

Blended Language Program Evaluation

Paul Gruba

Mónica S. Cárdenas-Claros

Ruslan Suvorov

Katherine Rick



Blended Language Program Evaluation

This page intentionally left blank

Blended Language Program Evaluation

Paul Gruba

University of Melbourne, Australia

Mónica S. Cárdenas-Claros

Pontificia Universidad Católica de Valparaíso, Chile

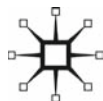
Ruslan Suvorov

University of Hawai'i at Mānoa, USA

Katherine Rick

Lincoln College International, Saudi Arabia

palgrave
macmillan



BLENDED LANGUAGE PROGRAM EVALUATION

Copyright © Paul Gruba, Mónica S. Cárdenas-Claros, Ruslan Suvorov, and Katherine Rick, 2016.

Foreword © Carol A. Chapelle, 2016.

Softcover reprint of the hardcover 1st edition 2016 978-1-137-51436-3

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No portion of this publication may be reproduced, copied or transmitted save with written permission. In accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, Saffron House, 6–10 Kirby Street, London EC1N 8TS.

Any person who does any unauthorized act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

First published 2016 by
PALGRAVE MACMILLAN

The authors have asserted their rights to be identified as the authors of this work in accordance with the Copyright, Designs and Patents Act 1988.

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of Nature America, Inc., One New York Plaza, Suite 4500 New York, NY 10004–1562.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

ISBN 978–1–349–70304–3

E-PDF ISBN: 978–1–137–51437–0

DOI: 10.1057/9781137514370

Distribution in the UK, Europe and the rest of the world is by Palgrave Macmillan®, a division of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

A catalog record for this book is available from the Library of Congress

A catalogue record for the book is available from the British Library

Library of Congress Cataloging-in-Publication Data

Names: Gruba, Paul, author. | Cárdenas-Claros, Mónica, author. | Suvorov, Ruslan, author. | Rick, Katherine, author.

Title: Blended language program evaluation / Paul Gruba, Monica Cardenas-Claros, Ruslan Suvorov, Katherine Rick.

Description: Houndmills, Basingstoke, Hampshire ; New York : Palgrave Macmillan, 2016. | Includes bibliographical references and index.

Identifiers: LCCN 2015038911 |

Subjects: LCSH: Languages, Modern – Study and teaching (Higher) | Blended learning. | Language and education. | BISAC: EDUCATION / Bilingual Education. | EDUCATION / Teaching Methods & Materials / Science & Technology. | FOREIGN LANGUAGE STUDY / English as a Second Language. | LANGUAGE ARTS & DISCIPLINES / Study & Teaching.

Classification: LCC PB35 .B424 2016 | DDC 418.0071/1—dc23

LC record available at <http://lcn.loc.gov/2015038911>

To our families

This page intentionally left blank

Contents

<i>List of Figures</i>	viii
<i>List of Tables</i>	ix
<i>Foreword</i>	xi
Carol A. Chapelle	
<i>Preface</i>	xiii
<i>Acknowledgments</i>	xix
1 Situating Blended Language Program Evaluation	1
2 Essentials of Blended Language Program Evaluation: Towards an Argument-Based Approach	20
3 Micro-Level Evaluation	46
4 Meso-Level Evaluation	65
5 Macro-Level Evaluation	85
6 A Micro-Level Case Study in Chile	105
7 A Meso-Level Case Study in the USA	135
8 A Meso-Level Case Study in Vietnam	164
9 A Macro-Level Case Study in Australia	192
10 Reflecting on Blended Language Program Evaluation	213
<i>References</i>	226
<i>Name Index</i>	241
<i>Subject Index</i>	246

List of Figures

2.1	An illustration of claims and inferences in the structure of an argument for blended language program evaluation	40
3.1	Inferences and claims in a sample argument at the micro level	52
4.1	Inferences and claims in a sample argument at the meso level	73
5.1	Inferences and claims in a sample argument at the macro level	94
6.1	Inferences and claims in the argument for the micro-level evaluation of blended learning in the ELT program at PUCV	110
7.1	Inferences and claims in the argument for the meso-level evaluation of blended learning in the ELI	142
8.1	Stakeholder influence and interest grid	170
8.2	The structure of the argument for the meso-level evaluation of blended learning in AEP	172
9.1	Inferences, warrants, and assumptions in the argument for macro-level evaluation at the University of Melbourne	199

List of Tables

1.1	Blended learning implementation categories	6
1.2	Stages of institutional adoption of blended learning	7
2.1	Proposed differential characteristics of micro, meso, and macro levels	27
2.2	Example variations of level characteristics by project scope	29
2.3	Example goals in blended program evaluation by level	30
2.4	Preliminary considerations in blended language learning	30
2.5	Sample questions for initiating discussion about blended learning evaluation	31
2.6	Sample blended program evaluation goals by consideration at the meso level	32
2.7	Sample vertical case study matrix design for blended language evaluation	33
2.8	Four stages in argument-based blended language program evaluation	44
3.1	Dimensions of technology use at the micro level	48
3.2	Discussion questions for a micro-level evaluation at the institutional level	50
3.3	Idealized goals at the micro level	51
3.4	Ethical considerations for micro-level evaluation projects	53
3.5	Ethical considerations in processing and analysis of qualitative data	62
4.1	Questions for discussion in meso-level evaluation projects	70
4.2	Idealized goals at the meso level	72
4.3	Ethical considerations at the meso level	75
4.4	Participation roles in a meso-level evaluation	76
4.5	A structure and example of pedagogical claims analysis	78
4.6	Example of a pedagogical claim analysis	79
4.7	Idealized curriculum for a blended four-year language program	80
4.8	Prospective evaluation questions in a focus group discussion	82
5.1	Possible claims and questions to initiate macro-level evaluations	92
5.2	Ideal goals against four considerations at the macro level	93

5.3	Ethical considerations at the macro level	96
5.4	Document types and purpose	98
5.5	Framework to analyze documents as a resource	100
6.1	Summary of the warrants, assumptions, and backing for the micro-level evaluation of the ELT program at PUCV	111
6.2	Teacher participants	115
6.3	Theme 1: Contextual factors	121
6.4	Pedagogical actions from classroom observations	122
6.5	Results from the students' questionnaire (n = 41)	124
6.6	Theme 2: Teacher factors	125
6.7	Theme 3: Student factors as perceived by teachers	128
6.8	Sources of data for analyses	129
7.1	ELI courses offered in Spring 2015	139
7.2	Summary of the warrants, assumptions, and backing for the meso-level evaluation of the ELI blended language learning	143
7.3	Summary of the ELI documentation used in the study	148
7.4	Length of semi-structured interviews and transcribed written corpus	149
7.5	Descriptive statistics for the ELI students' responses to Question 6 from the online survey (n = 34)	152
7.6	Reported types of technology used in the ELI classes and percentage of class time when technology is used	154
7.7	ELI instructors' perceived purposes of using technology	155
8.1	Primary and other interested stakeholders	171
8.2	Warrants, assumptions, and backing associated with each inference in the argument for the meso-level evaluation of blended language learning in AEP	174
8.3	Thematic analysis of interview data	179
8.4	Intercoder reliability percentage of agreement	180
8.5	Subthemes of strategic planning, pedagogy, and student engagement	180
8.6	Subthemes of professional development and technology	183
9.1	Clarification of macro-level documents	201
9.2	Representative documents for macro-level evaluation	202
9.3	Themes of blended learning purpose in university documents	205

Foreword

Blended learning has clearly emerged as the most promising use of technology for second language learning. Blended language learning provides students with both the human support of the classroom context and the opportunities for engagement with carefully planned learning materials and social communication tasks. But blended language learning can refer to many different configurations of classrooms, learning materials and human participants. Moreover, program and course constraints, as well as teachers' knowledge of blended learning, all come into play in affecting the way that blended language learning is carried out. In this environment of opportunity for learners and options for the design of curriculum and materials, the need has never been greater for the field to embrace the challenge of conducting evaluations of technology for language learning in use. This book is therefore a welcome exploration of how such evaluations can be accomplished. Moreover, like any good work in applied linguistics, it starts with a real-world problem, draws upon the relevant theory and experience to forge a way forward, tries out a principled approach, and reflects on the meaning of the experience in a manner that invites further contributions. Paul Gruba, Mónica Cárdenas-Claros, Ruslan Suvorov, and Katherine Rick have each crafted their contribution to the field in a manner that explains and illustrates each of these aspects of applied linguistics work.

Questions from the field about blended second language programs and courses are many and multi-faceted. Evaluation methods need to be responsive to the people asking the questions in the contexts where they arise. It is obvious to most stakeholders that the difficult questions cannot be addressed with laboratory-style control-treatment designs. What has been less clear is exactly how to evaluate the innovations in teaching and learning. The journals on technology and language learning are full of research articles examining particular technologies using a variety of research methods. But how can such research methods and findings inform the design and conduct of evaluation for real programs?

The authors have provided an extended response to this pertinent question drawing upon argument-based program evaluation. They demonstrate how such an evaluation approach can work in real

language programs where evaluations are needed at multiple levels. The case studies they present illustrate argument-based evaluation in action at three levels of program evaluation. The case studies are important in showing the variety of forms that argument-based evaluation can take depending on the needs of the stakeholders. Readers can therefore grasp the critical character of argument-based evaluation – that it is a way of framing an evaluation in a context-appropriate manner rather than a template for checking off evaluation criteria. Evaluators are key to the design of their own argument-based evaluations, but with the background, tools, and examples provided by the authors, evaluators have much more to work with than they did in the past.

Carol A. Chapelle
Distinguished Professor, Iowa State University

Preface

Throughout a range of educational settings, the push to integrate technologies in face-to-face environments in ways that promote a ‘blended approach’ to teaching and learning is now commonplace. Despite the widespread adoption of blended approaches, many of us still struggle to justify and evaluate the integrated use of technologies in our language programs. Concerns about the effectiveness of blended approaches in modern language programs touch on a range of issues. An increased use of technologies in language programs can shift pedagogical approaches, spur the need for greater faculty collaboration, and stress the use of policies to explain program outcomes. Blended language programs are complex, and part of the challenge in understanding such complexity arises out of understanding the wide range of factors that influence technology integration and, indeed, what techniques could be used to evaluate the many activities that contribute to successful modern language programs. To meet some of these many challenges, the aim of this book is to set out an argument-based framework for blended language program evaluation.

Throughout our framing, we maintain the view that evaluation projects are pragmatic undertakings designed to help develop and improve language programs (Kiely & Rea-Dickins, 2005; Lynch, 2003; Norris, Davis, Sinicrope, & Watanabe, 2009). Our effort to build a framework here rests on four main concepts: (a) blended language programs are influenced by actions across levels of an institution – micro, meso, and macro – that must take into account considerations of purpose, appropriateness, multimodality, and sustainability (Gruba & Hinkelman, 2012); (b) arguments can be used as a basis for blended language program evaluation (Chapelle, 2014); (c) blended learning is an emerging approach that can be investigated through action research designs that rely primarily on qualitative measures (Bax, 2011); and (d) ethical evaluation projects keep people at the core, seek to develop capacity, and strive to enhance the utility of outcomes for key evaluation stakeholders (Norris et al., 2009; Patton, 2011). To be clear, the intent of our work here is to assist evaluators in ensuring that their *evaluation findings are defensible*, rather than helping language program administrators to *defend the effectiveness of their program*. The structure of our work is

loosely based on the four-phase cycle of argument-based evaluation put forward by Chapelle (2014): planning an evaluation as argument, gathering evidence, presenting an argument, and appraising the argument.

Planning an evaluation

In the first part of this book, we have two tasks: (a) to situate blended language program evaluation across a range of disciplinary influences (Chapter 1), and (b) to propose a framework for program evaluation that places the use of arguments at its core (Chapter 2). The two tasks present challenges.

Blended learning, or the integrated and principled use of technologies in face-to-face settings, is an emerging approach in a number of modern language programs. Because it is emergent, there has been relatively little work done on exploring how integration does actually take place and, indeed, if integration of technologies actually helps to develop and improve a language program.

Distinct from approaches that require the close inspection of human–computer interactions to evaluate the *effect* of technologies in language learning, work on the evaluation of the *use* of technologies in language programs not as fully developed. Arguably, few guidelines exist to date that can assist with the evaluation of the integration, or blending, of technologies in language programs. In addition, few books have appeared in recent years with a focus on with language program evaluation; of these, none have made technology integration a prominent factor in their work. Our work in Chapters 1 and 2 seeks to address contemporary challenges.

In Chapter 1, we review work in three areas: blended learning, evaluation of educational technologies (including CALL evaluation), and language program evaluation. To do this, we draw first on relevant material from ‘mainstream’ disciplines before focusing more closely on issues inherent in modern language programs specifically. Eventually, we come to discuss a way forward and adopt concepts from several sources. We conclude that efforts to evaluate blended programs would be best served by an ‘action research’ approach (Bax, 2011) that is developmental rather than formative or summative (Patton, 2011). To examine blended learning, we rely on work by Gruba and Hinkelman (2012) that provides a way to see programs as layered at macro, meso, and micro levels. Additionally, their work sets out four considerations – purpose, appropriateness, multimodality, and sustainability – that can

help to focus an evaluation project. Finally, we note that the use of arguments as a way to examine technologies can help to situate an evaluation project (Chapelle, Cotos, & Lee, 2015; Gleason, 2013), especially as CALL continues to spread (Chapelle, 2001a).

Based on what we learned earlier, Chapter 2 of the book sets out our framework for blended language program evaluation. In turn, we explain how program evaluators alongside with key evaluation stakeholders can conduct evaluation by determining a level, a key consideration, and then developing an argument in a logical and purposeful manner.

Gathering evidence

In the next part of this book, we present three chapters that seek to help evaluation teams work at the micro, meso, and macro levels of a project.

Throughout Chapters 3–5, we discuss aspects of *gathering the evidence* as we take into account ethical issues, the evaluators' positions, possible stakeholder concerns, and the many early aspects that need to be considered when preparing for an evaluation. In these chapters, we describe how to conduct interviews and observations, select documents and policies, and make and deliver surveys. Following this, we present ways to analyze qualitative and quantitative data.

Our work across each of these three chapters in some way blends with the work of the other two chapters. In writing this material, we acknowledge that a survey, for example, could be implemented not just at the micro level, but also at the meso and macro levels. Interviews could take place with a wide range of people in an institution. Documents, whether they be lesson plans or national policies, can be analyzed at the macro, meso, and micro levels. We understand that methods and techniques are adaptable; nonetheless, we thought that emphasizing one level at one time might help to focus the efforts of an evaluation team. To reduce the complexity of a project and focus efforts, we believe that it is best to break projects into three levels – the distinction we attempt to demonstrate in Chapters 3–5.

Early in Chapter 2, we adopt a four-stage approach for our framework: that is, planning an argument, gathering the evidence, presenting the argument, and finally appraising the argument. In Chapters 3–5, we focus our efforts on the first two stages: planning an argument and gathering the evidence. We do this to illustrate how planning an argument-based evaluation and gathering evidence could be approached at the micro, meso, and macro levels. Here, we set out ways to plan a project

and gather data in order to help an evaluation team get started in their collaboration with key evaluation stakeholders.

Each of the three chapters has a similar structure. First, we express the reasons an evaluator may want to focus on a particular level. Next, we discuss ethical considerations before mapping out a sample structure of an argument. In the latter part of each chapter, we present techniques for gathering evidence and discuss analyses.

In Chapter 3, we focus our attention on micro-level evaluation. In our current framing, we see the micro level concerned with the activities of the classroom, and the attitudes of teachers and students towards the use and integration of technologies. Chapter 4 addresses the challenges of a meso-level evaluation project, and highlights the unique nature of academic department who may become involved through focus groups and participatory design approaches. The focus of Chapter 5 is on the macro level – an area often filled with policies and senior administration; in this chapter, our focus is on document and policy analysis that can be informed through online interviews and member checks with key evaluation stakeholders.

Presenting an argument

In the next four chapters (Chapters 6–9) our aim is to investigate the ways that an argument-based approach to blended language program evaluation actually works in practice. To this point, our book is largely theoretical as we took on the development of a preliminary framework and begs a central question: Does the framework work with real people in real programs across a range of diverse contexts? To address this question, we put an argument at the core of our efforts in case studies across three program levels.

Importantly, note that the purpose of the four case studies is to examine the viability of the proposed framework. We do not, in the strictest sense, conduct developmental evaluation projects with our key evaluation stakeholders; furthermore, we do not set out recommendations for our institutions and programs on ways that we would suggest their programs could be developed. Throughout, our work is reflexive and experimental: Do our own ideas work in practice?

One strength of these case studies, we believe, is that each is situated in a context that differs from the others. We believe, too, that another strength is that we see our own ideas a bit differently and thus interpret the concepts differently. We hope that our readers will often be able to ‘see themselves’ in our various contexts, and be inspired to transfer ideas

found here into the evaluation of their own programs. We are similar to our intended readers: Each of us has had extensive language teaching experience; each of us has been alternatively thrilled and frustrated with technology; and each of us has limited time and resources to devote to language program evaluation. Perhaps like many of our readers, too, we are new to language program evaluation, but understand that it is an important and integral part of development.

Micro-level evaluation is the focus of Chapter 6. Working in a Chilean university, Cárdenas-Claros evaluates blended approaches to language learning and teaching at the English Language Teaching program at Pontificia Universidad Católica de Valparaíso (PUCV) in Chile, using our proposed argument-based evaluation framework.

Chapters 7 and 8 are situated in Hawai'i and Vietnam, respectively, and each takes on the challenges of meso-level evaluation. In Hawai'i, Suvorov conducts an evaluation in the English Language Institute at the University of Hawai'i at Mānoa, USA, focusing on sustainability of blended approaches.

The case study of Vietnam in Chapter 8, conducted by Rick, also looks at the meso level. Here, in a 'transnational' context, Rick evaluates an English for Academic Purposes program at RMIT University Vietnam, an offshore branch campus of RMIT University, Australia. Her focus is on issues of implementation and sustainability of blended approaches.

The University of Melbourne in Australia is the site of the case study of macro-level evaluation presented in Chapter 9. In this chapter, Gruba looks at a range of national and institutional policies to investigate issues of purpose and sustainability in blended language programs at the macro level.

Appraising an argument

The last chapter takes on two questions: 'What have we learned?' and 'What is the way forward?' To address the first question, we summarize the key points of our findings, point out the benefits of the proposed argument-based approach to blended language program evaluation, and reflect on the challenges and limitations associated with this approach.

The second question has been building as we worked on the book for some time. We know that our efforts here can be seen as preliminary, and clearly they need to be further improved, developed, and investigated. In response to this question, we propose our vision of the way forward by outlining several directions for future work in the area of blended language program evaluation.

Audience

We hope that a range of language professionals will make use of this book. Specifically, we are aiming to help practitioners who are seeking to improve a current program, but researchers interested in blended learning may also find the book useful for their investigations. For the practitioner, we understand that the resources for an evaluation may be limited, and that often evaluation is done by an individual or small team who is considered an ‘insider’ conducting an internal evaluation.

This book can be used in both formal and informal settings. In formal situations, graduate students may use it as a basis for a course in language program evaluation. Administrators and managers may also like to read it to gain an in-depth understanding of the issues that may arise, and what is likely to be involved to undertake a complex evaluation. Informally, experienced language professionals may want to consult the work as a complement to their existing practices and techniques. Blended learning is an emerging area that may be new to an evaluator, and we hope that this book can provide a solid introduction to a challenging area of applied linguistics.

Paul Gruba
Mónica S. Cárdenas-Claros
Ruslan Suvorov
Katherine Rick

Acknowledgments

As a group, we would like to acknowledge the work of our colleagues at each of the four universities who participated in our case studies of program evaluation that we conducted to check our work and to provide examples of blended approaches. We are very grateful to Carol Chapelle for her insightful feedback on an earlier draft of the argument-based framework used in this book, as well as for her generous encouragement, support, and motivation. Cathie Elder, too, helped us with her generous advice and assistance throughout our process of writing.

On a personal note, Paul would like to thank Tobias and Noriko for their love and support; and Andrew Boileau was a constant source of inspiration. Paul appreciates the support of several colleagues at the School of Languages and Linguistics, and the generous funding from the Faculty of Arts, The University of Melbourne.

At the PUCV, Mónica would like to thank colleagues at the Institute of Literature and Language Sciences who helped her throughout the different stages of this project, and particularly Marianna Oyanedel for her work at earlier stages of the project and for her involvement in Chapter 6. She would also like to express her thanks to Mariángela Sardes for her meticulous work with references. Mónica is extremely grateful to Esteban Gil, her husband, for the innumerable late-night conversations about the book and for simply being there and taking care of the girls. Special thanks go to Isabella and Valentina for all the weekends and laughs Mónica still owes them. Mónica's early work on this project was funded by an internal grant Proyecto PUCV de Iniciación Colaborativa No. 37.0/2014.

Ruslan would like to say a big Mahalo to the faculty and administrators of the English Language Institute and his colleagues at the Center for Language and Technology, University of Hawai'i at Mānoa, for their support. In addition, Ruslan would like to thank Paul Gruba for an opportunity to be part of this exciting project.

Katherine would like to thank her husband, Brendan, for his unwavering love and support. He knows far more about this topic than he rightfully should, which will no doubt manifest in an abundance of interesting dinner conversation! She would also like to thank colleagues

in the Centre of English Language at RMIT University Vietnam for their support. Finally, Katherine would like to express her sincere gratitude to Paul Gruba for inviting her to be a part of this project. It is probably every graduate student's dream to one day collaborate with their supervisor – and for Katherine the dream became reality.

1

Situating Blended Language Program Evaluation

Blended language program evaluation can be situated at the crossroads of blended language learning, evaluation of educational technology, and program evaluation. In this chapter, our aim is to introduce each of these areas in a way that provides historical context, highlights key concepts, and seeks a synergy among the three areas that can inform blended language program evaluation.

Blended language learning

Blended language learning is a relatively young area, and its roots can be found in a range of work dealing with the use of educational technology for the purposes of learning an additional language that is widely known as Computer-Assisted Language Learning, or CALL (Davies, Otto, & Rüschoff, 2012). Recent prominent books in CALL include those by Chapelle (2001a); Levy and Stockwell (2006); Stockwell (2012); Thomas, Reinders, and Warschauer (2012); and Walker and White (2013). Specialists in CALL regularly publish their research in journals that include *Language Learning & Technology*, *Computer Assisted Language Learning*, *ReCALL*, and *CALICO Journal*.

The trajectory of CALL has traditionally moved in tandem with pedagogical trends in the larger field of applied linguistics that were grounded in behaviorist, cognitive, and social views of language learning (Chapelle, 2001a). Behaviorist perspectives, which highlighted the rote learning of grammar and vocabulary as central to success, saw computers as ‘patient tutors’ that delivered material to students to enhance memorization. Cognitive views of language learning were enacted through the use of computer applications that allowed for the practice of learning strategies to develop key skills in language. Many CALL specialists now underpin

their work with social views of language acquisition that promote the use of technology to enable purposeful interaction between learners and computers, and learners and their peers in projects that may involve a range of global partners (Thomas et al., 2012).

As social views of CALL were increasing and technology became more commonplace, Bax (2003) promoted a view that ‘normalisation’ should be a central goal of CALL. To provide context for his argument, Bax (2003) saw that trends in technology use could be divided into three general categories: Restricted, Open, and Integrated CALL. In the first two approaches, the use of technology appears to have been promoted ahead of learner needs such that entire lessons were devoted solely to computer-based activities. In Integrated CALL, learner needs are highlighted and technology forms a small part of lessons. Ideally, Bax (2003) argued, technology should ‘disappear’ into language teaching and learning and thus be no more distinguished than everyday items like pencil and paper.

As computers became more commonplace, Garrett (2009) defined CALL as ‘the full integration of technology in language learning’ that could be understood within ‘a dynamic complex in which technology, theory, and pedagogy are inseparably interwoven’ (pp. 719–720). Increasingly then, distinctions between language teaching approaches that use, or do not use, technology are meaningless (Chapelle, 2010, 2014); as Bax (2003) predicted, technology did indeed rise, but is disappearing into the language curriculum. A key challenge now, according to Oxford and Jung (2006), is to change approaches to language teacher education and TESOL programs themselves in ways that would align practices with prevailing US national guidelines for technology integration.

What holds back such deep integration? According to Bax (2003), one barrier to integration is that language educators tend to be either in ‘awe’ or ‘fear’ of technology. If language teachers are in awe, they tend to see technology as a powerful agent that alone can change an otherwise staid curriculum; if in fear, teachers tend to remain hesitant to use technology in ways that are productive with learners. Similarly, Oxford and Jung (2006) found that instructor attitudes, a lack of experience with technology, and poor technological infrastructure was holding back integration. To encourage normalization, and thus foster the dissemination of technology in language education, Chambers and Bax (2006) set out three reasons to promote integration as a central goal of CALL. First, they argued, the concept of normalization allows CALL specialists to make use of a wider range of literature concerned with educational

change, which sees technology as one of many ongoing innovations in education, and thus not unique. Secondly, adopting normalization as a fundamental perspective encourages researchers to consider work on innovation into their own studies, and hence profit from the many insights offered in that rich body of literature. Perhaps most importantly, though, an emphasis on normalization would focus CALL practitioners on a clear aim and agenda that eventually would come to 'mean that computers will be at the centre of no lessons, but will play a part in almost all' (Chambers & Bax, 2006, p. 467).

Based on a longitudinal study of two language programs, Chambers and Bax (2006) recommended that normalization promoted by attending to program logistics, stakeholder understanding and abilities, changes in the curriculum, and an improved training and support of educators. Importantly, they concluded, no single factor alone would foster blended learning because language programs were complex ecologies in which a wide range of factors influenced innovation and change.

One way to situate blended language learning is to understand that much work in the area seeks to foster the normalization and integrated use of technology (e.g., Gruba & Hinkelman, 2012; Sharma & Barrett, 2009; Tomlinson & Whittaker, 2013). In addition to these goals, one central motivation for deep integration of technology is to respond to persistent calls for language educators to integrate 'new literacies', 'multiliteracies', or 'digital literacies' to bolster student abilities that are essential to 21st century learning (R. Jones & Hafner, 2012; Lotherington, 2007). Briefly, the term 'new literacies' (and its associated phrases) embraces a range of concepts that pertain to the abilities of understanding, navigating, and creating digital texts. Without guidance and exposure to new technology, Lotherington and Jenson (2011) argue, students are trapped in a traditional curriculum that emphasizes 'traditional' or 'flat' literacies that privilege print materials and discourage productive questioning, exploration, and participation amongst learners. To move away from staid approaches, educators will need to redefine how they teach, learn, and assess languages in the 21st century (Lotherington & Jenson, 2011). Recent examples describe how students produce digital video clips as they learn to make sense of texts as 'multimodal ensembles' (Hafner, 2014); in German and Italian universities, Marenzi (2014) provides a comprehensive study of ways in which multiliteracies are taught in programs grounded in Content and Language Integrated Learning (CLIL) and English for Specific Purposes (ESP) approaches.

In sum, blended language learning is situated in the wider field of CALL that itself is an established sub-discipline of applied linguistics,

concerned with the uses of technology in language learning. Blended language learning has a primary focus on technology integration and thus responds to calls for ‘normalisation’ (Bax, 2003) that is now a defining purpose of the field (Garrett, 2009).

Alongside face-to-face interactions and activities, technology is a major part of blended language learning. An understanding of the efforts to evaluate technology, whether in isolation or as part of a wider program, is the basis of our next section, as we continue to situate blended language program evaluation.

Evaluating technology

As we start to review the evaluation of educational technology, we take note of Zhao (2003) who reminds us that ‘technology is an ill-defined concept that encompasses a wide range of tools, artifacts, and practices, from multimedia computers to the Internet, from videotapes to online chat rooms, from web pages to interactive audio conferencing’ (p. 8). Because technologies vary so widely in their capacity, and in the ways they are used, and in how they are accessed, it is, for example, ‘misleading to think the effects of videotapes are the same as those of the online chat rooms just because they are all called “technology”’ (Zhao, 2003, p. 8). Debates about what is, or the extent of, a ‘technology’ are ongoing. In their discussion within the context of blended learning, Gruba and Hinkelman (2012, pp. 14–16) argue for a wide view of educational technology as a way to understand that it is not simply a tool, but rather provides entire environments designed for the stimulation of learning.

The use of technology in teaching and learning can be understood along a continuum (Chapelle, 2007). At one end of the spectrum, concerns are focused on human–computer interactions to examine how effective technology is in meeting educational principles and objectives; at the other end, technology is examined to determine how well it is integrated (or ‘normalized’) into an overall educational program. The two ends of the spectrum are complimentary, and an awareness of the differing emphases in the uses of technology can help to guide our reading of studies in the evaluation of educational technology.

As we continue to develop our understanding, it is useful to examine concepts and ideas from other areas of education including those by Garrison and Anderson (2011); Luckin (2013); Phillips, McNaught, and G. Kennedy (2011); Ruhe and Zumbo (2009); and Spector, Merrill, Elen, and Bishop (2014). Journals that publish research on educational technology, as well as on their evaluation, include *British Journal of Educational*

Technology, Journal of Educational Technology and Society, Journal of Educational Technology Systems, and Journal of Educational Media.

An example of contemporary approaches to the evaluation of technology can be found in the work of Phillips et al. (2011). Throughout their work, they intermingle concerns of research and evaluation, noting that the two areas often intersect and share similar concerns in an effort to determine the 'effectiveness of e-learning artefacts and learning environments' (Phillips et al., 2011, p. 46). To do this, they suggest that projects focus on the evaluation of one of four areas: (a) the potential of new technology based on a judgement of potential, (b) the design and development of artifacts based on project quality criteria, (c) the improvement of an e-learning environment based on criteria for improvement or formative measures, or (d) the overall effectiveness of an e-learning environment based on measures of impact for a summative purpose. Phillips et al. (2011) identify eight types of stakeholder in an e-learning evaluation that includes learners, members of the university chancellor, employers, and professional accrediting bodies; of note, they argue that a 'project cannot be considered successful if its outcomes are not disseminated widely' (p. 164). Accordingly, in addition to producing research publications, an evaluation team should seek to discuss project outcomes and engage stakeholders in ways that encourage sustainable practices, build capacity, and enhance the quality of the teaching and learning with technology.

Spector (2014) adopts a logic model to frame and organize an evaluation project. Logic models have a long history in program evaluation (Gugiu & Rodríguez-Campos, 2007; Mertens & A. Wilson, 2012), and are used to portray the inputs, outputs (in terms of activities and participation), and the short-, medium-, and long-term outcomes of a project that is often presented through a schematic drawing. Spector (2014) suggests that evaluation studies focus on the fidelity of an implementation, or one that develops measures of adequacy to be used for formative purposes, or an impact study that seeks to determine how well a project has performed. Logic models help to depict a theory of change showing ways in which a problematic situation can be transformed to a more desirable state; for Spector (2014), one responsibility of an evaluator is to detect problematic areas of a project and seek to resolve them.

Similar to the work of Chambers and Bax (2006), Graham, Woodfield, and Harrison (2013) sought to identify factors that may promote or impede blended learning in three focal areas: strategy, structure, and support (Table 1.1)

Table 1.1 Blended learning implementation categories

Theme	Description
Strategy	Issues relating to the overall design of blended learning, such as definition of blended learning, forms of advocacy, degree of implementation, purposes of blended learning, and policies surrounding it
Structure	Issues relating to the technological, pedagogical, and administrative framework facilitating the blended learning environment, including governance, models, scheduling structures, and evaluation
Support	Issues relating to the manner in which an institution facilitates the implementation and maintenance of its blended learning design, incorporating technical support, pedagogical support, and faculty incentives

Source: Adopted from Graham et al., 2013.

For the evaluator, the work of Graham et al. (2013) can serve as a way to orient the direction and focus of a project. Is the lack of a clear institutional strategy to blame for the lack of uptake of blended approaches? If faculty is provided with a range of support mechanisms, why isn't technology better integrated into modern language courses? Graham and colleagues (2013) point out that it is important to take into account the current stage of adoption for an institution, as shown in Table 1.2.

The identification of these stages implies that adoption of blended learning approaches is a gradual process that cannot be easily characterized by clear start and end points. Therefore, for Graham et al. (2013), the institutional adoption of blended approaches is best seen as an ongoing, developmental process rather than a finite event. For the evaluator then, a blended learning evaluation project cannot be 'summative'; rather, it is better to see evaluation activities as one part of a larger effort to develop a blended approach to learning.

Building on the work of Graham et al. (2013), Porter, Graham, Spring, and K. Welch (2014) examined the ways in which US higher education institutions were making the transition from the stages of 'awareness and exploration' to one of 'adoption and early implementation'. In their study, they noted that institutions oftentimes ignore the evaluation of blended programs: Of the 11 universities under investigation, the majority had not adapted their evaluation protocols to suit the specific requirements for evaluating blended courses. Eight of the universities had relied on existing course evaluation methods, three created specific