cogent
education

*Corresponding author: Lindsay Mack, Center for Language Education, Ritsumeikan Asia Pacific University, Beppu, Japan
E-mail: lindsay.mack287@gmail.com

EDUCATIONAL ASSESSMENT & EVALUATION | RESEARCH ARTICLE

Validating the normalization of vocabulary systems in a university EFL program

Lindsay Mack[1]*, Paul Sevigny[1], Malcolm Larking[1] and Lance Stilp[1]

Abstract: The integration and normalization of multiple CALL systems with more traditional face-to-face learning, or blended learning, is an emerging trend of research. Evaluators are urged to investigate the processes involved in normalization of language learning tools in the classroom. Grounded in blended learning evaluation, this paper adopts an argument-based approach to interrogate the single claim: blended vocabulary learning systems can be normalized in English as a foreign language (EFL) programs. Using needs analysis, survey data, and document analysis, the authors examine what factors contribute to implementation and normalization of a blended language vocabulary program in a private Japanese university EFL program. Results reveal that transparency, constructive alignment, and coordinator's and teacher's knowledge, skills, and attitudes are important factors. Finally, it is argued that the normalization of blended learning programs seems to be predicated upon careful alignment with well-defined learning objectives, and on the availability of transparent analytics from online systems for the learner, teacher, and coordinator to ensure the diffusion of alignment from the meso to the micro level of the program.

Subjects: Higher Education; Bilingualism / ESL; Curriculum Studies

Keywords: Normalization; blended language learning; vocabulary learning; constructive alignment; technology integration

ABOUT THE AUTHORS

Lindsay Mack, Paul Sevigny, Malcolm Larking, and Lance Stilp share a variety of research interests including vocabulary acquisition, academic writing, pedagogical stylistics, discussion, reading strategies, bilingualism, and conversation analysis. The research reported in this paper was part of a larger review of an entire English Program's learning objectives and goals based on an Assurance of Learning (AOL) framework. When they started this project, they noticed a lack of studies on vocabulary assurance of learning and evaluation and hope this article contributes to that field. They hope to work together on future research projects for other skills besides vocabulary.

PUBLIC INTEREST STATEMENT

Coordinators of large English as a Foreign Language (EFL) programs are tasked with determining vocabulary learning goals for language courses and programs, and then blending digital and paper vocabulary learning systems to deliver relevant practice and assessment to individual learners. This case study followed an argument-based evaluation method to determine whether EFL blended vocabulary learning programs can be normalized and what factors impact the normalization process. The authors found that vocabulary programs have the potential to be normalized when careful consideration is given to the curriculum, teacher training, and the needs of the students. This approach of argument-based program evaluation may provide a blueprint for others interested in evaluating a blended vocabulary learning curriculum.

cogent··oa

cogent ·· education

1. Introduction

Language learners have an ever-expanding number of online learning tools for the study of vocabulary. English vocabulary, however, covering hundreds of thousands of head words, each with its own family of derivational and inflectional morphemes, has remained somewhat impervious to being broken down into manageable units for standardizing student learning objectives. To complicate matters further, individual learners may exhibit a wide variety of differences in their vocabulary knowledge. Due to these circumstances, Computer Assisted Language Learning (CALL) systems have intervened with spaced repetition systems (SRS), and individualized learning algorithms are leading the way in CALL vocabulary systems, but is it possible to integrate CALL vocabulary systems for a university EFL program as a whole? The full integration of a particular CALL system into the language learning classroom, *normalization*, is characterized as when a technological tool becomes invisible, or completely embedded in the learning practice (Bax, 2003, p. 23).

This article investigates the claim that blended vocabulary learning systems can be normalized in EFL language programs. This is not a given because, in short, not all aspects of vocabulary learning can easily be supported by computer applications, and it is often unclear or difficult to ascertain which aspect of vocabulary learning an online vocabulary system supports. The article will follow an argument-based approach (Gruba et al., 2016) for a blended language vocabulary program evaluation. Such approaches have been recommended especially for investigating specific layers in language programs, such as the micro level (classroom), meso level (coordinator level), or the macro level (upper administrator and above) (The Douglas Fir Group, 2016). This paper starts with a review of relevant literature related to program evaluation and blended vocabulary learning programs and the context of this study. Next, we follow the four phases of an argument-based approach focused at the meso level: planning the argument, gathering the evidence, presenting the argument, and appraising the argument (Gruba et al., 2016).

1.1. Background on blended language systems for vocabulary

Commonly in EFL classrooms, because time is limited, vocabulary study and practice is often relegated to the student outside of class. Therefore, to help learners acquire a mastery of vocabulary, online spaced repetition programs are used as one method to expedite this process. In the current market, a plethora of online vocabulary learning systems exist each with their own uniqueness and claim for effectiveness. In spite of the autonomous learning nature of online vocabulary learning systems, many institutions are now integrating these vocabulary systems in their EFL curriculum through blended learning.

The term blended learning can be defined in a number of ways, but for this paper, the term means any system or curriculum that combines both traditional classroom instruction with online components (Bonk & Graham, 2006). The majority of studies on blended learning vocabulary have supported positive effects on improving vocabulary knowledge (Tosun, 2015). Bielawaski and Metcalf (2003) argue that true blended learning emphasizes achievement of learning objectives through careful consideration of individual learning styles and the application of various technologies that promote learning in a variety of ways. More institutions are adopting the idea to use technology both in and outside the classroom as a way to cover the gaps between learner differences (Marsh, 2012). For large post-secondary institutions in particular, this approach to learning is particularly attractive to achieve broad curriculum objectives for a large body of students. Blended learning has become the new norm; it is not a radical approach to constructing curriculum, but a necessary part in an age of high technology and online interactional learning environments.

1.2. Background on argument-based approach to blended language program evaluation

The evaluation of a blended language program, like other forms of evaluation and assessment, starts with concerns of validity. Validity arguments were first developed in pursuit of lending credibility to interpretations for norm-referenced proficiency tests (Chapell et al., 2010; Chapelle et al., 2008; Kane, 2006). More recently, validity arguments have been transformed into the argument-based approach to evaluation as first postulated with the following four phases: planning an argument, gathering evidence, presenting an argument, and appraising an argument (Chapelle, 2014). Furthermore, argument-based

cogent ·· education

validation methods have been successfully adapted to the evaluation of technology in learning environments, first theoretically, through applying neo-Vygotskian frameworks (Bax, 2011), then with ethnographic study, and then as applied as rationale for a blended learning program design (Gleason, 2013).

Recently, there has been an emerging trend of research in applied linguistics to apply the argument-based approach to the evaluation of technology in blended learning programs advocated by Gruba et al. (2016). Employing this approach, the integration of technology was recently evaluated at a leading university in Vietnam (Gruba & Nguyen, 2019). These researchers found that "failure to understand the complex nature of meso level influences may lead to the poor uptake of recommendations for program improvement" (Gruba & Nguyen, 2019, p. 634), and concluded that these difficulties warrant further development of new approaches to evaluate blended learning or CALL systems (p. 634). More relevant to this study was a recent argument-based investigation of the claim that materials are constructively aligned in modern language programs, with the assumption that if blended technology materials are constructively aligned, then they will normalize (Yoon & Gruba, 2019). Unfortunately, the main finding of this study was a lack of ability of technology use to translate into aligned pedagogical achievements. They argue that in all likelihood, true normalization is actually predicated upon careful constructive alignment of material design and language outcomes. These researchers, like Bax (2011), question whether there is a surface-level appearance of normalization that is not aligned with actual learning outcomes.

Despite the plethora of studies on online learning and CALL, "few to date have focused on integration or blending of technology in face-to-face language programs" (Gruba et al., 2016, p. 18). Furthermore, there are very few studies on language program evaluation in general. According to Norris (2016), language learning outcome evaluation "is currently lacking in the midst of rapid (rabid?) innovation and deployment of technology-mediated language courses and programs" (p. 180). This paper emerged from the desire to contribute to this important topic, blended language program evaluation.

2. Situating the study

The blended language vocabulary curriculum that was under review was part of the English Program at a midsize private international university in Japan titled X University (pseudonym). XU is located in a small city in southern Japan. XU is an international university with a 50/50 ratio of international and domestic (Japanese) students. A core focus of the students' degrees is on foreign language education. Participants in this study are Japanese-basis students (mostly domestic) who must take EFL courses up to an Upper Intermediate level, which corresponds to the B1 level of the Common European Framework of Reference for Languages (CEFR), to pass the requirements of the English standard track. Each level of the English Program from elementary to advanced is split into an A course with four classes per week, focusing broadly on productive language skills, and a B course of two classes per week, focusing more on receptive language skills.

The motivation to evaluate online vocabulary learning tools came out of a larger review of the learning objectives and goals of the entire English Program, based on an Assurance of Learning (AOL) framework. The XU English Program is managed by meso-level administrators, with associate professors and senior lecturers, who are responsible for curriculum design. From 2018, the management team embarked on an Assurance of Learning Project, with a commitment to structural change. A vocabulary team was established to review all the vocabulary methods of instruction, including a thorough review of online learning tools currently used, and other potential online systems to be trialed, evaluated, and implemented.

3. Planning an argument-based evaluation

3.1. Establishing purpose, scope and stakeholders

The primary purpose of our evaluation was to examine the vocabulary learning program at our university, both blended and face-to-face, and to explore the factors affecting implementation and normalization. We focused our evaluation on the meso level, the English Program in general,

instead of focusing on individual classes at the university. One reason we focused on the coordinator level is because it is "a uniquely autonomous and powerful yet also uniquely interdependent structure within a very complex system" (Walvoord et al., 2000, p. 33). Limiting our analysis to the meso level seemed especially apt since the majority of the decisions about curriculum and blended learning take place at the meso level by the course coordinators. Even though the document analysis and member checking take place at the meso level, the data is still triangulated by eliciting students' and teachers' opinions and students' performance at the micro level.

The primary stakeholders for this evaluation are four course coordinators who comprise the vocabulary team at XU university, and the authors of this paper, whose research set out to improve a blended learning vocabulary curriculum. For our evaluation we followed the process use, which means instead of limiting our evaluation findings to make desired program changes at the end, we instead employed the process of evaluation itself to achieve desired program changes (Norris et al., 2009). As taking the dual role of key stakeholders and the researchers, we were specifically conducting insider research. Since the 2000s, evaluation researchers have advocated for local ownership of the evaluation undertaking. Although this created a risk for potential bias, given the specific purpose of our research to improve vocabulary blended learning practice through understanding, influencing, and changing the direction of the program, we felt conducting insider research was justified. In line with insider research (Fleming, 2018), we believed our role as insiders was more of a strength than a hindrance. Moreover, we could use our unique perspective as insiders to enable "a deep level of understanding and interpretation which outsiders may not be able to uncover" (Fleming, 2018, p. 312). The other stakeholders comprise the students and teachers in the English Program, whose opinions and performance were extremely important to triangulate our data, but also to accurately understand the factors of successful implementation.

3.2. Focal questions and main claims for evaluation

To begin our research, the primary stakeholders, the vocabulary team, collaboratively decided the purpose of our study and the main claim to interrogate.

Main Claim:

• Blended learning for vocabulary can be normalized in EFL university contexts.

Focal Question:

• What are the factors affecting the implementation and normalization of a blended learning program for vocabulary in a University EFL program?

By choosing one focal question we intended to elucidate what aspects of vocabulary learning can be handled by online systems and what can be handled by supplemental methods and if blended learning vocabulary study can indeed be normalized. This investigation aligns with Bax's 2003 research in which he promotes that normalization should be the central goal of CALL.

Following Gruba et al. (2016), our vocabulary evaluation follows four specific stages: planning an argument, gathering evidence, presenting the argument, and appraising it. After establishing the purpose and main claim to interrogate, we planned the argument (see Table 1). The argument-based approach employed has five inferences: domain definition, evaluation, explanation, utilization, and ramification. The claims were developed in line with Gruba et al. (2016, p. 141) "by brainstorming the types of conclusions that could be drawn from each inference in the context of meso-level evaluation" of blended learning for vocabulary.

As shown in Table 1 in planning our domain definition, our main warrant rested on the key assumption that the methods of data collection are appropriate and triangulated to successfully elucidate the factors affecting blended learning normalization for vocabulary. The evaluation

Table 1. Warrants, assumptions, and appropriate evidence backed by each inference			
Inference	**Warrant**	**Assumptions**	**Examples**
Ramification	The XU specific meso-level evaluation findings will have broader implications for other universities when adopting blended learning systems for vocabulary.	Findings are broadly disseminated in academic contexts. University language programs that have not developed explicit objectives for blended vocabulary programs will collaborate with universities that have done so. Online systems will develop more transparent analytics for items, item bands, and productive practice.	Dissemination through presentations and research articles instructs meso-level administrations of important micro-level impacts of adopting a blended learning program for vocabulary.
Utilization	The explained findings help stakeholders to understand online systems in use and to make changes to improve blended learning normalization for vocabulary.	The findings are interpretable by stakeholders.The instructors can use the findings to improve constructive alignment of vocabulary materials and online tools.Meso-level administrators can use findings to refine development for teachers and make decisions on how to improve normalization of blended learning for vocabulary.	Instructors and coordinators can refine or adjust range of vocabulary for a level. Instructors and coordinators can identify hard-to-acquire items and create alternative practice systems.Meso level led changes to systematic improvements in calibration of learning targets and achievement for vocabulary.
Explanation	The evaluation findings are explained with consideration of the context of XU at the meso level.	The findings can be explained within the context of the English department of XU at the meso level and will explain what factors meso-level administrators should consider when implementing new blended learning initiatives for vocabulary.	Thematic analysis of survey responses, meso-level curriculum documents, and needs analysis documentation was conducted.
Evaluation	Documents relevant to the language program, online tools observed in use, and survey data from instructors and students are analyzed to provide comprehensive findings.	The analysis of findings from the data can provide the factors affecting blended learning implementation and normalization for vocabulary.The analysis is conducted accurately, rigorously, and ethically.	Research ethics and methods are systematically followed. Descriptive statistics are used for student and teacher responses to survey questions. Thematic analysis of qualitative data from online teacher and student survey was conducted.
Domain Definition	The XU English Program and data gathered from administrators, instructors, and students provide adequate insights on the factors and variables contributing to blended learning normalization for vocabulary.	Appropriate sources on blended learning for vocabulary have been identified.The methods of data collection are appropriate and triangulated to provide a view of blended learning normalization for vocabulary.	Review of literature on blended learning programs for vocabulary, surveys, needs analysis, and document analysis was conducted.

inference links this collected data to the results of the analysis of the data, thus this inference is where our main evaluation will be located. Next, the explanation inference links the evaluation findings to the reasons that can explain these findings in the context of our program, while the utilization inference links these findings to how we can utilize the findings to improve our program. Finally, the ramification inference links our local findings to a discussion of broader implications for other blended language programs.

4. Gathering the evidence

4.1. Methods

In order to gather evidence for the evaluation inference, efforts were undertaken including the following methods: a) needs analysis b) surveys about online vocabulary learning c) surveys about productive vocabulary learning with paper systems, and d) AOL document gathering and analysis. See Table 2 for details on each method of inquiry. As mentioned previously, our evaluation was a process in use, therefore we used the data collected in phase one and two to inform phase three and four. These methods were employed over four phases.

Table 2. Methods of enquiry		
Method of Enquiry	**Details**	**Reason to Include**
Surveys about online learning	Phase 1—Fall 2018 administered the survey on the old online vocabulary blended learning system (n = 258)Phase 2—Spring 2019 administered surveys related to competing online systems as tested for utility as platforms for NGSL study (n = 40)Phase 3—Fall 2019 administered survey on the new online learning system to students (n = 182) and teachers (n = 10).Phase 4—Spring 2020 administered the same survey in phase 2 about the online learning systems to Intermediate, Upper Intermediate, and Advanced English students (n = 645)	The surveys were used to elicit responses from students and instructors on a range of views of the online vocabulary program including implementation and integration.
Surveys about productive vocabulary learning	Administered in phase two, Spring 2019 to Elementary English A course students (n = 54) and Upper Intermediate English A course students (n = 270).	The surveys were used to evaluate the effectiveness of the productive lists at increasing the students' productive vocabulary knowledge and to inform the wider implementation and normalization of the lists across the program.
Needs Analysis	Conducted in phase two with Pre-Intermediate students. Data included observation field notes, video recordings, student surveys, and test data. Other online systems were then analyzed with the criteria to compare multiple alternatives. (n = 42)	Data from two specific classes added thick descriptions of classroom realities (Geertz, 1973). Resulting criteria enabled evaluation of online alternatives prior to piloting.
Document Analysis	Conducted throughout the phases. Documents analyzed included AOL monthly vocabulary team plans and meeting minutes, AOL, curriculum documents, online learning tool information, productive vocabulary lists, teacher vocabulary material.	Analysis provided context about the objectives, practices, and resources for a blended learning program for vocabulary in XU.

- Phase 1: Evaluation of prior vocabulary systems (Fall of 2018)
- Phase 2: Needs analysis and single-class piloting (Spring of 2019)
- Phase 3: Single-level implementation of new online system (Fall of 2019)
- Phase 4: Wide-scale implementation and evaluation (Spring of 2020)

4.2. Anonymous surveys about online vocabulary learning

Anonymous surveys were carried out in all phases of the AOL initiative. The questions surveyed students on how useful they thought the online learning system was, whether the system had placed them in the correct level, how much they enjoyed using it, and their choice of platform to study (PC or smartphone). At the end of the survey, two open-ended questions were asked: if they had any general comments about the online system, and what they thought about vocabulary learning in general in their English courses. In addition to the student surveys, teachers were also surveyed for their opinions which included questions about how easily they could adopt blended learning in their classroom. In sum, the surveys remained consistent across all phases to ensure results were interpretable in the best means possible.

4.3. Anonymous surveys about productive vocabulary lists

In order to assess the efficacy of the productive vocabulary lists, student surveys were administered. The productive vocabulary lists provided the students an opportunity to practice their productive vocabulary skills in an offline context. The lists contained 20 target words which corresponded to the textbook unit of each course and were supplemented with topic-specific words from the relevant band of the New General Service List (NGSL) (Browne et al., 2013). In the Spring 2019 and 2020 semesters, anonymous online surveys about the productive vocabulary lists were administered to the elementary English A (N = 54) and upper-intermediate English A course students (N = 270). The surveys were used to evaluate the effectiveness of the lists at increasing the students' productive vocabulary knowledge, and to inform the wider implementation of the lists across the program. The surveys asked questions about the usefulness of the lists and whether the students could successfully use the words in their writing and speaking. The surveys also asked about the students' preference for the lists over other vocabulary study methods and included open-ended questions regarding their opinions about the lists, and vocabulary study in their courses in general. The students responded to Likert-scale survey questions that were quantified as follows: 1 = not at all, 2 = slightly, 3 = moderately, 4 = very much, 5 = definitely. A few questions on each survey were multiple choice and were converted to a 5-point Likert scale for consistency and analysis.

4.4. Needs analysis

A Needs Analysis (NA) was conducted in the fall of 2018 and spring of 2019, following Long's (2005) recommendation of utilizing insider experts and the triangulation of sources and methods. The NA included two initiatives: investigating the utility of the New General Service List (NGSL) for the Standard Track of the English Program and determining the utility of various online potential components for meeting student learning objectives as they were developed by the vocabulary team. NGSL recall tests, an item analysis on the NGSL recall test, and textbook vocabulary analysis were undertaken to gauge the breadth and depth of knowledge of learners in the standard track with respect to knowledge of the NGSL. To determine the utility of potential online components, the vocabulary team investigated alternative online flashcard systems and online vocabulary programs in the process of determining their functionality for supporting the learning of relevant learning objectives as they were specified by assurance of learning. As part of the needs analysis, two pre-intermediate English classes were vetted for equivalent average proficiency in English. The course content and teacher were the same except for the online vocabulary programs. In each class, volunteer participants were video recorded using different vocabulary systems on both mobile phones and PCs in a computer lab. Volunteers also described ways they used the systems and

cogent ·· education

the obstacles they experienced in the process. The teacher researcher extracted factors impacting online vocabulary program success from the data with the two online systems.

4.5. AOL documentation

The aim of collecting AOL and vocabulary material documentation was to provide context about the objectives, practices, and resources for a blended learning program for vocabulary in XU. As insiders to the department, the researchers had access to this material and in fact were the authors of some of the material. In line with Bowen (2009), documents were determined by their relevance to the research questions and ability to supplement and corroborate other sources of evidence. In the end, we collected five different types of documents: AOL Monthly Vocabulary Team Plan, AOL Curriculum Documents, Online Learning tool information, Teacher Vocabulary Material, and Productive Vocabulary Lists for each course.

4.6. Data analysis

Throughout the four phases, vocabulary team members met monthly to review the data from the surveys, needs analysis, document analysis, and relevant literature. This iterative process proved to develop a thick description of the needs of the learners and to specify student learning goals. We employed the process use evaluation (Norris et al., 2009) to achieve desired program changes throughout the four phases. Phases one and two of the research findings were used to evaluate the prior online system, pilot a new system, and decide our vocabulary learning objectives. In phases three and four, the data generated evidence for the inferences in our argument-based approach to evaluate and explain our program. The quantitative data from the survey was analyzed using simple descriptive statistics, such as frequency counts, mean, and mode (Nunan & Bailey, 2008). Descriptive statistics were useful for this study because they could be employed to illuminate the teachers' and students' perceptions of the blended vocabulary program and the new online learning system. The qualitative data from the open-response questions at the end of the survey was analyzed through an interpretivist framework in which we coded and analyzed the data for recurring patterns and thematic constructs and interpreted the data by drawing on past research and personal reflections (Creswell, 2008). Documents were first analyzed using content analysis, by skimming through the documents and deciphering what data is important and relevant and which data can be disregarded (Bowen, 2009). After the meaningful data was determined, thematic analysis was conducted, and patterns and themes emerged. The main themes that emerged from the data will be reported in the evaluation and explanation inferences.

5. Presenting the argument

The purpose of this section is to present the findings for each inference in the planned argument. The first five inferences in the argument are presented: domain definition, evaluation, explanation, utilization, and ramification.

5.1. Domain definition

The main assumptions at the domain level were that appropriate sources on blended learning for vocabulary have been identified and that the methods of data collection are appropriate and triangulated to provide a view of blended learning normalization for vocabulary.

The most important task during our needs analysis was deciding which online vocabulary learning tool to use. For normalization to occur, we needed a transparent reporting of detailed learner performance and control over NGSL list segmentation, practice, and weekly testing. Many systems were vetted to ensure they could be centered on NGSL recall and that transparent analytic reports and quiz score details would provide the necessary data to determine whether the system was working for the population and the extent to which learners were completing weekly learning goals. In this stage we were able to pilot (phase two) and implement (phase three) a new online

learning system, tool A. With the introduction of a new online system, our inference was appropriate enough to continue our argument-based approach.

5.2. Evaluation

We employed the evaluation inference to focus on focal question one: What are the factors affecting the implementation and normalization of a blended learning program for vocabulary in XU? Our evaluation inference was based on our assumption that the analysis of findings from phases three and four (after a new online vocabulary system was chosen), and documents relevant to the language program and the survey data, can provide the factors related to implementation, and that the analysis is conducted accurately, rigorously and ethically.

Our assumption that our data was collected accurately and rigorously is connected to employing multiple methods, members checking and our analysis of triangulated data (documents relevant to the language program, needs analysis and the survey data). In the end, we discovered the two key factors (with smaller subthemes connected) that affected the blended learning implementation and normalization for vocabulary learning: constructive alignment and coordinators' and teachers' knowledge, skills, and attitudes (KSA). Below in Table 3 are the themes and subthemes that emerged from our data analysis and some examples of the sample data that accounted for that code.

5.3. Constructive alignment

Constructive alignment defined by Biggs (2014) is an "outcomes-based approach to teaching in which the learning outcomes that students are intended to achieve are defined before teaching takes place" (p. 5). Assessment and instruction methods and material are then designed to best achieve those outcomes and to assess the quality at which they have been achieved. Policies and systems that make the goals clear to all stakeholders are what empower normalization and constructive alignment. After these are achieved the goals should be communicated in a transparent process. However, in our context, during phase one and two the vocabulary goals were not constructively aligned. As mentioned in our domain inference, we realized that our online learning system was not transparent, nor did we have clearly defined learning objectives. After phase one and two of our investigation, clear vocabulary goals were designated and presented in a transparent process. Further analysis identified two subthemes of constructive alignment: blended language policy and transparency. See Table 3 for more details.

5.4. Coordinator and teacher KSA

The context of this EFL program is a large program with approximately 40 teachers and 11 coordinators, each with their own preference and teaching style. The analysis of the data revealed that it is hard to create clear vocabulary policies and guidelines in a large program with different coordinators and teachers each with their own KSA for vocabulary practice. Since our program is highly coordinated to meet AOL goals, it is up to the coordinator to implement these goals using blended learning programs. Each coordinator and teacher had their own vocabulary learning tool. Preferences here are part of attitudes towards technology acceptance. There has been a prolific amount of research establishing attitudes towards technology as one of the main factors enabling or resisting technology (for example, see Oxford & Jung, 2007). Some teachers were very willing to grow their expertise in online vocabulary blended language programs. In other words, they have strong buy-in to the program. However, other teachers were resistant to the new blended language program and less willing to learn a new program. This resistance was a hindrance to normalizing the program. Further analysis identified two sub-themes of coordinator and teacher KSA: willingness and resistance. See Table 3 for more details.

cogent •• education

Table 3. Coding themes, definitions, and sample data		
Theme/Subtheme	**Working Definition**	**Sample Data**
1. Constructive Alignment	Creating vocabulary learning goals, standards, and assessments that are coherent and connected in a curriculum.	"The words we learned in our class were kind of easy for me, so I want to learn more difficult vocab."Online learning survey, Fall 2019, student"Once we determined the correct band of NGSL for the course level and the extent to which we should test recall and accuracy in terms of derivational and inflectional morphemes, then I knew our system was correctly aligned." Vocabulary team report, Vocabulary Team Member
1a. Blended learning policies	A set of specific policies or guidelines for teachers and administrators for how to use online learning tools.	NA
1b. Transparency	Clear blended language vocabulary goals and policies communicated to coordinators, teachers and students.	"Sometimes there are some problems with Vocab Tool A test's vocabulary list. Because I don't know how it works, it was hard to get a good score."Online learning survey, Spring 2020, student
2. Coordinator/ Teacher KSA	The role the coordinator takes depends on their KSAs (knowledge, skills, and attitudes) towards blended learning practices.	"I think it is great we are trying out different learning systems. But I am worried that now that coordinators have spent so much time on these two different vocabulary programs they won't want to change if we discover that one is working better than the other. I am wondering why one course is using a different vocabulary online system than another course."Vocabulary team report, Vocabulary Team Member
2a. Willingness	The willingness of coordinators and teachers to adopt and adapt to a new online learning tool.	"I'm a big fan of SRS systems for vocab learning, so (Tool A) definitely seems like a good route to follow." Online learning survey, Fall 2019, teacher
2b. Resistance	The resistance coordinators and teachers have toward adopting and adapting to a new online learning tool.	"It may be worth the time and effort for us to develop our own vocabulary tests from the NGSL. This would make delivery of the tests more flexible and give us more ready access to the performance data. It would also provide some continuity if we were ever to move away from Tool A as a study tool."Online learning survey, Fall 2019, teacher

5.5. Explanation

The explanation inference was founded on the warrant that the evaluation findings can be analyzed in consideration of the meso level of the XU English Program. This warrant was based on the assumption that the findings from a thematic analysis of surveys, curriculum documents, and needs analysis can explain what factors the XU meso-level coordinators should consider when implementing a blended learning vocabulary program and what factors affect normalization.

Furthermore, as insiders, our team ensured that analysis reflected "a deep understanding of the program context, the departmental culture and issues relating to technology integration" (Gruba & Nguyen, 2019, p. 629). Identifying such factors is essential for meso-level coordinators, as they are often the sole decision makers in this context. As Gruba and Nguyen (2019) note that "given the absence of macro level guidelines and policies our analysis shows that the bulk of decisions related to technology integration are made at the meso level" (p. 629). Our experience corroborates this division of administrative decision-making, as the curriculum changes made by the vocabulary team had very little oversight from the macro-level administrators.

The difficulty of constructive alignment across a large program was alleviated by using the AOL learning objectives to explain to macro-level administrators, teachers, and students the purpose of curriculum changes. Without such a framework, aligning vocabulary curriculum reform across all levels of the program may have been harder to justify to the stakeholders. Developing a sequence from program-wide objectives to course goals and finally to specific can-do statements for vocabulary, created a solid reasoning for program changes. Although a clear reasoning for adopting a blended learning program was lacking, the refocusing of the learning objectives to more productive vocabulary usage helped guide our transition to the kind of blended learning program we adopted.

The documentation of these steps also makes the reform measures transparent, as the other administrators were kept informed with regular reporting from the vocabulary team and had access to all the vocabulary AOL documentation. Without this transparency and reporting, a vocabulary curriculum runs the risk of having isolated course goals, without alignment between different program levels, coordinators, and teachers. Being able to clearly show the proposed sequencing of vocabulary instruction to the stakeholders certainly helped with the vertical integration of the curriculum changes.

A second factor regarding implementation is clearly coordinator and teacher KSAs. Multiple surveys of teachers over several semesters of AOL vocabulary work have shown that teacher knowledge and preferences are not only a limiting factor, but that coordinator KSA regarding online system alternatives carry a multiplier effect, because coordinator level decisions often determine the online applications implemented in entire levels and programs, affecting thousands of students at once. Furthermore, the enthusiasm and clear understanding of teachers lead to the magnitude and clarity of the message repeated to students by teachers. The themes of teacher/coordinator KSA, willingness, and resistance all speak to the importance of AOL vocabulary objective mapping and especially to the importance of teacher training time and space to ensure that teacher knowledge is accompanied by the skills and attitudes that will lead to carrying the program in a truly coordinated fashion.

5.6. Utilization

Firstly, the explained findings help stakeholders to understand online systems in use and to make changes to improve blended learning normalization for vocabulary. In phase one and phase two we were able to a) determine specific vocabulary objectives and b) identify which parts of the program could target our specified objectives for students. In our small-scale trial, feedback from students and instructors quickly revealed how the initial bands proposed in our vocabulary objectives were far above the level of the students.

From there, in phases three and four we were able to c) make decisions on how to improve normalization and d) refine teacher training and development. Our team decided that it was the role of the meso-level administrators to communicate the necessity of adopting new technology and study methods to teachers, as Gruba and Nguyen observe that "the teacher willingness or resistance to using technology in their classrooms is greatly affected by the departmental and sectional leadership" (Gruba & Nguyen, 2019, p. 629). From our experience, being able to reference

our AOL objectives to justify the purpose of change helped teachers' willingness to take on the vocabulary curriculum reforms and the new online learning tools we were implementing.

For teacher development, we used the data to inform professional development workshops we facilitated with online learning program providers. We discovered that having a strong relationship with the technology provider, with access to explanations and training for teachers, is essential in fostering a willingness to adopt new online learning tools.

5.7. Ramification

Two assumptions underlie the ramification claim, that the findings are disseminated and that it will help other meso-level administrators introduce new blended language programs. We are currently disseminating our research through conference presentations in Japan at National teaching and CALL conferences. As for the second assumption, it is our hope that the specific meso-level evaluation findings for XU will have broader implications for other universities when adopting blended learning systems for vocabulary.

6. Appraising the argument

To review, the main claim under interrogation in this article is whether a blended vocabulary learning program can be normalized and which factors are important in implementation and normalization.

The ultimate appraisal of the strength of a claim in an argument-based approach has been formulated as having three levels according to Golonka et al. (2014):

(1) Weak—based mainly on anecdotal evidence
(2) Moderate—based on one well-designed study
(3) Strong—based on multiple well-designed studies

Here we have evidence for the first three inferences: domain, evaluation and explanation. Analysis of the domain inference came from phases one and two of the study, which included survey data, document analysis, and needs analysis. Evidence at the domain level of the argument revealed that there was a lack of transparent analytic data from the former online vocabulary tool used in the program. Furthermore, the prior online system did not meet the domain definition for appropriacy because of lack of transparency. This transparency needs to be supported by online system dashboard designs and program policies for how to communicate goals and progress to teachers and students. This inference was moderate as it was based on one well-developed study.

Evidence for the evaluation inference was derived from data in phases two and three which included survey data from students and teachers about online vocabulary learning and in-class vocabulary learning and document analysis. Vocabulary test results verified the reliability of specified vocabulary bands and rigor for the specific program levels. Based on triangulated data, member checks, and repeated cycles of analysis, as well as our insider status, we view our work as trustworthy and thus our evaluation inference as moderate, because it is based on one study. Evidence for our explanation inference is moderate as well. The evidence came from using the themes discovered in the evaluation inference to explain the factors affecting implementation and normalization. Our study was designed to draw on our insider status, but it is precisely this insider status that might limit our research as it is impossible to escape personal biases. Insights from more coordinators or coordinators at other schools might have made our data more robust.

In conclusion, all three of our inferences, domain, evaluation, and explanation, were moderate mainly because, although they were based on multiple sources, these sources were from a single case study. As evidence was not gathered for utilization and ramification inference, these parts could not be appraised.

cogent •• education

7. Conclusion

As mentioned in the presenting the argument section, our main objective of this study was to present an argument-based approach to a vocabulary blended language program in a large EFL program. The analysis found that changes made at the meso level did lead to improved use, integration, and vocabulary learning. It is not surprising that the NGSL track of an EFL vocabulary program can be normalized over time; however, how all the different components of vocabulary learning, AOL objectives, and online systems fit into the core curriculum of an English program has not been previously researched and provides new insights to the field of implementing CALL vocabulary programs.

The normalization of blended learning programs seems to be predicated upon careful alignment with well-defined learning objectives and on the availability of transparent analytics from online systems for the learner, teacher, and coordinator to ensure the diffusion of alignment from the meso to the micro level of the program. There is still a lack of awareness on the part of online vocabulary system producers of the multifaceted needs of learners with regard to varying tracks of a language program and the types of lists, processes, and strategies that learners need. The lack of disciplined, transparent AOL descriptions communicated to the industry has created the need for leading universities to blend multiple online systems including in-house systems, resulting in more of a Rube Goldberg machine of multiple sign-on systems than one or two well-designed applications.

The sheer number of alternatives for online study platforms, even within individual platforms, and the number of alternative pricing models can make evaluating the utility of a platform complicated. AI systems that claim to know the needs of learners independent from school administrators perpetuate the Sole Agent fallacy (Bax, 2003), especially when not providing analytics to teachers and users, because school coordinators, teachers, and students are all agents whose own perceptions and needs are bypassed and not validated. Thus, a major point in this article is the need for schools to pursue partnerships with like-minded schools and online companies in order to protect the core functions of their programs.

A second main theme in the evidence is that constructive alignment is necessary for an online system to be normalized. Indeed, it appears that false normalization is a major problem for online vocabulary systems. There is the implication here that universities or consortia of universities which have established clearly aligned objectives could help to represent these needs to CALL companies in order to reduce the barriers to transparency and alignment. At present, lack of knowledge about student needs on the part of universities and the lack of detailed analytics on the part of CALL companies have obfuscated the process of alignment. Other obstacles to constructive alignment are coordinators' and teachers' KSA to the online vocabulary system. Similar to other studies (for example, see Yoon & Gruba, 2019), in this research coordinators' and teachers' KSA are linked to their own willingness and resistance.

The study found that face-to-face assessment of spoken vocabulary in use presents a much harder process to normalize. Our evidence gathering did not clearly explore this and thus spoken vocabulary and written use of vocabulary in meaningful writing appear to be left as future research pathways.

In conclusion, this paper has confirmed the necessity of validating the normalization of vocabulary systems for supporting the practice and testing of second language vocabulary acquisition. In other words, normalization of a blended vocabulary program is a domino effect. The validation process necessarily entails program evaluation, needs analysis, means analysis, and the specification of student learning objectives so that designated platforms for delivery and testing are aligned correctly and supply needed analytics for documenting student learning gains. When EFL program coordinators clearly define the student learning objectives for their program *AND* when online vocabulary learning systems provide transparent analytics, then and only then can constructive alignment be achieved for blended vocabulary learning systems.

cogent · education

Author details
Lindsay Mack[1]
E-mail: lindsay.mack287@gmail.com
ORCID ID: http://orcid.org/0000-0002-4758-0052
Paul Sevigny[1]
E-mail: paul.sevigny@outlook.com
ORCID ID: http://orcid.org/0000-0002-7812-4834
Malcolm Larking[1]
E-mail: larkingm@apu.ac.jp
ORCID ID: http://orcid.org/0000-0001-9502-2648
Lance Stilp[1]
E-mail: ljstilp@gmail.com
ORCID ID: http://orcid.org/0000-0002-7739-6408
[1] Center for Language Education, Ritsumeikan Asia Pacific University, Beppu, Japan.

Data
The data instruments that support the findings of this study are openly available in Mendeley data at https://data.mendeley.com//datasets/cf2b9kvdw6/1

Declaration of Interest and Ethical Statement
No potential conflict of interest was reported by the authors. None of the faculty participants work for any of the CALL companies whose online systems were used or vetted for this study. Additionally, CALL companies were also anonymized. All decisions about data collection were made in conjunction with regular Assurance of Learning (AOL) meetings and were part of a bigger AOL initiative. Data was collected with the permission and approval of the language center at XU (pseudonym); although not an independent ethics committee, it ensured we followed the ethical guidelines set forth by our university that are in line with Taylor and Francis's ethical guidelines. Great care was taken to ensure ethical procedures and guidelines set out by our university were followed for data collection and all permissions were granted. All students who were video recorded signed informed consent forms. All surveys were anonymous, voluntary, and any identifying information has been removed to protect confidentiality. The survey data which we quote specifically in our paper included informed consent. One survey in phase one was not voluntary because it was part of an overall evaluation and needs assessment of student vocabulary learning. However, students could choose whether their response could be used for publication.

Citation information
Cite this article as: Validating the normalization of vocabulary systems in a university EFL program, Lindsay Mack, Paul Sevigny, Malcolm Larking & Lance Stilp, *Cogent Education* (2021), 8: 1985688.

References
Bax, S. (2003). CALL - past, present, and future. *System, 31*(1), 13–28. https://doi.org/10.1016/s0346-251x(02)00071-4
Bax, S. (2011). Normalization revisited: The effective use of technology in language education. *International Journal of Computer-Assisted Language Learning and Teaching, 1*(2), 1–15. https://doi.org/10.4018/ijcallt.2011040101
Bielawaski, L., & Metcalf, D. (2003). *Blended e-learning: Integrating knowledge, performance support, and online learning*. HRD Press.
Biggs, J. (2014). Constructive alignment in university teaching. *HERDSA Review of Higher Education, 1*(1), 5–22. https://www.herdsa.org.au/herdsa-review-higher-education-vol-1/5-22
Bonk, C. J., & Graham, C. R. (Eds.). (2006). *Handbook of blended learning: Global perspectives, local designs*. Pfeiffer Publishing. https://doi.org/10.5465/amle.2008.31413871
Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40. https://doi.org/10.3316/QRJ0902027
Browne, C., Culligan, B., & Phillips, J. (2013). *The new general service list*. http://www.newgeneralservicelist.org
Chapell, C. A., Enright, M. K., & Jameison, J. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice, 29*(1), 3–13. https://doi.org/10.1111/j.1745-3992.2009.00165.x
Chapelle, C. A. (2014). Arguments for technology and language learning. *Keynote presentation at the EuroCALL 2014 Conference*. Groningen, Netherlands.
Chapelle, C. A., Enright, M. K., & Jameieson, J. M. (Eds.). (2008). *Building a validity argument for the test of English as a foreign language*. Rougledge. https://doi.org/10.4324/9780203937891
Creswell, J. W. (2008). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). SAGE Publications, Inc. https://doi.org/10.7748/nr.12.1.82.s2
Fleming, J. (2018). Recognizing and resolving the challenges of being an insider researcher in work-integrated learning. *International Journal of Work-integrated Learning, 19*(3), 311–320. https://www.ijwil.org
Geertz, C. (1973). *The interpretation of cultures: Selected essays*. Basic Books, Inc. https://doi.org/10.2307/2064031
Gleason, J. (2013). An interpretive argument for blended course design. *Foreign Language Annals, 46*(4), 588–609. https://doi.org/10.1111/flan.12050
Golonka, E., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning, 27*, 105–170.
Gruba, P., Cárdenas-Claros, M., Suvorov, R., & Rick, K. (2016). *Blended language program evaluation*. Palgrave Macmillan. https://doi.org/10.1057/9781137514370
Gruba, P., & Nguyen, N. B. C. (2019). Evaluating technology integration in a Vietnamese university language program. *Computer Assisted Language Learning, 32* (5–6), 619–637. https://doi.org/10.1080/09588221.2018.1527365
Kane, M. (2006). Validation. In R. Brennen (Ed.), *Educational measurement* (4th ed., pp. 17–64). Praeger.
Long, M. (Ed.). (2005). *Second language needs analysis (Cambridge applied linguistics)*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511667299
Marsh, D. (2012). *Blended learning: Creating learning opportunities for language learners*. Cambridge University Press.
Norris, J. M. (2016). Language Program Evaluation. *The Modern Language Journal, 100*(S1), 169–189. https://doi.org/10.1111/modl.12307

Norris, J. M., Davis, J. M., Sinicrope, C., & Watanabe, Y. (Eds.). (2009). *Toward useful program evaluation in college foreign language education*. National Foreign Language Resource Center.

Nunan, D., & Bailey, K. M. (2008). *Exploring second language classroom research: A comprehensive guide* (1st ed.). Heinle ELT.

Oxford, R., & Jung, S.-H. (2007). National guidelines for technology integration in TESOL programs: Factors affecting (non)implementation. In M. Peters, K. Murphy-Judy, R. Z. Lavine, & M. A. Kassen (Eds.), *Preparing and developing technology-proficient L2 teachers* (pp. 23–48). Computer Assisted Language Instruction Consortium.

The Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal*, *100*(S1), 19–47. https://doi.org/10.1111/modl.12301

Tosun, S. (2015). The effects of blended learning on EFL students' vocabulary enhancement. *Procedia: Social and Behavioral Sciences*, *199*, 641–647. https://doi.org/10.1016/j.sbspro.2015.07.592

Walvoord, B. E., Carey, A. K., Smith, H. L., Soled, S. W., Way, P. K., & Zorn, D. (2000). *Academic departments: How they work, how they change*. Jossey-Bass.

Yoon, S. J., & Gruba, P. (2019). Evaluating normalization: An argument-based approach. *System*, *83*, 4–12. https://doi.org/10.1016/j.system.2019.04.008

Cogent Education (ISSN: 2331-186X) is published by Cogent OA, part of Taylor & Francis Group.

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com