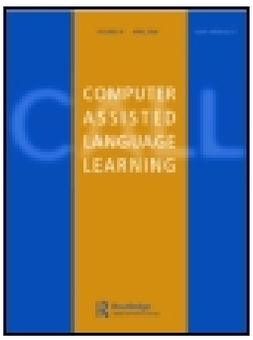

Computer Assisted Language Learning

ISSN: 0958-8221 (Print) 1744-3210 (Online) Journal homepage: https://www.tandfonline.com/loi/ncal20

Evaluating technology integration in a Vietnamese university language program

Paul Gruba & Ngoc Bao Chau Nguyen

To cite this article: Paul Gruba & Ngoc Bao Chau Nguyen (2019): Evaluating technology integration in a Vietnamese university language program, Computer Assisted Language Learning, DOI: 10.1080/09588221.2018.1527365

To link to this article: https://doi.org/10.1080/09588221.2018.1527365

Published online: 20 Jan 2019.

COMPUTER ASSISTED LANGUAGE LEARNING
https://doi.org/10.1080/09588221.2018.1527365

Routledge
Taylor & Francis Group

Evaluating technology integration in a Vietnamese university language program

Paul Gruba[a] and Ngoc Bao Chau Nguyen[b]

[a]School of Languages and Linguistics, University of Melbourne, Parkville, Australia; [b]English Department, Hue University of Foreign Languages, Hue City, Vietnam

ABSTRACT

In the main, evaluation in CALL has focused on the assessment of the individual use of technology in accord with the principles of second language acquisition. Now that CALL focuses on the full integration of technology in language programs, evaluators are urged to take into account a wide range of factors that span institutional policy, departmental culture and classroom practice. In this study, we use an argument-based approach to evaluate technology integration within an English language program at a leading university in Vietnam. Using a layered approach and qualitative techniques to gather data, we examine how administrators and lecturers at the departmental, or meso, level perceive technology integration in an English language program. Results of the study inform technology integration, stakeholder identification, resource allocation and curriculum interpretation, and concludes with a discussion of the utility of blended language program evaluation.

KEYWORDS

Language program evaluation; technology integration; interpretive arguments; Vietnam

Introduction

With CALL now defined as "the full integration of technology in language learning" in ways that "technology, theory, and pedagogy are inseparably interwoven" (Garrett, 2009, pp. 719–720), language professionals have widely adopted blended, or hybrid, approaches to teaching and learning in a range of innovative ways throughout their second language classrooms (Hinkelman, 2018; Sharma & Westbrook, 2016). Given the complexities that such full integration of technology entails (Kern, 2014), CALL educators increasingly face new challenges in ensuring their work is defensible and well-regarded (Chapelle, 2014). Long situated at its core, evaluation is key to advancement of CALL because professionals "want to make sure that the time expended using technology-related

CONTACT Paul Gruba p.gruba@unimelb.edu.au School of Languages and Linguistics, University of Melbourne, Parkville, Australia

language-learning activities is effective and well spent" (Levy & Stockwell, 2006, p. 246). The importance of evaluation was re-iterated in a recent agenda setting article for CALL: to continue our development, Levy, Hubbard, Stockwell, and Colpaert (2015) wrote, we must collectively make a "commitment to rigorous evaluation and testing using the most sophisticated and effective research tools" (p. 5). In this paper, we respond to such commitment for evaluation to meet the demands posed by defining CALL through its work in the full integration of technologies throughout contemporary language programs.

As Chapelle (2007) explains, the focus of evaluation in CALL is best thought along a continuum: at one end, language alone is of concern; towards the middle, there are questions about the optimum mix of language learning and technology; near the other end, there is a sole focus on the affordances of technology. Although much work in CALL evaluation has studied individual interactions with software and tasks (Caws & Heift, 2016), recent work on the evaluation of blended language learning has been conducted with an entire course in mind (Bueno-Alastuey & López Pérez, 2014; Gruba, Cárdenas-Claros, Suvorov, & Rick, 2016). More than ever, the widespread use and complexity of blended language learning demands that CALL professionals forge innovative methods of evaluation (Chapelle, 2014).

In this paper, we seek to evaluate technology integration within the English department at a major university in Vietnam. Of the three levels recently proposed to frame contemporary language learning (The Douglas Fir Group, 2016), our efforts here concern the 'meso', or departmental, level. After a review of blended language program evaluation theory, we follow the four stages of an argument-based approach (Kane, 2006) to assess the strength of claims made about the integration of technology. We conclude with a discussion of issues that arose in our evaluation project and suggest an agenda for further research.

Understanding blended language program evaluation

Language program evaluation can be seen as a "pragmatic mode of enquiry" that "provides a framework, heuristic, and methodologies for making sense of language teaching and learning (and other language-related endeavors) *in situ*, as well as for answering 'what works' questions and informing practical efforts at improvement (Norris, 2016, p. 169, italics original). Much of the previous work in language program evaluators (e.g., Davis, Sinicrope, & Watanabe, 2009; Kiely & Rea-Dickins, 2005; Lynch, 1996, 2003) has not taken technology into account with the exception of Lynch (2000) who promoted a view that CALL

professionals should make evaluation an "ongoing, integral part" (p. 438) of program development. Stakeholders, wrote Lynch, including students, teachers and administrators must be encouraged to participate regularly in the assessment of their efforts if technology was to be effectively integrated in language learning.

Because of the inherent complexity of factors involved in blended language learning, the use of checklists (Susser, 2001) is no longer sufficient to assess the effectiveness of technology; as Chapelle (2014) suggests, CALL professionals must now make use of interpretive arguments for the design of evaluation approaches. Briefly, interpretive arguments seek to establish the strength of claims through an interrogation of the warrants, assumptions and inferences (Kane, 2006). Such arguments have been used for example, to underpin the validity of a language test (Chapelle, Enright, & Jamieson, 2008), as the basis for blended course design (Gleason, 2013), and research in both second language writing (Ranalli, Link, & Chukharev-Hudilainen, 2017) and pragmatics (Youn, 2015).

Following Chapelle (2014), Gruba et al. (2016) developed an argument-based approach to blended language program evaluation. The evaluation of entire blended language learning programs, they proposed, was best understood when it could account for institutional and policy concerns at a macro level, department administration and culture at a meso level, and teacher and student classroom factors at the micro level. As in the transdisciplinary framework proposed by The Douglas Fir Group (2016), a three-level view can help identify ideological structures at the macro level, recognize sociocultural institutions and communities at the meso level, and inform views of social activity at the micro level. As CALL professionals are aware (Hinkelman, 2018), the three levels do not exist independently of each other but rather "persist only through constant interaction with each other and so exist in state of continuous change" (The Douglas Fir Group, 2016, p. 25).

CALL professionals are accustomed to holding a number of positions across the institution as they move amongst varied roles (Kern, 2014; Levy & Stockwell, 2006). In line with the ethos of blending technologies in language programs, it is clear that neither institutions, technology or educators have strictly bounded positions when working across global networks (Sharma & Westbrook, 2016); further, we know that faculty members do not solely occupy instructional roles but move across policy, administrative and teaching responsibilities (Mårtensson, Roxå, & Stensaker, 2014), and it is obvious that national and institutional policies do not pass uninterrupted through an institution to be simply enacted in the classroom (Liddicoat, 2014).

Table 1. Warrants, assumptions and backing associated with each inference in the argument for meso level investigation of technology use at the English Department.

Inference	Warrant supporting the inference	Assumptions underlying the warrant	Examples of backing sought to support assumptions
Ramification	Findings provide insights to the meso level and foster the utilitarian purpose of language program evaluations	Findings are disseminated via journal articles, conference presentations or other appropriate forms Meso level activities concern language program evaluators and researchers	Evaluation findings are published A call for further research in understanding the meso level in language program evaluation
Utilization	Project outcomes can provide stakeholders an improved understanding of their roles regarding technology integration	The findings can be interpreted and have motivational power Stakeholders develop technology integration in the department's language program	Debriefing with participants to report study findings Participants take initiatives to improve the use of technology in the language programs
Explanation	The findings of the study are situated within the specific context of the site of study, and align with stakeholder sociocultural and institutional expectations	The findings can be explained in context of the English Department at the meso level Participants' views can be explained with consideration to departmental roles and structure Participants' views give an insight on the meso level	Analysis of the questionnaire responses, semi-structured interviews, and program syllabi in relation to the program context, the departmental culture and program-level issues relating to technology implementation
Evaluation	Data analysis reveals stakeholder views on technology use and integration	The analyses are conducted rigorously, accurately and ethically The analyses give an accurate representation of the implementation of technology in the department	Research ethics, methods and analyses are followed; inter-coder reliability meets accepted standards Thematic analysis of qualitative data resonates with participate experiences
Domain definition	Data gathered from lecturers in the department provide insights on their roles and factors contributing to the use of technology in teaching and learning at the English Department	Technology integration is a priority, supported by policy and funding, in Vietnamese language programs Meso level sources of data to do with technology integration are available and can be accessed The methods for collecting data are appropriate to form a view of technology integration at the meso level	Reading the Vietnam Foreign Languages 2020 project brief Insider status of the research team at the site of study Review of literature on the use of document analysis, open-ended questionnaires and semi-structured interviews

Adopted from Gruba et al. (2016).

Positioned at the meso level, an academic department can be seen as "a uniquely *autonomous* and *powerful* yet also uniquely *interdependent* structure within a very complex system" (Walvoord et al., 2000, p. 33, italics original). Departments play a crucial role in the adoption and success of blended learning (Garrison & Vaughan, 2013). Given the

centrality of this role, this study seeks to examine the role of the department – that is, where work and interaction amongst faculty takes place as a sociocultural activity – as a locus for a meso level evaluation of technology integration. Following earlier work (Gruba et al., 2016), our evaluation project is structured in four stages: planning an argument, gathering the evidence, presenting the argument, and appraising it.

Planning the argument

In the early phases of a program evaluation, it is important to bring together the evaluation team with key stakeholders to discuss the purpose and scope of the project (Patton, 2008). In an argument-based approach, the task of initial discussions is to plan an argument to map a chain of inferences that links target domain to the final outcomes of the project. The team and stakeholders work together to set out the relevant warrants, assumptions and backing for each of the inferences. Eventually the team must anticipate how each inference is to be crossed, perhaps similar figuring out how to move from one side of a bridge to another (Chapelle, Enright, & Jamieson, 2008), through the constant assessment of the strength of evidence that has been gathered specifically to meet the warrant and its assumptions of a particular inference.

In blended language program evaluation, an argument must be located at one of three levels to determine what type of data is to be gathered, the methodology for data collection, and the analysis of the data itself (Gruba et al., 2016). As a result, project recommendations are tailored to the specific level of concern. To illustrate, Table 1 shows the plan we devised with the evaluation team at the start of the current project.

To explain the planning of this argument, note that technology integration as a sociocultural activity at the meso, or department, level was our primary domain. (Arguments start at a domain definition that is placed at the bottom of a table and move progressively upwards until potential ramifications are developed.) In planning our Domain definition, our warrant for motivating the need for an evaluation rested on the key assumption that the Vietnamese are keen to foster language learning (Nguyen, 2015). Characterized as one of Asia's "most dynamic economies, actively and deeply integrated into the world economy" and actively striving to globalize (Anh, 2017, n.p.), the National Foreign Languages 2020 Project was launched in 2008 to enhance the foreign language education (Nguyen, 2015). 'Project 2020', as it became known, received an initial funding of USD $440 million; nearly 75% of the funding was allocated to the upgrading of technological infrastructure in participating educational institutions (Nguyen, 2014).

In planning our second inference, Evaluation, we drew on the literature and previous research experience to set out criteria for data analysis so that our assessment of findings would ultimately be defensible (Chapelle, 2014). Such planning of both the data gathering and analytical processes helped us to also justify an estimated budget for the project for potential sponsors (Bryson, Patton, & Bowman, 2011; Patton, 2008) and university colleagues as well as members of the Institutional Review Board (IRB) who needed to approve a required human research ethics application.

A third stage of planning an argument-based approach considers the need to explain findings to stakeholders (Gruba et al., 2016). For this project, we began to anticipate the challenge of working with the Vietnamese such that the evaluation findings would be understood, and eventually, accepted and utilized; as well known in program evaluation, any stakeholder suspicion or misunderstanding of outcomes results in the lack of uptake of project recommendations (Lynch, 1996; Patton, 2008). The insider status of Nguyen on the project team fostered sensitivity to the need for clear explanations and transparency of techniques in this study.

In our final stages, we considered ways that our findings may have utility to the language program itself (Lynch, 2000; Morris, 2006); in short, we hoped that discussion in planning the argument would make it more transparent what was specifically being investigated, and that program stakeholders would view our efforts as 'insiders helping insiders' that would foster an uptake of any recommendations (Patton, 2008). With argument planning now completed, we proceeded to gather evidence that could support, or counter, our preliminary set of claims.

Gathering the evidence

On the ground in Vietnam, Nguyen led the evaluation project at the university. As one of five regional institutions that were involved in Project 2020, the university received funding to equip several rooms with workstation computers, smartboards, and educational software. Of note, the university does not manage a central Learning Management System (LMS) or provide email addresses for staff and students.

Although other modern languages are taught, the English has the vast majority of enrolments with approximately 2500 students per year. The English department, broken into four discipline areas, is led by a dean and two deputy-deans. Each discipline area is then managed by a section head and deputy-head. The majority of the 49 lecturers in the department have earned graduate degrees from American, British or Australian universities. The workload for each lecturer consists of teaching six to 12

Table 2. Methods used to gather the evidence.

Method; focal source	Reason	Associated references
Ethnography; Nguyen (lead investigator on the evaluation project)	Allows for an 'insider' view of the meso level; academic departments are little researched, socially complex and require personal trust and connection to investigate	Heigham & Sakui (2009); Hinkelman (2018)
Document analysis; 110 subject guides in the English department	Provides evidence to ascertain the extent to which technology is formally embedded in the curriculum	Bowen (2009); Gruba et al. (2016); Lynch (2000)
Questionnaire; sent to 49 lecturers, received 10 full responses	Provides an opportunity to gather a range of opinions regarding technology integration	Brown (2009); Dörnyei & Taguchi (2010); Phakiti (2014)
Semi-structured interviews; four of the six senior academic administrators	Gives insights into complex decisions regarding leadership and administration of technology initiatives	Brinkmann & Kvale (2015)

classes per semester, and there are 40 to 60 students per class. Nguyen, an insider, reported that each lecturer has a high degree of autonomy; importantly, all faculty members understand that it is their responsibility to prepare students to sit a common exam that is held at the end of the semester.

In line with recent calls to conduct 'layered' investigations (King & Mackey, 2016), we gathered data through ethnography, the collection of departmental documents, an online open-ended questionnaire and semi-structured interviews as shown in Table 2. Done with care and sensitivity (Levy, 2015), a layering of qualitative methods can offer a useful way of exploring perspectives and experiences of program participants (Patton, 2002; Phakiti, 2014).

Ethnography, and the related method of 'participatory action research', has been widely used research concerning technology use and integration in language programs (e.g., Bax, 2011; Hinkelman, 2018; Kawamura, 2006; Lynch, 2000). In this study, Nguyen took on the role of an 'insider' to the department not only to navigate what to ask and whom to talk to about technology integration, but also to share observations as a colleague on possible points of improvement. Later, moving from data collection to analysis, Nguyen shared insights on what particular aspects of evidence meant – such as comments about the overall university leadership – in meso level project.

As Lynch (2003) suggested that document analysis be an entry point to program evaluation, we first collected the subject guides (syllabus) for each of the 110 subjects taught in the English department with the explicit purpose of seeing how technology appeared within the formal curriculum. Using an open-ended questionnaire (Brown, 2009; Dörnyei & Taguchi, 2010) in the form of five short-answer questions, we then asked all English department lecturers to briefly explain how they used technology, viewed the appropriateness of it in their teaching, or thought about the long-term use of technology in the program; of 49 lecturers in the

department, ten of them responded to our questionnaire. Our third set of data came out of individual semi-structured interviews (Brinkmann & Kvale, 2015) with four of the six senior departmental administrators. In these half hour talks, Nguyen elicited views on departmental culture in relation to the effectiveness of technology integration. All of the data was gathered in English.

To begin the process of analysis, we first prepared the data: each subject guide was sorted, all ten questionnaire responses were compiled, and the interview data was transcribed. Our principles of qualitative data analysis were based on work in thematic analysis (Saldaña, 2013). For the subject guides, we followed Bowen (2009) to categorize the documents by level, topic and use of technology. Analysis of the open-ended questionnaires was grounded in work by Brown (2009) as we formed a view of faculty perspectives. As with the transcriptions of the four semi-structured interviews, we coded the questionnaire data for themes that emerged as they related to funding, the uses of technology and view of academic work. Following cycles of discussion and refinement, we achieved an inter-coder reliability agreement of .88 at the conclusion of our analysis (Miles, Huberman, & Saldaña, 2014).

Presenting the argument

In the third stage of an argument-based approach, the results of the data analysis are presented to determine if the evidence is sufficient to meet the assumptions and backing that are needed to support each inference (Chapelle, 2014). To start the process, we began to work from our initial stage, planning the argument, and assessed each of the key sections beginning with the definition of the domain.

Domain definition

In line with the flow of logic in an argument-based approach, our domain definition inference was based on the warrant (or, reason) that the data we had collected in the department provides insights on technology integration. We took into account the three assumptions to back this warrant, starting with the assumption that technology use in language teaching and learning is a focal concern of the Vietnamese government and, by extension, of the university itself. Funded by the Project 2020 initiative, it is clear that the university has devoted significant resources to install the infrastructure that is required to fully integrate technology in language programs.

As we continued, the second assumption we needed to meet was that English department work and insights could be accessed. With Nguyen

was an insider, we are able to meet this assumption through privileged access to program documents and fellow colleagues. A third assumption, based on a need to justify our data collection techniques, was supported by our thorough understanding of ethnography, document analysis, open-ended questionnaires, and semi-structured interviews.

Evaluation inference

As we set out to meet the evaluation inference, we based our reasoning on the point that our data analyses revealed participants' view on departmental factors regarding technology use and implementation. The warrant was based on two assumptions: (1) the analyses were conducted rigorously, accurately and ethically, and (2) the analyses gave an accurate representation of technology integration in the department. In meeting these assumptions, we followed a rigorous series of analyses to produce an acceptable inter-rater reliability indice of .88 that had begun with document analysis.

Analyzing program documents can be a powerful method to reveal important information about the language programs' processes and activities (Bowen, 2009; Kiely & Rea-Dickins, 2005; Lynch, 1996). We gathered syllabi of the subjects offered by the department (PDF file format). Besides the syllabi, there was no official documentation regarding the use of technology in teaching. The deputy dean provided subject descriptions of all undergraduate and postgraduate subjects offered by the department. An analysis of the documents found was no mention of technology use or integration, aside from the focal topic of two subjects (e.g., *Technology for Language Teaching; Computer Assisted Translation*). No documents encouraged the use of technology in teaching and learning, and none did provide guidelines for technology integration.

Overall, document analysis identified no official policies and directives from the higher, or macro level, of the university regarding technology; at the meso level, the analysis of documents showed that the languages curriculum requires no use of technology in teaching and learning aside from the two specialized subjects.

Responses to the questionnaire provided insights on how technology is used at the micro level. Each lecturer reported that there was no guidance on how to integrate technology. They listed CD players, PowerPoint slides and email as their main sources of technology in their teaching. A few teachers made use of video-sharing and news sites, and two teachers used social media sites to interact with students. No technology was used for assessment purposes beyond simple spreadsheets for record keeping.

Table 3. Coding themes, definitions, and sample data.

Theme/subtheme	Working definition	Sample data
Funding and support	The use of funds and other resources to promote technology	*"I don't know where the money comes from. I don't know who was involved in the process of deciding what to buy … And that is public property and lecturers and students do not have any rights to decide or use it their own way"* (Interview 5, p. 2)
Curriculum and assessment	The role of curriculum and assessment designs in the use of technology	*We have a committee of senior lecturers. They are the experts in deciding the mainstream methods of teaching many courses in the department.* (Interview 1, p. 7)
Individual expertise power	The influence of individuals with expertise on other lecturers' use of technology	*There is one, that is Mr. X. He really likes technology and he tries to help everybody with it. But he will retire in two months and I'm really worried about that.* (Interview 2, p. 3)
Departmental leadership	The influence of meso leaders' actions on the effectiveness of technology integration	
Multiple roles	Meso level leaders' identification with the roles of their positions	*As a lecturer who directly teaches the students … From the date that I work as a vice dean, I have had a lot of meetings with the superior level, as well as with other universities and foreign experts … And in our department, at this level, we also have some kind of discussion with other teachers …* (Interview 1, p. 1)
Power	Meso level leaders' regard and exertion of their power	*Oh but no, we don't have that kind of power. We are asked to do things. We are not invited to say "here, you work with us as a team … ". So if I had the power to do that, just to voice myself, …* (Interview 5, p. 5)
Responsibility	Meso level leaders' views and ownership of their responsibilities	*Interviewer: How do you think we can encourage these teachers (of the older generation) to use technology? Interviewee: I don't know, ask Mr. Y (department dean). We are going to retire and so why change now?* (Interview 4, p. 6)

In line with findings from Egbert (2010), lecturers reported that professional development and student low digital literacy were prominent challenges. Although technology was available (i.e., classrooms were equipped with a computer, speakers and projectors), teachers needed further training to integrate such resources. Low student digital literacy, including problems with typing in English and navigating search results, was also seen as a barrier to language learning. Further analysis identified four themes that identify reasons for poor technology integration (Table 3).

Explanation inference

The explanation inference was based on the warrant that the findings of the study align with the context of the English department. Underlying this warrant are assumptions that (a) findings can be explained in

relation to the context of the English department and that (b) participants' views can be explained with consideration to their departmental roles, and relationships, and (c) the participants' views provide insights to work at the meso level. As an insider, Nguyen ensured that the analyses reflected a deep understanding of the program context, the departmental culture and issues relating to technology integration.

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. Because of the fluid nature of interaction amongst levels (Gruba et al., 2016; The Douglas Fir Group, 2016), it is not surprising then to see that that teacher willingness or resistance to using technology *in their classrooms* is greatly affected by the departmental and sectional leadership. For CALL professionals, this is an important point: as noted by Chapelle (2007), our field must attend to a level above the classroom if we are to better understand how teachers make decisions regarding technology use and integration.

Not surprisingly, our analysis also revealed variation amongst administrator views of their roles and responsibilities. In this case, the deputy-dean showed a high level of responsibility, but the section leaders did not indicate they felt a need to have a similar level of accountability. Given this point, we would argue that issues of technology integration in language programs may arise when responsibilities are not evenly distributed. In this setting, for example, the deputy-dean managed a major university project and taught four to six classes per semester as is expected by senior departmental administrators. In their roles, however, the section leaders were not involved in university projects and thus felt themselves to be powerless and unable to contribute to macro level decisions. In this university, technology integration is seen to be a top-down process that does not need to recognize insights grounded in meso and micro level experiences, and arguably such lack of consultation would result in minimal uptake of Project 2020 goals.

Utilization inference

The purpose of the utilization inference is to link any claims about the local program improvement to the rationale; for our work, the key warrant here is to provide stakeholders of meso level challenges facing technology integration within English language program. The assumptions that support this warrant include that (a) the findings are understood, and that (b) meso level stakeholders are motivated to use the evaluation outcomes. In the brief life span of this project, it was not possible to collect the long-term evidence needed to check assumptions underpinning

utilization. We can conclude that meso level stakeholders, the departmental lecturers, are keen to improve technology integration but a continued lack of consultation with macro level leaders would likely thwart any uptake of recommendations.

Ramification inference

The final inference in the argument-based approach, ramification, links claims about implications suited to the local context to their broader contribution in the field. For this study, the key warrant for this inference is that results of the present study contribute to a better understanding of programs at the meso level, thus increasing the utilitarian purpose of research in language program evaluation. The assumptions backing this warrant are that (a) the findings are disseminated via journal articles, conference presentations or other appropriate forums, and (b) the meso level is an area of concern for language program evaluators and researchers. Here, in line calls for further research in language program evaluation (Norris, 2016) with a focus on technology integration in particular (Chapelle, 2014; Gruba et al., 2016; Levy et al., 2015), this study may foster the further development of evaluation in CALL.

Appraising the argument

The final step in the argument-based approach seeks to determine the strength of claims, clarity of the argument and whether or not inferences and assumptions are plausible (Chapelle, 2014). Rebuttals, or counterarguments, to any outcomes can also be made in this final stage. In line with Golonka, Bowles, Frank, Richardson, and Freynik (2014), judgments were based on criteria for strong, moderate and weak arguments in CALL. In this view, an argument can be considered strong if it based on evidence from multiple well-designed studies; it is moderate if evidence comes from a single well-designed study, for example; an argument is weak if the evidence is derived from a flawed study or non-empirical data such as anecdotes (Golonka et al., 2014). Here, we have evidence for the first three inferences: domain definition, evaluation and explanation; with this in mind, the appraisal focuses solely on these three inferences.

Evidence for the first inference, the domain definition, rests on our in-depth understanding of the English language programs in Vietnamese higher education experience with methods of document analysis, open-ended questionnaires and semi-structured interviews. The claim at this point is strong in that data collection and analyses were robust and

involved multiple sources (Patton, 2008; Youn, 2015); a possible rebuttal, however, is that the small sample of data used in this evaluation does not represent the meso level sufficiently.

We believe that the strength of the evaluation inference, seeking to reveal perceptions of meso level participants on their roles, power relations, responsibilities, was moderate. Our perception of this strength comes from an inter-coder reliability coefficient (.88), but is weakened as it is based within this single case study.

The explanation inference is partially supported and thus moderate. The warrant for the explanation inference was that findings can be explained in the local context, and our study was designed to draw on the insider status of Nguyen. Aside from sporadic mentions of the Project 2020, the evaluation lacks the wider 'macro' discussion of the Vietnamese sociocultural context. As noted, an appraisal of the utilization and ramification are premature in light of the short-term nature of this study.

To summarize, the argument-based approach to the evaluation of technology integration in this university English department has found there to be a strong warrant for CALL, and the domain inference has sufficient evidence to support a strong claim that the institution has prepared for integration. However, despite a second claim that data gathering was sufficiently strong, inferences concerning evaluation were moderate. This point, as well as a need to enhance the utilization of any outcomes, could become the focus of efforts to strengthen and improve the English language program.

Limitations and discussion

As with many small-scale studies, this evaluation study has limitations: the project, despite benefitting from the insider status of Nguyen, was completed in relatively short time that would have benefitted from a longer time that is recommended for language program evaluations (Lynch, 2003). One shortcoming of a short-term project is an inability to engage stakeholders in such ways that eventually promote the greater utility of evaluation outcomes (Patton, 2008). Although we are confident that our layered approach and qualitative methods are sound, critics may point out that quantitative data including test results, quality of teaching scores, and wide-scale survey would help to bolster the evidence needed to ground conclusions (Norris, 2016).

Potentially, as Norris (2016) concludes, work in language program evaluation may well "provide a much needed and heretofore missing epistemological space to applied linguists for deliberating, understanding,

and (critically) acting upon the specific problems of language learning" (p. 184). Traditionally, CALL specialists have been hesitant to engage in evaluating entire language programs in preference for focused work on applications (Chapelle, 2014); with CALL now encompassing the full integration of technology in language learning (Garrett, 2009), professionals will need new tools to be able to justify ways in which technology promotes improved language learning outcomes (Chapelle, 2007). Further work could extend the adoption and utility of an argument-based approach (Chapelle, 2014), provide a greater understanding of stakeholder perspectives (Bryson et al., 2011), and help ensure a stronger commitment to CALL evaluation (Caws & Heift, 2016; Levy et al., 2015).

The results of this study may inform language program evaluation. Traditional procedures of determining which stakeholders to include in the evaluation process have relied on array of techniques and toolkits such as power versus interest grid (e.g., Bryson et al., 2011). Historically, the view of the meso level was as 'the level in between' (Liljenström & Svedin, 2005) that evoked an image of a middle-layer where individuals at this level were constantly involved mediating/translating/intermediating between the macro and micro levels. Contemporary studies now suggest that that earlier views did not fully capture the intricacies and fluidity of the meso level (The Douglas Fir Group, 2016). However, without an adequate recognition of the multiple roles that stakeholders often hold, there is a risk that any static view of stakeholders ignores the fact that individuals work in a fluid manner across levels as they participate in policy making, administration and teaching. Further attention is needed to discern how stakeholders identify with specific roles and thus affect technology integration, and a sensitivity to multiple stakeholder roles may help identify factors that influence evaluation.

Our study further motivates the need for professional development. For language program administrators, stakeholder analysis is crucial in the allocation of resources in an efficient manner (Kiely & Rea-Dickins, 2005). As this study points out, further understanding of the sociocultural activity at the meso, or departmental, level may sharpen decisions regarding technology integration. CALL specialists, often themselves working in the administration of language departments, may continue to benefit from management and leadership training in areas of communication, workload balance and changes in technology (Gimeno-Sanz, 2016; Kern, 2014). Potentially, CALL program administrators may require additional sensitivity regarding the influence of their roles.

Based on our work, we would suggest that two central orientations exist at the meso level: one is shaped by a formal view of the hierarchy amongst department and section heads, and the other can be understood

in seeing how individuals may identify with their leadership position and level of expertise. If the individual is given responsibility or is knowledgeable about the area of concern, he/she can be empowered to identify more strongly with the leadership position. The leadership of a CALL program, we believe, may benefit by taking into consideration a stronger awareness of the intertwined network of factors.

Towards an agenda for future research

With CALL evaluation now more focused on technology integration than human-computer interaction (Chapelle, 2014), future investigations may seek to identify the characteristics of the different levels to determine the multiple influences that influence language program success. Perhaps keeping in mind that the "most important lesson" for the continued development of our field is that " … language professionals need to consider their contributions at a level that goes beyond each person's own classroom" (Chapelle, 2001b, p. 12), CALL evaluation professionals may build their projects in ways that can better account for influences rather than seeing language programs as a unitary whole. A three-level framing directs attention to important influences in language programs that range from national policy through departmental responsibility to lesson planning (Gruba et al., 2016). More sophisticated understanding of the levels, particularly to do with the oft-neglected meso level, may lead to a greater utility of outcomes across the entirety of an institution. Indeed, future work in understanding the three levels can inform transdisciplinary CALL within the wider field of applied linguistics, such as language testing or curriculum design (Norris, 2016; The Douglas Fir Group, 2016).

With a focus on meso level influences, future studies may seek to extend work concerning the interactions within academic departments (Walvoord et al., 2000): at present, we know little about the ways colleagues may influence technology integration (Bax, 2011). Methodologies could be based on ethnographic, or auto-ethnographic, approaches (Heigham & Sakui, 2009; Hinkelman, 2018; Kawamura, 2006) that could be then coupled with surveys and interviews in line with recent suggestions to promote layered approaches in applied linguistics research (King & Mackey, 2016).

Research at the macro level, perhaps motivated by a close look at how work such as the TESOL Technology Standards may impact language programs, is also an area rich for further evaluation. Because these Standards specifically seek to promote "pedagogically solid ways of integrating and using technology in teaching methods" (Healey et al., 2008,

p.17), they clearly align with a contemporary definition of CALL (Garrett, 2009). Although difficult to construct and often greeted with suspicion, standards are part of a growing emphasis on competency-based education that require further "healthy discussions" (Kessler, 2016, p. 65) amongst teacher educators and CALL professionals to be developed. Macro-level evaluation, in that they inform such discussions, could contribute to the development of CALL as a distinct field of study.

Past efforts to investigate CALL in limited technology contexts (Egbert, 2010) can be extended to include language program evaluation. Further studies in developing countries, for example, would inform work in language program evaluation research that has largely been situated in British, North American and Australian settings (e.g., Davis et al., 2009; Kiely & Rea-Dickins, 2005; Lynch, 1996, 2000). If our experiences in Vietnam may serve as a guide, CALL evaluation studies that are grounded in the argument-based approach may also help to benefit stakeholders who must often defend their decisions to integrate technology (Gruba et al., 2016).

To conclude, this study was motivated by a need to reconsider how we evaluate our technology-rich language program (Caws & Heift, 2016; Norris, 2016) in light of CALL being defined as the "full integration of technology in language teaching" (Garrett, 2009, p. 719). Further, the study refined our conceptualization of language departments; once seen as merely a transit point, it is better to see them as key sites of attention. The failure to understand the complex nature of meso level influences may lead to the poor uptake of recommendations for program improvement (Patton, 2008). Attention to the meso level as core site of sociocultural interaction for research in a multilingual world (The Douglas Fir Group, 2016) further heightens the importance of developing new approaches in CALL evaluation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Paul Gruba is an associate professor in applied linguistics in the School of Languages and Linguistics, University of Melbourne, Australia. He researches language program evaluation, multimodality, and second language listening.

Ngoc Bao Chau Nguyen is a lecturer in the Department of English, Hue University of Foreign Languages, Vietnam. Her research interests include the role of technology in language teaching, assessment and language program evaluation.

ORCID

Paul Gruba http://orcid.org/0000-0002-6616-9568
Ngoc Bao Chau Nguyen http://orcid.org/0000-0002-4326-8563

References

Anh, M. (2017). *Vietnam economy overcomes challenges*. Retrieved from http://vccinews.com/news_detail.asp?news_id=34272

Bax, S. (2011). Normalisation revisited: The effective use of technology in language education. *International Journal of Computer-Assisted Language Learning and Teaching*, *1*, 1–15.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*, 27–40. doi:10.3316/QRJ0902027

Brinkmann, S., & Kvale, S. (2015). *InterViews: Learning the craft of qualitative research interviewing* (3rd ed.). Thousand Oaks, CA: Sage.

Brown, J. D. (2009). Open-response items in questionnaires. In J. Heigham & R. A. Croker (Eds.), *Qualitative research in applied linguistics* (pp. 200–219). New York: Palgrave Macmillan.

Bueno-Alastuey, M. C., & López Pérez, M. V. (2014). Evaluation of a blended learning language course: Students' perceptions of appropriateness for the development of skills and language areas. *Computer Assisted Language Learning*, *27*, 509–527. doi: 10.1080/09588221.2013.770037

Bryson, J. M., Patton, M. Q., & Bowman, R. A. (2011). Working with evaluation stakeholders: A rationale, step-wise approach and toolkit. *Evaluation and Program Planning*, *34*, 1–12.

Caws, C., & Heift, T. (2016). Evaluation in CALL: Tools, interactions, outcomes. In L. Murray, & F. Farr (Eds.), *Handbook of language learning and technology*, (pp. 127–140). New York: Routledge.

Chapelle, C. A. (2001a). *Computer applications in second language acquisition*. Cambridge: Cambridge University Press.

Chapelle, C. (2001b). Innovative language learning: Achieving the vision. *ReCALL*, *13*, 3–14.

Chapelle, C. A. (2007). Challenges in the evaluation of innovation: Observations from technology research. *Innovation in Language Learning and Teaching*, *1*, 30–45.

Chapelle, C. A. (2014). Arguments for technology and language learning. Keynote presentation at the EUROCALL 2014 Conference. Groningen, the Netherlands. Retrieved from http://www.eurocall2014.nl/

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). *Building a validity argument for the Test of English as a Foreign Language*^{TM}. New York: Routledge.

Davis, J. M., Sinicrope, C., & Watanabe, Y. (2009). College foreign language program evaluation: Current practice, future directions. In J. 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 Hawai'i, National Foreign Language Resource Center.

Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing*. New York: Routledge.

Egbert, J. (Ed.) (2010). *CALL in limited technology contexts*. San Marcos, TX: CALICO.

Garrett, N. (2009). Computer-Assisted Language Learning trends and issues revisited: Integrating innovation. *The Modern Language Journal*, *93*, 719–740. doi:10.1111/j.1540-4781.2009.00969.x

Garrison, D. R., & Vaughan, N. D. (2013). Institutional change and leadership associated with blended learning innovation: Two case studies. *The Internet and Higher Education*, *18*, 24–28. doi:10.1016/j.iheduc.2012.09.001

Gleason, J. (2013). An interpretive argument for blended course design. *Foreign Language Annals*, *46*, 588–609.

Golonka, E. M., 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*, 70–105. doi:10.1080/09588221.2012.700315

Gimeno-Sanz, A. (2016). Moving a step further from "integrative CALL". What's to come? *Computer Assisted Language Learning*, *29*, 1102–1115. doi:10.1080/09588221.2015.1103271

Gruba, P., Cárdenas-Claros, M., Suvorov, R., & Rick, K. (2016). *Blended language program evaluation*. New York: Palgrave Macmillan.

Healey, D., Hegelheimer, V., Hubbard, P., Ioannou-Georgiou, S., Kessler, G., & Ware, P. (2008). *TESOL technology standards framework*. Alexandria, VA: TESOL.

Heigham, J., & Sakui, K. (2009). Ethnography. In J. Heigham & R. A. Croker (Eds.), *Qualitative research in applied linguistics: A practical introduction* (pp. 91–111). Basingstoke, UK: Palgrave Mcmillan.

Hinkelman, D. (2018). *Blending technologies in second language classrooms* (2nd ed.). New York: Palgrave Macmillan.

Kane, M. (2006). Validation. In R. Brennen (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: Praeger.

Kawamura, H. (2006). Program evaluation as ethnography. In D. Birckbichler (Ed.), *Evaluating foreign language programs* (pp. 15–28). Columbus, OH: The Ohio State University.

Kern, R. (2014). Technology as *pharmakon*: The promise and perils of the internet for foreign language education. *The Modern Language Journal*, *98*, 340–357. doi:10.1111/j.1540-4781.2014.12065.x

Kessler, G. (2016). Technology standards for language teacher preparation. In F. Farr & L. Murray (Eds.), *The Routledge Handbook of Language Learning and Technology* (pp. 57–70). London: Routledge Taylor & Francis Group.

Kiely, R., & Rea-Dickins, P. (2005). *Program evaluation in language education*. New York: Palgrave Macmillan.

King, K. A., & Mackey, A. (2016). Research methodology in second language studies: Trends, concerns, and new directions. *The Modern Language Journal*, *100*, 209–227. doi:10.1111/modl.12309

Levy, M. (2015). The role of qualitative approaches to research in CALL contexts: Closing in on the learner's experience. *CALICO Journal*, *32*, 554–568.

Levy, M., Hubbard, P., Stockwell, G., & Colpaert, J. (2015). Research challenges in CALL. *Computer Assisted Language Learning*, *28*, 1–6. doi:10.1080/09588221.2014.987035

Levy, M., & Stockwell, G. (2006). *CALL dimensions: Options and issues in computer-assisted language learning*. Mahwah, NJ: Lawrence Erlbaum.

Liddicoat, A. J. (2014). The interface between macro and micro-level language policy and the place of language pedagogies. *International Journal of Pedagogies and Learning*, *9*, 118–129.

Liljenström, H., & Svedin, U. (2005). *Micro, meso, macro: Addressing complex systems couplings*. River Edge, NJ: World Scientific.

Lynch, B. K. (1996). *Language program evaluation: Theory and practice*. Cambridge: Cambridge University Press.

Lynch, B. (2000). Evaluating a project-oriented CALL innovation. *Computer Assisted Language Learning*, *13*, 417–440. doi:10.1076/0958-8221(200012)13:4-5;1-E;FT417

Lynch, B. K. (2003). *Language assessment and programme evaluation*. Edinburgh: Edinburgh University Press.

Mårtensson, K., Roxå, T., & Stensaker, B. (2014). From quality assurance to quality practices: An investigation of strong microcultures in teaching and learning. *Studies in Higher Education*, *39*, 534–545. doi:10.1080/03075079.2012.709493

Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.

Morris, M. (2006). Addressing the challenges of program evaluation: One department's experience after two years. *The Modern Language Journal*, *90*, 585–588.

Nguyen, N. H. (2015). Vietnam's national Foreign Language 2020 Project: Challenges, opportunities, and solutions. In Bigalke, T. W. & Sharbawi, S. (Eds.), *English for ASEAN integration: Policies and practices in the region*. Bandar Seri Begawan: Universiti Brunei Darussalam.

Nguyen, T. (2014). Vietnam to revise doomed multi-million-dollar foreign language project. Retrieved from http://www.thanhniennews.com/education-youth/vietnam-to-revise-doomed-multimillion-dollar-foreign-language-project-27201.html

Norris, J. M. (2016). Language program evaluation. *The Modern Language Journal*, *100*, 169–189. doi:10.1111/modl.12307

Pardo-Ballester, C. (2012). CALL evaluation: Students' perception and use of LoMasTV. *CALICO Journal*, *29*, 532–547.

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.

Phakiti, A. (2014). *Experimental research methods in language learning*. London: Bloomsbury.

Porter, W., & Graham, C. R. (2016). Institutional drivers and barriers to faculty adoption of blended learning in higher education. *British Journal of Educational Technology*, *47*, 748–762. doi:10.1111/bjet.12269

Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: Investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, *37*, 8. doi:10.1080/01443410.2015.1136407

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). Los Angeles: Sage.

Sharma, P., & Westbrook, K. (2016). Online and blended language learning. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning and technology* (pp. 320–333). New York: Routledge.

Smith, N. L. (2015). Using action design research to research and develop evaluation practice. *New Directions for Evaluation*, *2015*, 57–72. doi:10.1002/ev.20157

Susser, B. (2001). A defense of checklists for courseware evaluation. *ReCALL*, *13*, 261–276.

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

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*. San Francisco: Jossey-Bass.

Youn, S. J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing*, *32*, 199–225. doi:10.1177/0265532214557113