastery of performance objectives related to target language use, and their assessment is continuous and ongoing. *Carrier topics* are used each week to make the use of this formulaic language meaningful, genuinely communicative, and psychologically authentic. As each lesson focuses on a specific topic module chosen for its suitability to deliver explicit instruction of a set of lexical clusters that realize a specific competency, the EDC syllabus can also be situated in the *theme-based* section of the CBI spectrum as conceptualized by Brinton and Snow (2017).

Table 2. A selection of EDC cognitive objectives and their matching lexical clusters.

Number in Hurling's (2012) List of Cognitive Objectives	Cognitive Objective Derived from Dörnyei & Thurrell (1992; 1994) and Kehe & Kehe (1994)	Discussion Skill and Example Lexical Cluster in the EDC Syllabus	Competency in Celce-Murcia's (2007) Model
1	Volunteer to join a discussion	Joining a Discussion (Can I say something?)	Interactional (conversational)
2	Mark connections between their ideas and others' ideas	Connecting Ideas (As ___ said,...)	Discourse
25	Paraphrase	Paraphrasing (Do you mean...?)	Strategic

The EDC syllabus can therefore be described as both functional-notional and theme-based, while the instructional approach that accompanies it is communicative and competency-based. It is important to emphasize "that the application of any theoretical model of communicative competence is relative rather than absolute" (Celce-Murcia, Dörnyei, & Thurrell, 1995, p. 30) and that a working model and the principles derived from them "must be compatible with the local context" (Richards, 2006, p. 15). A course such as EDC must therefore be flexible and responsive to students' shifting needs. Ongoing program evaluation by EDC's program managers has provided some ability to identify such needs and adapt to them accordingly. Program evaluation projects that directly investigated the EDC syllabus design in part or total are Brereton, Schaefer, Bordilovskaya, and Reid (2019), Doe (2014), Schaefer (2018), Lesley and West (2019), and Young (2017, in press), while other projects have provided indirect insight. Whenever revisions have been made to the syllabus as a result of an evaluation project in the past, program managers have had an array of tools to investigate the merits and demerits of the revision. These typically included surveying teachers regarding their perception of the revision, analyzing results to the student survey administered across the course at the end of each semester, and analyzing students' grades pre- and post-revision.

From 2020 onwards, there will be opportunities for course administrators to continue conducting syllabus analysis and projects evaluating discrete course components. However, it is currently unknown who will be responsible for conducting these projects. As such, it is impossible to determine at present what projects may or not may be conducted. Additionally, as the student survey administered at the end of each semester has been a valuable tool in determining how students view the syllabus, the loss of the quantitative survey items will surely be felt if and when the new center's leadership deem it worthwhile to evaluate the syllabus they will be responsible for managing.

Instructional Materials

EDC uses two course-specific textbooks: one for the spring semester and another for the fall. Since the course was required for all first-year Rikkyo students in 2010, students have been streamed into one of four proficiency levels based on a TOEIC placement test. Each EDC textbook therefore has four level-specific versions graded for readability with the same carrier topics, but with slight variations in target skills and lexical clusters. These textbooks were written by the course creators in the 2009 and 2010 academic years, and have been revised by program managers every year since. EDC instructors often create their own supplementary materials to accompany these textbooks, and these are shared with other instructors via an assortment of Google Drive folders created for this purpose.

In terms of QA and program evaluation of instructional materials, a number of mechanisms have been employed in EDC over the years. An evaluation project related to textbook development with specific regard to readability was carried out in the 2015 academic year (Young, 2016). The same year, a pilot study into student perceptions of the textbooks' 24 carrier topics was conducted, with a fuller investigation surveying over 1,550 students in the 2017 academic year (Young, in press). These findings have helped inform the selection and sequence of topics in each subsequent edition of the textbook, as well as the new textbook created for the condensed 2020 EDC course. Far more useful than these projects has been regular feedback from EDC instructors about each edition of the textbook. The importance of feedback from teachers on instructional materials is well established (Stillwell et al., 2011), and in EDC one of all instructors' regular non-teaching duties is to provide feedback on individual components of the textbook at the end of each semester. This feedback is then integrated into annual textbook revision in a principled manner using an evaluative checklist for each lesson component (Young, 2017).

The 2020 EDC textbook (Fearn-Wannan, Kita, Sturges, & Young, 2020) was created based on findings from the projects mentioned above, as well as on feedback from instructors on the most recently published editions of the earlier version of the textbooks (Brereton, Lesley, Schaefer, & Young, 2019; Fearn-Wannan, Kita, Sturges, & Young, 2019). As Brereton (2019) found that most EDC instructors value the collaborative culture resulting from EDC's strongly unified syllabus, the quality and evaluation cultures with respect to textbook development will hopefully persist through the curriculum revision, at least in the short term. As such, future course evaluators should be able to collect an ample amount of feedback from instructors on the new textbook, though this feedback will not be mandatory under the management of the new Center for Foreign Language Education and Research. As such, there is some concern that the number of instructors who provide feedback will drop once the task becomes an optional one.

What has the potential to be far more problematic, however, is that from 2020 onwards, textbook development will shift from program managers to a Discussion Committee. As the members of this committee will be different from year to year, it will be extremely difficult to create and maintain a skill and knowledge base with respect to long-term projects such as textbook development, as the cycle of gathering feedback, drafting, and revising the textbook as described in Young (2017) takes several months and requires a high degree of organization and communication across stakeholders, including the publishing company. The EDC administrative staff and program managers have spent a decade building and maintaining a relationship with the publishers and copyeditors, and the final stages of textbook production involve three rounds of revisions with the publisher before the textbook can be printed. Fortunately, two of the current program managers will serve on this committee in 2020, and so will be able to help mitigate potential problems in the short term, as well as establish procedures for future Discussion Committees to follow in order to streamline their duties and maintain some degree of quality assurance. However, the makeup of this committee beyond 2020 is as yet undetermined.

Finally, the new Center for Foreign Language Education and Research plans to shift away from using Google Drive as a forum for teachers to share supplementary instructional materials, opting instead to use Blackboard. The new system may require instructors to submit their materials to the Discussion Committee for approval before they can be posted on Blackboard. This may result in materials for a specific lesson being made available to other teachers only after the lesson has concluded. It also creates work for instructors on the committee that would be redundant if the previous means of sharing materials were allowed to continue. More importantly, it may lead to gatekeeping on the part of committee members or alienate instructors accustomed to a freer mode of collaboration. More than likely, teachers will continue to share materials outside of Blackboard as a workaround. As Rikkyo University contracts with Google to provide email addresses for the institution, all teachers have access to Drive by default. As Brereton (2019) found that EDC instructors generally value using Drive as a way to collaborate on lesson planning and materials creation, these instructors may eschew submitting materials for committee review if required to do so, and instead continue using Drive after the 2020 curriculum revision regardless.

Assessment

EDC students are assessed on an ongoing basis using two different rubrics. In regular, non-test lessons, instructors use a criterion-referenced rubric that provides a score of zero to four across five criteria: attendance, quiz score, use of discussion skills, use of communication skills, and participation. Using the regular lesson rubric, teachers assess students on the last three of these criteria as they complete a sequence of discussion-based tasks over the course of the lesson. The teacher typically monitors two groups concurrently (three groups of three in nine-student classes), and provides formative feedback between each discussion. Students' performance in these lessons accounts for 70% of their overall grade. Historically, teachers have been trained to use the regular lesson rubric as part of their orientation onboarding after being hired to teach the course. As part of ongoing QA in assessment, additional training sessions have occurred irregularly in instances of syllabus or rubric revision.

The remaining 30% of a student's grade consists of three equally weighted discussion tests administered in lessons five, nine, and 13 of each semester. The test lesson rubric is also criterion referenced, but with a finer grain of detail than the regular lesson rubric. Each test yields a score for every student out of 25 points, with five points each in the categories of communication skills, questions, content, listener-sided discussion skills, and speaker-sided discussion skills. The test is completed by one group at a time so that the teacher can focus on their performance exclusively. Teachers have historically undergone extensive rater-norming upon being hired by the Center for English Discussion Class, as well as participated in an additional norming session before each test lesson. In past years, program managers usually conducted an interrater reliability check after each norming session, allowing program managers to follow up with any misfitting raters. Doe (2012) found this method to be successful in decreasing the number of such raters in EDC over the course of an academic year.

As assessment in both regular and test lessons happens in real time, teachers are able to assign a grade to every student for each criterion as soon as the class concludes. Teachers can immediately record these grades on an assessment form distributed to all teachers by the EDC administrative staff at the beginning of each lesson week. These assessment forms are printed from the Lecturer Reporting System (LRS). The LRS is a unique database created specifically for EDC to record and report grades consistently and efficiently. Reporting the grades for a single class on the LRS typically takes between one and two minutes. After each lesson, instructors enter their students' grades into this system before submitting the assessment forms to the administrative staff. Students' completed reading quizzes, which are administered at the start of each lesson, are

also submitted every day after being recorded on the assessment form and entered into the LRS. Such double reporting allows the administrative staff to check for human errors. When such errors are found, program managers confer with the teacher to identify the nature of the error and ensure that the student receives the proper grade on the LRS, as this database also compiles all students' final grades for reporting to Rikkyo University's Academic Affairs Office at the end of each semester.

As the LRS tracks every student's performance across all lessons and scoring criteria, the quantitative data it provides has been an invaluable resource for assessment-related QA and program evaluation projects. For example, program managers were able to identify a drift towards more lenient application of the rubric from 2012 to 2016, and to then revise each rubric in stages over a two-year period to correct for this rater drift (Schaefer & Young, 2019). The project was successful, but analysis of the grading data through the spring semester of the 2019 academic year suggests that without consistent norming, such a drift towards leniency is likely to recur. (Spring semester grade distribution from the 2010 to 2019 academics years are included as Appendix B.) Another example of the functionality of the LRS with respect to quality assurance and evaluative purposes was an item analysis of the reading quizzes taken by students in the spring semester of the 2018 academic year (Young & Nicklin, 2019). This study found that the reading quizzes that have traditionally begun each EDC lesson are successful in preparing students to discuss the given lesson topic, but also identified ten problematic items that could then be revised for the spring 2019 course.

After the 2020 curriculum revision, the LRS will be abandoned in favor of Blackboard, and so QA of assessment in EDC will be drastically curtailed. As there will be no further double reporting of grades, any grading inconsistencies resulting from human error will go unnoticed. Furthermore, for each student, teachers will submit only a final grade at the end of the semester rather than submitting grades after each lesson. Therefore, the detailed quantitative data tracked by the LRS will no longer be available if there is to be any attempt at ascertaining the impact of the curriculum revision on both actual student learning and instructors' interpretations of the rubric on a lesson-by-lesson or domain-specific basis.

As the newly formed Center for Foreign Language Education and Research will not hold any mandatory norming of assessment instruments, EDC program managers have had to greatly reimagine these instruments and how assessment appears in the broader syllabus. With respect to the new syllabus, there will still be three discussion tests weighted as 30% of the students' final grades, but these tests will be conducted in the same format as regular lessons using the same impressionistic rubric that is used in regular lessons. This rubric maintains clear band descriptors, but these now assess fewer criteria across more lesson stages in an attempt to maintain a degree of reliability that does not require intensive and regular rater-norming. These revisions were also made with consideration of the increase in class size from eight students to ten from 2020 onward, as Moroi (2014) found that EDC instructors do not feel that they can reliably assess students in classes with nine or more students.

Support Systems for Students

The final critical guidepost for evaluating the quality of instruction is support systems for students, which is a key aspect of any curriculum for linking the quality of instruction with the quality of learning. There are two sides to the support systems coin: communicating expectations for performance to the students, and providing a venue for students to request support should the need arise. A curriculum provides a spectrum of support only if both elements are present.

Communicating realistic expectations for achievable standards of performance to students is the first step in getting students to meet those standards. To this end, EDC students

have received a Student Handbook at the beginning of each semester. This document provides detailed information on what students are expected to learn and be able to do by the end of the term, along with rubrics for regular and test lessons and advice on how to succeed in the classroom. All of this information is provided in both English and Japanese, and instructors are expected to outline the main points of the Student Handbook in the first lesson of the semester. Additionally, the Student Handbook has information on how to request support from either their instructor (e.g. by listing their instructor's email and office hours) or the EDC administrative staff (e.g. to request a grade change or exemption from class after contracting a communicable disease).

In addition to the support provided to all students through the distribution of a Student Handbook, students with disabilities receive additional support through the implementation of a context-specific framework for providing such students with a continuum of services. Providing systematized support for students with disabilities is a critical component of ensuring quality education for a diversity of students (Dunn, Pratt, & Van Putten, 2020). A longitudinal attendance and grade analysis revealed that EDC's framework for supporting students with disabilities had thus far been successful in helping students with disabilities meet lesson and course aims at a rate commensurate with their abled peers (Young, Schaefer, & Lesley, 2019). Furthermore, instructors have expressed satisfaction with how this framework helps them prepare to teach and provide a continuum of services for the students with disabilities in their classes (Young & Schaefer, 2019).

A third, vital QA measure for support systems for students throughout EDC's history has been the student survey administered at the end of each semester. As had been noted above, this instrument has provided a final opportunity for students to voice concerns before they exit the course via an open comments section. With respect to support systems, this comment section provides a space for students to request support or lodge a complaint if they are not getting the support they need.

The new Center for Foreign Language Education and Research at Rikkyo University will not provide a course-specific student handbook after the 2020 curriculum revision. This means that performance expectations and standards for student learning will not be officially communicated to students as a matter of course, and instructors will not have an official document to refer students to should the need arise. Students will still be able to ask instructors for support or to meet during office hours, though this pales in comparison to the loss of support systems that preempt the need for the student to ask for support in the first place. Fortunately, the framework for supporting students with disabilities will be modified and retained. However, it is unclear who will be responsible for overseeing discrete stages of the framework's implementation, and so it is difficult to say with certainty that EDC students with disabilities will receive a full spectrum of services in the 2020 academic year. Furthermore, the loss of the LRS will complicate any attempts to conduct an attendance and grade comparison to determine the efficacy of this framework's implementation in the new center. The open response item on the new, 2020 survey will allow space for students to lodge a complaint if they have not received support, but by the time the survey has been administered the course will have concluded, and so any such complaints could only be used to inform future practice.

Quality of Learning

Perceived Learning

The student survey administered at the end of each semester and already mentioned several times throughout this paper is most useful as a QA instrument in its capacity to gauge students' perceived learning, which is the first of two critical guideposts for quality of learning in EDC. The quantitative results of the spring 2019 survey are included as Appendix A. Here, 93.5% of the 2,499 respondents to this survey agreed or strongly agreed that their English speaking skills

improved after taking the course. Additionally, 97.2% agreed or strongly agreed that EDC lessons made them feel more comfortable using English to communicate, which supports the course structure in meeting its affective aims as outlined by Hurling (2012). These findings concord with a more detailed analysis of the quantitative and qualitative results of the spring 2018 survey (Brereton, Schaefer, Bordilovskaya, & Reid, 2019). Additionally but to a far lesser extent compared to these regularly administered student surveys, lesson observations provide a window into the classroom from a QA perspective to gain a sense of how students perceive their learning experience and environment.

The absence of any kind of quantitative data in the 2020 exit survey will severely inhibit course evaluators' ability to gauge students' perceived learning in English discussion class after the curriculum revision. It is likely that this critical guidepost will be impossible to evaluate within the new Center for Foreign Language Education and Research.

Actual Learning

Quality assurance in actual learning is currently performed by norming the assessment instruments for both regular and test lessons and then analyzing grading data recorded in the LRS. As outlined above, the Lecture Reporting System allows instructors to input individual student grades for discrete criteria on both rubrics after each lesson. This gives program evaluators an immense amount of data to determine students' actual learning, primarily with respect to the target language.

After the 2020 curriculum revision, instructors will simply report a single letter grade for each student to the university's Academic Affairs Office. Without rater norming of assessment instruments or the level of detail in recording students' grades using a dedicated database, program evaluators will have a proportionately vague impression of students' actual learning after the 2020 curriculum revision. As EDC will be reduced from two semesters to one after the curriculum revision takes effect, it seems sensible to predict that students' actual learning with respect to course aims will decrease compared to previous iterations of the course, though there will be no way to measure this.

DISCUSSION

Over the years, evaluating English Discussion Class has been both process- and product-focused. A number of hybridized evaluation instruments have collected information from a variety of sources to gain insight into program implementation and investigate the pedagogy underpinning course delivery, while simultaneously ensuring both program quality and accountability with respect to the eight critical guideposts discussed above. As the application of Lynch's (1996) connoisseurship model of curriculum evaluation has revealed, a majority of these evaluation instruments and data sources will not be available within the new English language curriculum at Rikkyo University when it launches in 2020, and those that do remain will be minimized in scope or strength. As such, it will be comparatively difficult to make confident judgments about the quality of any discrete course components or the effect of the curriculum revision in either general or specific terms. The present study failed to identify a rationale for abandoning the QA instruments that have been the status quo in administering English Discussion Class, though it appears fair to speculate that these losses are incidental to the transition from the existing management structure to a new one.

The new center will not maintain a program manager position tasked with conducting program evaluation. Instead, these responsibilities will be decentralized and shift to either a number of previous and newly hired professors charged with overseeing the new center's administration, or to the new Discussion Committee. On the eve of the new center's launch, detailed responsibilities for program implementation and accountability have yet to be designated

to any specific people or positions, which is cause for some concern not only with respect to quality assurance, but also to general course administration and the prolongation of a work culture that values quality and program evaluation.

The abandonment of so many underpinning QA measures is especially troubling, as it indicates that those driving the curriculum revision had to actively dismantle systems for QA in the Center for English Discussion Class rather than simply allowing them to continue. In some cases, this was done against the express recommendation of program managers and administrative staff at the Center for English Discussion Class. Certain QA measures, such as conducting classroom observations, take time and manpower, and so one could make a reasonable assumption that these measures were not deemed valuable enough for such resource apportionment. However, systems such as the LRS and Student Handbook require no such time or manpower to keep and continue, and so the decision to abolish them without a functional alternative is puzzling at best. The net result will be a far less unified syllabus that lacks clear performance standards for both teachers and students, and by extension has little accountability from curricular, institutional, or policy perspectives. Additionally, as the presence of faculty development schemes has been directly linked to the effective QA (Iijime, Takahashi, Watanabe, & Watari, 2020), the decision to diminish the preexisting faculty development will further compromise overall course quality and risk eroding EDC's quality culture (Rose, 2007).

Perhaps the most telling decision made by the creators of the new center is the resolve to remove any quantitative data collection in the exit survey of students. An analysis of the spring 2018 student survey data by Brereton, Schaefer, Bordilovskaya, and Reid (2019) powerfully demonstrates just how useful this survey is for conducting quality assurance of every critical guidepost for both quality of instruction and quality of learning in EDC, and the results of the spring 2019 survey appended to this paper further illustrate their usefulness in understanding how students perceive the course. The decision to diminish students' voices from any discussion about the 2020 curriculum revision is to minimize the very people the revision is meant to serve.

While growing pains during curriculum change are to be expected, many such pains could have been easily avoided in the case outlined here. With the removal of so many QA measures after the 2020 revision, course evaluators may never again have the capabilities to meaningfully evaluate EDC or any other courses overseen by Rikkyo University's Center for Foreign Language Education and Research. As such, EDC's evaluation culture is also at risk of vanishing. Leadership is an important part of maintaining an evaluation culture (Norris & Mills, 2014). If and when specific roles and responsibilities are ever articulated and assigned within the new Center's management structure, those charged with overseeing the new center and the Discussion Committee should both make a concerted effort to continue this culture in spirit if not in actual practice.

Failing such an effort, Rikkyo University's 2020 English language curriculum revision may prove to be yet one more "fresh façade" (Chapple, 2014) plaguing Japanese institutes of higher education preoccupied with marketing to prospective students rather than focusing on the educational experience of the students already enrolled there. Taking a commercial view of QA positions lesson delivery as a product and students as the consumers. From such a standpoint, QA eliminates defects to increase reliability and leads to higher student satisfaction and trust. This is critical if an institution stakes its reputation on its quality of instruction and learning. Therefore, if curriculum revisers are unwilling to listen to the stakeholders conducting program evaluation, or to their own students, then such revision can only serve to erode trust and investment in the curriculum and those who must implement or study within it.

CONCLUSION

This paper used Lynch's (1996) connoisseurship model of curriculum evaluation to inventory and assess the feasibility of implementing a number of quality assurance measures in an English discussion course at Rikkyo University before and after a major curriculum revision taking effect in April of 2020. The assessment of the QA measures' feasibility revealed that nearly all of those used by the Center for English Discussion Class from 2009 to 2019 will be unfeasible within the newly created Center for Foreign Language Education and Research. Despite the rather grim portrayal above, there is scope for several effective QA measures to be implemented if stakeholders in the new center are willing and organized.

With respect to teaching staff and teaching methodology, such measures could include reinstating intensive orientation onboarding and mandating regular faculty development and lesson observations for newly hired instructors, as well as instituting a peer observation or reflective professional development scheme for more senior instructors. If these are not organized in an official capacity by program administration, instructors could continue these evaluative practices in a teacher-driven, bottom-up way, though such an approach may not garner buy-in from the entirety of the teaching staff, thus dampening their efficacy as a QA measure. Administrators and teachers alike could also conduct independent evaluations of the new syllabus.

With respect to instructional materials, Discussion Committee members should continue collecting feedback on the course-specific textbook, but would do well to assign clear roles and deadlines for the completion of any revisions made to the textbooks. Additionally, norming of assessment instruments will be badly needed if students are to be assessed equally across the course, and the rubrics and expectations for student performance in the classroom should be articulated clearly and transparently to students. Ideally this would be in the form of an official Student Handbook like the one traditionally used in the Center for English Discussion Class. Barring that, teachers can determine collectively to pass this information to students on their own. Concerning the critical guidepost of support systems for students, the new center's management should assign responsibilities for implementing individual stages of the amended framework for supporting students with disabilities, and include accountability checks to ensure that no student with a disability goes without support.

With regard to the quality of student learning, perceived learning could be evaluated by reinstating the same exit survey used in and prior to 2019. Actual learning would benefit from rater norming of the revised EDC rubric, though the interrater reliability achieved by earlier iterations of the course will be a thing of the past.

Finally, it will more than likely fall to instructors in the new center to be change agents and maintain EDC's quality and evaluation cultures, especially considering that some of these instructors will serve on the Discussion Committee in one year or another. Curriculum revision can be a trying process for practitioners in any position, replete with unanticipated challenges and hidden impacts on quality of instruction and learning. When taken lightly by educators, it is ultimately the learners who lose out. It is strongly hoped, therefore, that EDC instructors continuing in the new center will remain committed to offering a learning experience held to the highest possible standards.

REFERENCES

- Auerbach, E. R. (1986). Competency-based ESL: One step forward or two steps back? *TESOL Quarterly*, 20(3), 411-30.
- Bailey, K. M. (2006). *Language teacher supervision: A case-based approach*. Cambridge, England: Cambridge University Press.
- Bloom, B. S. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*.

- New York, NY: David McKay Co Inc.
- Brereton, P. (2019). Teacher perceptions of working within a strongly unified curriculum. *New Directions in Teaching and Learning English Discussion*, 7, 257-271.
- Brereton, P., Lesley, J., Schaefer, M. Y., & Young, D. (2019). *What do you think? Interactive skills for effective discussion 1* (10th ed.). Tokyo, Japan: DTP Publishing.
- Brereton, P., Schaefer, M. Y., Bordilovskaya, A., & Reid, S. (2019). An analysis of student survey results on classroom experiences and learning outcomes within a large-scale English discussion program. *New Directions in Teaching and Learning English Discussion*, 7, 272-297.
- Brereton, P., & Young, D. (2018, May). *Towards autonomy in reflective teacher development*. Paper presented at Pan-SIG 2018 in Tokyo, Japan.
- Brinton, D. M., & Snow, M. A. (2017). *The content-based classroom* (2nd ed.). Ann Arbor, MI: University of Michigan Press.
- Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. In E.A. Soler and M.P.S. Jorda (Eds.), *Intercultural language use and language learning* (pp. 41-57). Dordrecht, Netherlands: Springer.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6(2), 5-35.
- Chapple, J. (2014). Finally feasible or fresh façade? Analyzing the internationalization of Japanese universities. *International Journal of Research Studies in Education*, 3(4), 15-28.
- Davis, J. M., Sinicrope, C., & Watanabe, Y. (2009). College foreign language program evaluation: Current practice, future directions. In J. M. Norris, J. M. Davis, C. Sinicrope, & Y. Watanabe (Eds.), *Toward useful program evaluation in college foreign language education* (pp. 209-226). Honolulu, HI: University of Hawaii, National Foreign Language Resource Center.
- Doe, T. (2012). Assessment: Improving rater reliability on the EDC discussion test. *New Directions in Teaching and Learning English Discussion*, 1(1), 1.11-1.18.
- Doe, T. (2014). Willingness to communicate and confidence in English discussion class. *New Directions in Teaching and Learning English Discussion*, 2, 3-10.
- Dörnyei, Z. (2009). The 2010s: Communicative language teaching in the 21st century: The 'principled communicative approach'. *Perspectives*, 36(2), 33-43.
- Dörnyei, Z., & Thurrell, S. (1992). *Conversations and dialogues in action*. New York, NY: Prentice Hall.
- Dörnyei, Z., & Thurrell, S. (1994). Teaching conversational skills intensively: Course content and rationale. *ELT Journal*, 48(1), 40-49.
- Dunn, M., Pratt, K., & Van Putten, F. (2020). Quality language education for diversity and inclusion: Challenges and proposals for quality improvement. In J. de Dios Martinez Agudo (Ed.), *Quality in TESOL and teacher education* (pp. 150-158). New York, NY: Routledge.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143-188.
- Farrell, T. S. C. (2016). The practices of encouraging TESOL teachers to engage in reflective practice: An appraisal of recent research contributions. *Language Teaching Research*, 20(2), 223-247.
- Farrell, T. S. C. (2018). *Research on reflective practice in TESOL*. New York, NY: Routledge.

- Fearn-Wannan, J., Kita, S., Sturges, J., & Young, D. (2019). *What do you think? Interactive skills for effective discussion 2* (10th ed.). Tokyo, Japan: DTP Publishing.
- Fearn-Wannan, J., Kita, S., Sturges, J., & Young, D. (2020). *What's your opinion? Interactive skills for effective discussion* (1st ed.). Tokyo, Japan: DTP Publishing.
- Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly*, 22(3), 473-492.
- Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *The Canadian Modern Language Review*, 61(3), 325-353.
- Honisz-Green, J. (2013). Critical pedagogy: Quality practices in coordinated EAP programmes. *The Journal of Rikkyo University Language Center*, 30, 59-72.
- Hurling, S. (2012). Introduction to EDC. *New Directions in Teaching and Learning English Discussion*, 1(1), 1.2-1.10.
- Iijima, Y., Takahashi, S., Watanabe, A., & Watari, H. (2020). EAP in Japan. In H. Terauchi, J. Noguchi, & A. Tajino (Eds.). *Towards a new paradigm for English language teaching: English for specific purposes* (pp. 79-92). Oxon, England: Routledge.
- Kamada, Y. (2014). Japanese university students' perceptions of native and non-native teachers of English. *New Directions in Teaching and Learning English Discussion*, 2, 11-20.
- Kehe, D., & Kehe, P. D. (1994). *Conversation strategies*. Brattleboro, VT: Pro Lingua.
- Krahnke, K. (1987). *Approaches to syllabus design for foreign language teaching*. Englewood Cliffs, NJ: Prentice-Hall.
- Lesley, J. (2017). Managing professional development in a unified curriculum: The FD planning and review cycle. *New Directions in Teaching and Learning English Discussion*, 5, 303-312.
- Lesley, J. (2018). Reflective teacher training: Uniting theory with practice. *New Directions in Teaching and Learning English Discussion*, 6, 261-272.
- Lesley, J., & West, T. (2019). Assessing the effects of increased task complexity on high-level students' academic discussion performance. *New Directions in Teaching and Learning English Discussion*, 7, 298-319.
- Livingston, M., & Moroi, T. (2015). Instructors' views on tasks outside the classroom: A report on a questionnaire survey. *New Directions in Teaching and Learning English Discussion*, 3, 333-351.
- Lynch, B. K. (1996). *Language program evaluation: Theory and practice*. Cambridge, England: Cambridge University Press.
- Mann, S., & Walsh, S. (2017). *Reflective practice in English language teaching*. New York, NY: Routledge.
- Mercado, L. A. (2012). Guarantor of quality assurance. In M. A. Christison & F. L. Stoller (Eds.), *A Handbook for language program administrators* (2nd ed., pp. 117-136). Miami Beach, FL: Alta Book Center.
- MEXT (2009). *Quality assurance framework of higher education in Japan*. Retrieved from https://www.mext.go.jp/component/english/_icsFiles/afieldfile/2011/06/20/1307397_1.pdf
- Moroi, T. (2014). Instructors' views on class size. *New Directions in Teaching and Learning English Discussion*, 2, 21-36.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.
- Nation, I. S. P., & Macalister, J. (2010). *Language curriculum design*. New York, NY: Routledge.
- Norris, J., & Mills, N. (2014). *Innovation and accountability in foreign language program*

- evaluation. In J. Norris & N. Mills (Eds.), *Issues in language program direction: Innovation and accountability in language program evaluation* (pp. 1-14). Boston, MA: Cengage.
- O’Leary, M. (2014). *Classroom observation: A guide to the effective observation of teaching and learning*. Oxford, England: Routledge.
- Richards, J. C. (2006). Materials development and research—Making the connection. *Regional Language Center Journal*, 37(1), 5-26.
- Richards, J. C. (2013). Curriculum approaches in language teaching: Forward, central, and backward design. *RELC Journal*, 44(1), 5-33.
- Richards, J. C. (2017). *Curriculum development in language teaching* (2nd ed.). Cambridge, England: Cambridge University Press.
- Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2nd ed.). Cambridge, England: Cambridge University Press.
- Rose, M. (2007). Leadership and a culture of quality. In L. Muresan, F. Heyworth, G. Mateva, and M. Rose (Eds.), *QualiTraining: A training guide for quality assurance in language learning* (pp. 23-31). Strasbourg, Germany: Council of Europe Publishing.
- Savignon, S. J. (1983). *Communicative competence: Theory and classroom practice*. Reading, MA: Addison-Wesley.
- Schaefer, M. Y. (2018). Communication skills for strategic competence. *New Directions in Teaching and Learning English Discussion*, 6, 273-279.
- Schaefer, M. Y. & Lesley, J. (2019). In-house journals as a form of professional development. *2018 PanSIG Journal*, 210-216.
- Schaefer, M. Y., & Young, D. (2019, December). *Revising assessment tools for a large-scale and strongly unified speaking course syllabus*. Paper presented at New Directions 2019, Yokohama, Japan.
- Stillwell, C., Kidd, A., Alexander, K., McIlroy, T., Roloff, J., & Stone, P. (2011). Mutual benefits of feedback on materials through collaborative materials evaluation. In Tomlinson, B. & Masuhara, H. (Eds.). *Research for materials development in language learning*, pp. 257-272. London, England: Bloomsbury Academic.
- Underwood, P. (2012). The course of study for senior high school English: Recent developments, implementation to date, and considerations for future research. *Toyo Eiwa Journal of Humanities and Social Science*, 30, 115-145.
- Widdowson, H. G. (1989). Knowledge of language and ability for use. *Applied Linguistics*, 10(2), 128-137.
- Young, D. (2016). Textbook revision in the EDC context: Readability and topic interest. *New Directions in Teaching and Learning English Discussion*, 4, 295-302.
- Young, D. (2017). Textbook revision in the EDC context: Lesson activities. *New Directions in Teaching and Learning English Discussion*, 5, 313-319.
- Young, D. (in press). Student perceptions of textbook topics in a discussion course. *OnCUE Journal*.
- Young, D., & Nicklin, C. (2019). Item analysis of the EDC reading quizzes. *New Directions in Teaching and Learning English Discussion*, 7, 320-333.
- Young, D., & Schaefer, M. Y. (2019). Collaborative support for students with disabilities. In P. Clements, A. Krause, & P. Bennett (Eds.). *Diversity and inclusion*, Tokyo: JALT, pp. 136-142.
- Young, D., Schaefer, M. Y., & Lesley, J. (2019). Accommodating students with disabilities studying English as a foreign language. *Journal of Post-secondary Education and Disability*, 32(3), 311-319.

APPENDIX A

English Discussion Class spring 2019 student survey results (n=2,499).

Item	Strongly agree	Agree	Disagree	Strongly disagree
1. EDC lessons made me feel more comfortable using English for communication.	49.7%	47.5%	2.3%	0.6%
2. I deepened my understanding of the topics we discussed in the course.	39.5%	56.4%	3.7%	0.4%
3. The homework reading in the textbook helped me prepare for lessons.	59.9%	43.8%	4.6%	0.7%
4. The textbook activities helped me prepare for discussions.	51.4%	45.7%	2.5%	0.4%
5. The Discussion Skills (e.g. Opinions) helped me to participate in discussions.	64.2%	34.1%	1.2%	0.4%
6. The Communication Skills (e.g. Paraphrasing) helped me to participate in discussions.	54.6%	41.8%	3.2%	0.4%
7. The discussion test was a good way to assess my discussion ability.	42.0%	50.2%	6.5%	1.2%
8. The teacher's feedback during lessons helped me to understand what skills I needed to use more.	56.5%	40.4%	2.8%	0.3%
9. The teacher's feedback on the EDC website helped me to understand what skills I needed to use more.	40.4%	52.3%	6.7%	0.6%
10. The teacher gave clear instructions for using skills and completing activities.	65.0%	33.6%	1.2%	0.2%
11. The teacher gave enough time to discuss their opinions and ideas.	72.6%	26.1%	1.1%	0.2%
12. EDC's small class size helped improve my discussion ability.	67.7%	30.0%	1.8%	0.5%
13. After taking EDC, my English speaking skills have improved.	48.0%	45.5%	5.6%	0.9%

APPENDIX B

English Discussion Class spring semester grade distribution, 2010-2019.

